Diss. ETH No. 14360

Constraint Optimization Networks for Visual Motion Perception - Analysis and Synthesis

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZÜRICH
for the degree of
DOCTOR OF NATURAL SCIENCES

presented by
ALAN ALFRED STOCKER
Dipl. Masch.-Ing. ETH
born January 12th 1970
citizen of Schönenberg, Switzerland

accepted on the recommendation of
Prof. Dr. Rodney Douglas
Prof. Dr. Eric Vittoz

2001

Acknowledgments

As everything in life, this thesis did not emerge out of nothing and by isolation but rather reflects the enormously rich and fruitful environment it grew on. Many institutions and individuals were responsible to generate this environment of which I would like to thank in the following the most important and influential ones:

I specially thank my advisor Prof. Dr. Rodney Douglas for generously providing me with excellent facilities to explore my ideas in far reaching and splendid freedom. In many ways, he showed me what computational neuroscience could be all about – and what not.

Many thanks go to my co-supervisor Prof. Dr. Eric Vittoz for his time and interest for day-lasting discussions in topics of this thesis. I always very much appreciated his hospitality at CSEM.

I am very grateful to Dr. Jörg Kramer who became more than just my office mate over the past years. He introduced me to analog circuits and circuit design and exhibited an immense patience to answer all my many questions. He provided valuable suggestions to the manuscript and did a careful proof reading.

Not less, I would like to express my gratitude to Dr. Tobias Delbrück for his contagious enthusiasm and his very profound expertise in circuits and circuit technology that he generously shared with me.

Nevertheless, my thanks go to all the other members of the Institute of Neuroinformatics for jointly providing such a stimulating research environment.

I am deeply thankful to Caroline for sharing with me all the good and bad moments during this period of my life.
I would like to thank also Andreas, Martin and Marc for being good friends with a common fate which made it often easier to carry on.
And finally, I would like to express my very deep gratitude to my family who always believed in me.

Contents

Zusammenfassung

Die Extraktion der visuellen Bewegungsinformation ist ein Beispiel sensorischer Informationsverarbeitung, das typischerweise eine hohe Datenbandbreite und schnelle Verarbeitung verlangt. In der Natur existieren viele Beispiele von Systemen, die visuelle Bewegung bei weitem effizienter verarbeiten als jedes von Menschen geschaffene künstliche System. Biologische Systeme verwenden physikalische Berechnungsarchitekturen, die aus Netzwerken von stark miteinander verknüpften, parallel arbeitenden einfachen Einheiten bestehen. Diese Netzwerke repräsentieren eine grundlegende Art der Informationsverarbeitung, die nicht nur für sensorische sondern für jegliche informationsverarbeitenden neuronalen Strukturen gilt – wie etwa für unser Gehirn. Die Motivation, die Berechnungsaspekte solcher Netzwerke zu verstehen ist demzufolge zweifach begründet: Erstens ist es eine grosse Herausforderung unserer Zeit zu verstehen, wie unser Gehirn funktioniert. Das Studium kleiner Netzwerke, die sehr spezifische Aufgaben lösen, kann uns helfen, die grundlegenden Rechenprinzipien von komplexeren Systemen besser zu verstehen. Und zweitens macht es die offensichtliche leistungsmässige Überlegenheit dieser biologischen Netzwerke sehr reizvoll, diese Rechenarchitekturen in moderne Technologie zu übertragen um eine bestimmte Art von angewandten Rechenproblemen effizient zu lösen.

Die vorliegende Dissertation präsentiert und analysiert einfache Netzwerklösungen für verschiedene Probleme der Bewegungswahrnehmung. Es wird postuliert, dass die Wahrnehmung visueller Bewegung ein Optimierungsproblem darstellt, das notwendigerweise durch die Mehrdeutigkeit der visuellen Information in Bezug auf die Wahrnehmung der visuellen Bewegung existiert. Der Optimierungsprozess besteht darin, eine Interpretation der visuellen Bewegung zu finden, die der beobachteten visuellen Information am besten entspricht, unter Beachtung der im Netzwerk konstituierten internen Bewegungsmodelle. Eine nicht hierarchische analoge Netzwerkarchitektur wird vorgestellt, die kontinuierlich eine optimale Schätzung der lokalen visuellen Bewegung (auch optischer Fluss genannt) liefert. Es wird ausführlich und genau gezeigt, dass sich dieses Netzwerk immer in einem einzigartigen und klar definierten optimalen Zustand befindet, unabhängig aller möglichen visuellen Eingangsbedingungen, was es von anderen, vorgängig vorgeschlagenen Lösungen unterscheidet. Simulationen des Netzwerkes mit realistischem visuellen Eingangssignal zeigen eine plausible und robuste Schätzung des optischen Flusses. Die Charakteristik

des geschätzten optischen Flusses hängt dabei stark von zwei globalen Parametern σ und ρ ab, die die isotrope Verbindungsstärke zwischen den einzelnen Elementen des Netzwerks und den Einfluss von *a priori* Annahmen bezüglich der wahrgenommenen visuellen Bewegung bestimmen. Je nach Wert dieser Parameter variiert die Charakteristik des optischen Flusses kontinuierlich zwischen einer globalen Schätzung der visuellen Bewegung, wobei in diesem Fall das Aperturproblem für ein einzelnes visuelles Objekt gelöst wird, und der Schätzung von senkrechtem optischen Fluss, die dann resultiert, wenn die Verbindungen zwischen den einzelnen Elementen des Netzwerkes vollständig unterbrochen sind.

Zwei erweiterte Systeme werden vorgestellt, in denen zusätzliche Netzwerke rückläufig mit dem zuerst vorgestellten Netzwerk zur Schätzung des optischen Flusses verbunden sind, um dynamisch die lokalen Werte der Parameter ρ und σ zu kontrollieren. Das erste dieser erweiterten Systeme, das bewegungssegmentierende Netzwerk, findet Bewegungsunstetigkeiten und beschränkt in Folge davon den kollektiven Schätzungsvorgang des optischen Flusses auf diejenigen Elemente im Netzwerk, von denen angenommen wird, dass sie visuelle Information desselben Objekts im Raum erhalten. Das Netzwerk ist fähig, nahezu optimale Lösungen des rechentheoretisch harten Problems der Bewegungssegmentation zu finden. Das zweite System, das bewegungsselektive Netzwerk, liefert die gezielte Wahrnehmung nur derjenigen visuellen Bewegung, die einer vorgegebenen Bewegungspräferenz entspricht. Dies ermöglicht eine aufmerksamkeitskontrollierte Wahrnehmung der visuellen Bewegungen.

Der zweite Teil dieser Dissertation vervollständigt den Analyse-Synthese-Kreislauf, indem gezeigt wird, wie zwei der vorgeschlagenen Netzwerkarchitekturen in einem physikalischen Substrat eingebettet werden können. Analoge, in sehr grossem Massstab integrierte (aVLSI) elektronische Schaltkreise des Netzwerks zur Schätzung des optischen Flusses und des bewegungssegmentierenden Netzwerkes werden vorgestellt, die unter natürlichen visuellen Bedingungen voll funktionsfähig sind. Diese beiden integrierten Schaltungen gehören zu den wenigen Beispielen von angewandten kollektiven Berechnungsprozessen in Sensorsystemen, wobei sie wahrscheinlich die leistungsfähigsten und komplexesten Systeme ihrer Art darstellen. Die gemessenen Charakteristiken der Schaltungen bestätigen die Möglichkeit, solche analogen Netzwerke für die effiziente Lösung bestimmter Klassen von Problemen der Wahrnehmung in praktischen Anwendungen einzusetzen.

Abstract

The extraction of visual motion information is an example of sensory information processing which typically has to deal with high data bandwidths and fast processing requirements. Nature provides us with many examples of visual motion processing systems that are far more efficient than any man-made artificial system. Biological systems use a physical computational architecture that consists of networks of highly interconnected simple units that all work in parallel. These networks represent a generic way of information processing that applies not only to sensory processing (sub-)systems but to any processing in neural structures such as our brain. The motivation to understand the computational aspects of such networks is therefore two-fold: Firstly, it is one of the great challenges of our time to understand how the brain works. Studying small networks that solve particular tasks might help us to understand the basic computational principles of more complex systems. Secondly, the obvious superiority in performance of these networks makes it very appealing to transfer such computational architectures into technology in order to provide efficient solutions to a particular class of computational problems.

This thesis presents and analyzes simple network solutions for different motion perception problems. It postulates that the perception of visual motion must be understood as an optimization problem. Optimization is necessary to deal with the ambiguity between the visual information and the perception of visual motion. It is meant to find the interpretation of visual motion that best fits the observed visual information according to the motion model built into the network. A non-hierarchical analog network architecture is proposed that is continuously providing an optimal estimate of the local visual motion (also called optical flow). It is rigorously shown that this network is always in a unique and well-determined optimal state independent of its visual input conditions, which is in contrast to other, previously suggested solutions. Simulations of the network behavior with realistic visual input show a plausible and robust estimation of the optical flow field. The characteristics of the estimate depend strongly on two global parameters ρ and σ, that determine the isotropic connection strengths between units in the network and the influence of some *a priori* assumption about the perceived visual motion. According to the values of these parameters, the characteristics of the estimated optical flow field vary continuously between a global motion estimate, in which case the aperture problem is

solved for a single visual object, and a normal optical flow estimate that results when the connections between units of the network are completely disabled.

Two extended systems are introduced where additional networks are recurrently connected to the basic optical flow network in order to dynamically control the local values of the parameters ρ and σ. The first system, the motion segmentation network, finds motion discontinuities and restricts the collective estimation process of the optical flow to those units that are considered to receive visual information of the same object in space. The network is able to provide close-to-optimal solutions for the computational hard problem of motion segmentation. The second system, the motion selective network, provides a selective perception of visual motion according to a given motion preference. This provides the means for an attentional control of the perception of visual motion.

In the second part of this thesis, the analysis-synthesis loop is closed by demonstrating how some of the proposed network architectures can be embedded in a physical substrate. Analog Very Large Scale Integrated (aVLSI) circuit implementations of the optical flow and motion segmentation networks are presented that are fully functional under real-world conditions. The two circuits are among the few examples of sensory systems that apply collective computation and represent probably the most powerful and complex implementations of their kind. The measured characteristics of these circuits prove the feasibility of using physical analog network architectures to solve a particular class of perceptual problems efficiently in practical applications.

Chapter 1

Introduction

Our world is a **visual world**; our visual sense is by far the most important one to gather information from our environment. Light reflected from objects in our environment is a very rich source of information. The short wavelength and high transmission speed of light allow a spatially accurate and fast localization of the reflecting surfaces in space. Spectral variations in wavelength and intensity of the reflected light resemble the physical properties of objects and provide means to recognize them. The light-sources we are exposed to are usually inhomogeneous. Thus shadows and reflectances are highly correlated with the spatial dimension of the objects which help us to identify our environment.

Our world is also a **world of motion**. We are moving creatures. We have to be able to navigate successfully through our environment in order to survive. And we predominantly use our visual perception to do so. It is crucial for us to have a sense of motion, to perceive our own motion in relation to the environment as well as the relative motion of objects to each other. We need to determine quickly what is moving where, in which direction, and at which speed. It is clear that there is a strong requirement for a motion system that is able to process visual information quickly and extract the important motion information. Therefore it is not surprising that the visual system in general, and its motion processing stage in particular, occupy a substantial fraction of our central processing unit, the brain.

When mankind started the enterprise to build artificial autonomously behaving agents, it was a common believe that by year 2001, autonomous agents would be part of our everyday life in various aspects as proposed in numberless science-fiction stories and movies (*e.g.* 'R2-D2' in star wars)[1]. We now are wiser. Whoever pursued NASA's recent mars Pathfinder mission[2], or demonstrations of 'artificial pets'[3], immediately acknowledges that these brittle and slow 'state-of-the-art' agents are far from fulfilling our expectations.

There are two key reasons that are strongly interrelated which are responsible for the

[1] www.starwars.com

[2] http://sse.jpl.nasa.gov/missions/mars_missns/mp.html

[3] *e.g.* the robot-dog AIBO from SONY ,http://www.world.sony.com/Electronics/aibo/index.html

poor advances so far. Firstly, it is not yet clearly understood what it means to perceive; what the computational basis of perception is and, so what intelligence is. And secondly, it seems that the computational machinery used in artificial autonomous agents is not efficient in solving perceptual tasks. Up to now, computer vision and robotics have almost exclusively relied on sequential and binary computational architectures that are based on *von Neumann's* [von Neumann 1945] general purpose computing device. Presumably, all computable problems can be formulated such that they are solvable by such computing devices. However, it has to be seriously doubted that such devices are computationally efficient in solving perceptual tasks. In particular, nature provides us with various examples of wonderfully engineered, efficient systems that are the existence proof of other means of computation and computational architectures. Efficiency is crucial in biological systems and promotes solutions that consume the smallest amount of resources available, namely **time, space and energy**. Biological computational structures simply cannot afford to have a centralized structure where memory and hardware, algorithm and computational machinery are physically separated. Here, the function is the architecture – and *vice versa*. A fair amount of the computational power and efficiency is due to the collective computation in networks of single processing elements. Furthermore neuronal structures use the intrinsic physical properties of their single units and connections, their membrane capacitances and resistances as computational primitives to solve computationally demanding tasks.
These insights motivated scientists about 15 years ago to begin the endeavor of **neuromorphic engineering**, in which they aim not only to understand the nature of physical computation in biological networks but also to exploit these principles and create artificial electronic systems able to solve tasks such as visual motion perception computationally efficiently.

The **mission** of this thesis is to investigate the possibilities to perform 2D visual motion processing in analog networks of simple computational units. It proposes a general understanding of perception as solving optimization problems. Furthermore, the thesis tries to demonstrate that these networks can successfully be implemented in analog VLSI hardware to provide computationally efficient solutions to solve the tasks. Emulating rather than simulating such networks confronts us directly with the physical constraints of the computational substrate and thus forces a different consideration of, and a better understanding of, physical computation in analog networks.

Chapter 2

Visual Motion Processing

Visual motion processing extracts information about the relative motion between an observer and its environment using the visual information the observer can access from the environment. In general, the visual information is provided by an imaging device, embodied in the observer. Since there is a causal relationship between the brightness distribution in the image and the objects in space, the spatiotemporal brightness changes induce a perception of relative motion in the image which in the following will be referred to as visual or **apparent motion**[1].

The representation or encoding of apparent motion can be arbitrary and depends on the computational machinery, the affordable bandwidth and the needs of further processing stages. A dense and thus very flexible form of representation is a vector field $\mathbf{v}(x, y, t)$ that characterizes the direction and speed of apparent motion at each particular time and image location. According to [Gibson 1950] such a flow field is referred to as optic or **optical flow**.

The interdependence between apparent motion and the two-dimensional projection (or 2D motion) of the real three-dimensional motion is not a direct mapping. Only in particular cases is the mapping identical [Verri and Poggio 1989]. The reasons are twofold: Firstly, the information provided by the spatiotemporal brightness changes might be ambiguous or missing altogether. Secondly, the spatiotemporal brightness changes do not have to originate from real physical motion alone. Consider for example a static scene illuminated by a light-beam from outside the visual field of the observer such that shadows of the objects are generated. If the light-beam is moving, the shadows will also move and induce apparent motion, although no real motion is taking place. In general, reflectance properties and scene illumination are dominant factors for how well apparent motion matches 2D motion. However, this also raises the question what a possible benchmark for

[1] This definition is in contrast to some of the psychophysical literature where apparent motion is referred to as the perception of continuous motion of a sequence of image frames with discrete displacements [Braddick 1980].

visual motion system should be. Is a best possible matching of the 2D motion [Barron et al. 1994] the preferred goal? Obviously, the apparent motion of the moving shadows in the above example provides useful information to the observer, namely that the light source is moving. Hence, resolving such ambiguities of apparent motion should be performed preferably in higher level stages.

It is well-known that specified areas of primate's visual cortex encode local image motion in a manner similar to optical flow. The dominant motion sensitive Medial Temporal (MT) area in macaque monkeys, for example, is retinotopically organized in columns of directional and speed sensitive neurons very similar to the orientation columns in the primary visual cortex V1 (see Lappe [2000] for a review). Thus with respect to the computational power of such biological systems it seems rather fair to assume that the estimation of optical flow is a sensible step in the process of building a complete visual processing system.

2.1 Matching Models

In order to define a representation of apparent motion we have to apply some computational model that extracts a quantitative measure given the image data.

Optical flow cannot be observed, it must be computed.

The basic assumption of every visual motion processing system is that a feature in the image at some point in time will also be present somewhere in the image at a later time. Thus, a quantitative description of apparent motion requires a matching process that measures the local image displacements between two successive moments in time. In the following we will distinguish two different classes of models.

2.1.1 Explicit matching models

The first class includes all those models that apply matching explicitly: They extract features and track them in time. The displacement per time unit in image space serves as the measure for apparent motion. Features can be thought of as being extracted on different levels of abstraction, starting from the raw brightness patches to low-level features like edges and up to objects. Typically, these models can be described by the following matching operation:

$$M(x, y; d_x, d_y) = \phi\big(E(x, y, t_0), E(x + d_x, y + d_y, t_1)\big) \tag{2.1}$$

where E is the representation of some features in image space and time, d_x and d_y are the displacement in image space and ϕ is some correlation function that is maximal if $E(x, y, t_0)$ and $E(x + d_x, y + d_y, t_1)$ are most similar. The task reduces to find d_x and

d_y such that $M(x, y; d_x, d_y)$ is maximal. Knowing the relative time difference, the image velocity is directly proportional to the displacements.

Unfortunately, maximizing M might be an ambiguous problem and mathematically *ill-posed*[2]. That is, there may be several solutions maximizing M. We are faced with the so-called *correspondence problem* [Ullman 1979]: A single location on an object in our environment is not necessarily expressed by an unique image feature in our projection. One can object that the application of a competitive selection mechanism always ensures a solution. However, in this case the correspondence problem is just hidden in a way that the selection of the motion estimate in an ambiguous situation is driven by noise and thus *ill-conditioned*. Unfortunately, the extraction and tracking of higher level spatial features does also not circumvent the problem. The correspondence problem is in this case just partially shifted to the extraction process which is ill-posed by itself [Bertero et al. 1987] and partially to the tracking depending on the level of feature extraction; *e.g.* tracking edges might be ambiguous if occlusion happens.

The extraction of optical flow in 2D image space is at least an ill-conditioned if not an ill-posed problem.

Several reasons do not favor a complex spatial feature extraction stage for low-level motion estimation. Fault tolerance decreases because once a complex feature is misclassified, the motion information related to the complete feature is wrong; whereas outliers on low level features might be discarded by some confidence measure, or simply averaged out. Furthermore, since spatial feature extraction takes place before any motion is computed, motion cannot serve as a cue to enhance the feature extraction process.

2.1.2 Implicit matching models

The second class of models contains those that rely on the continuous interdependence between apparent motion and the spatiotemporal pattern observed at some image location. The matching process is only implicitly present.

Gradient based methods assume that the brightness of an image point remains constant while undergoing visual motion [Fennema and Thompson 1979]. Let $E(x, y, t)$ describe the brightness distribution in the image on a Cartesian coordinate system. The Taylor expansion of $E(x, y, t)$ leads to

$$E(x, y, t) = E(x_0, y_0, t_0) + \frac{\partial E}{\partial x}dx + \frac{\partial E}{\partial y}dy + \frac{\partial E}{\partial t}dt + \epsilon \tag{2.2}$$

where ϵ contains higher order terms that are neglected. Assuming the brightness of a moving image point to be constant *i.e.* $E(x, y, t) = E(x_0, y_0, t_0)$, and dividing by dt leads

[2] see Appendix A

to

$$\frac{\partial}{\partial x}E(x,y,t)\ u + \frac{\partial}{\partial y}E(x,y,t)\ v + \frac{\partial}{\partial t}E(x,y,t) = 0 \tag{2.3}$$

where $u = dx/dt$ and $v = dy/dt$ represent the two components of the local optical flow vector.

Equation (2.3) represents the *brightness constraint equation*,[3] which was first introduced in this form by Horn and Schunck [1981]. Obviously, the brightness constraint equation is almost never true. For example, it requires that every change in brightness is due to motion and thus the illumination to remain constant; that object surfaces are opaque and scatter light equally in all directions; and that no occlusions occur. Many of these objections are inherent problems of the estimation of optical flow. Even if the brightness constraint were to hold perfectly, we could not extract a dense optical flow field because Equation(2.3) contains two unknowns u and v. The computation of visual motion using the brightness constraint is thus said to be ill-posed, as are many other tasks in visual processing [Poggio et al. 1985]. Nevertheless, the brightness constraint equation grasps the basic relation between apparent motion and brightness variations and has shown to be a valid first order model. It provides a constraint that continuously relates apparent motion to the spatiotemporal structure of the image data.

A second group of implicit matching models characterize the spatiotemporal nature of visual motion by the response to spatially and temporally oriented filters. A relatively simple model in one spatial dimension was proposed by Hassenstein and Reichardt [1956] after studying the visual motion system of the beetle species *Chlorophanus*. This **correlation method**, which turns out to be of general nature in insect vision, correlates the temporally low-pass filtered response of a spatial feature detector with the temporally high-pass filtered output from its neighbor. Correlation will be maximal if the observed stimulus matches the time constant of the filters. A similar arrangement was also found in the rabbit retina [Barlow and Levick 1965]. Here the output of a particular detector is inhibited by the delayed output of its neighbor located in the preferred moving direction. A stimulus moving in the preferred direction will thus elicit a response which is shut down after the output of the neighboring detector signals the arrival of the stimulus and passed the delay element. In the null direction, the detector is inhibited by its neighbor if the stimulus matches the time constant of the delay element. Correlation methods do not explicitly report velocity. Furthermore, their motion response is phase-dependent.

It has been shown that the correlation based models are computationally equivalent to the first stage of a more general family of models [Van Santen and Sperling 1984]. These so-called **motion energy models** [Adelson and Bergen 1985] apply odd and even type Gabor-filters in the spatiotemporal frequency domain such that their combined output is phase-invariant and reaches a maximum for stimuli of a particular spatial and temporal

[3]Sometimes it is also referred to as the *motion constraint equation*.

frequency. Many of these filters tuned to different combinations of spatial and temporal frequencies are integrated such that they support a particular image velocity [Heeger 1987a, Grzywacz and Yuille 1990]. The motion energy models are similarly affected by the correspondence problem as the explicit matching models. However, since the 'features' are characterized by their spatial and temporal frequencies such an approach is possibly more robust. In any case, none of the above matching models can circumvent the inherent correspondence problem.

2.2 Flow Models

To resolve the ambiguity of optical flow estimation additional constraints have to be imposed on the motion field. Additional constraints can result in parametric flow models and reflect our expectations of the observed type of visual motion. Therefore these models represent *a priori* assumptions of the environment and might be formed by adaptation processes taking place on various time-scales [Rao and Ballard 1996, Rao and Ballard 1999, Mead 1990]. The more complex a model is and thus the more accurate it can describe a particular flow field, the more it lacks generality. The choice of the model is determined by the motion field expected, the type of motion information required, and the complexity of the system allowed. Furthermore, depending on their complexity and thus the required accuracy of the model, flow models permit a very compact and sparse representation of the motion field. Such sparse representation is important *e.g.* in efficient video compression standards that are able to cut down the huge bandwidth in video transmission.

2.2.1 Global motion

First, we consider the modeling of the flow field induced by movements of the observer with respect to its environment. If the environment remains stationary the apparent motion is directly related to the **ego-motion** of the observer. In this case, the observer does not have to perceive his environment as a collection of single objects but rather sees it as a spatiotemporally structured background that allows him to sense his own motion [Sundareswaran 1991].

Three fundamental models can be associated with relative global motion to a fronto-parallel oriented background:

- The simplest flow field imaginable results from pure *translational* motion (Figure 2.1a). The induced flow field $\boldsymbol{v}(x, y, t)$ does not contain any source nor rotation. Thus the divergence $\mathrm{div}(\boldsymbol{v})$, the rotation $\mathrm{rot}(\boldsymbol{v})$ and the gradient $\mathrm{grad}(\boldsymbol{v})$ of the flow field are zero. Such global translational motion can be represented by a single flow vector.

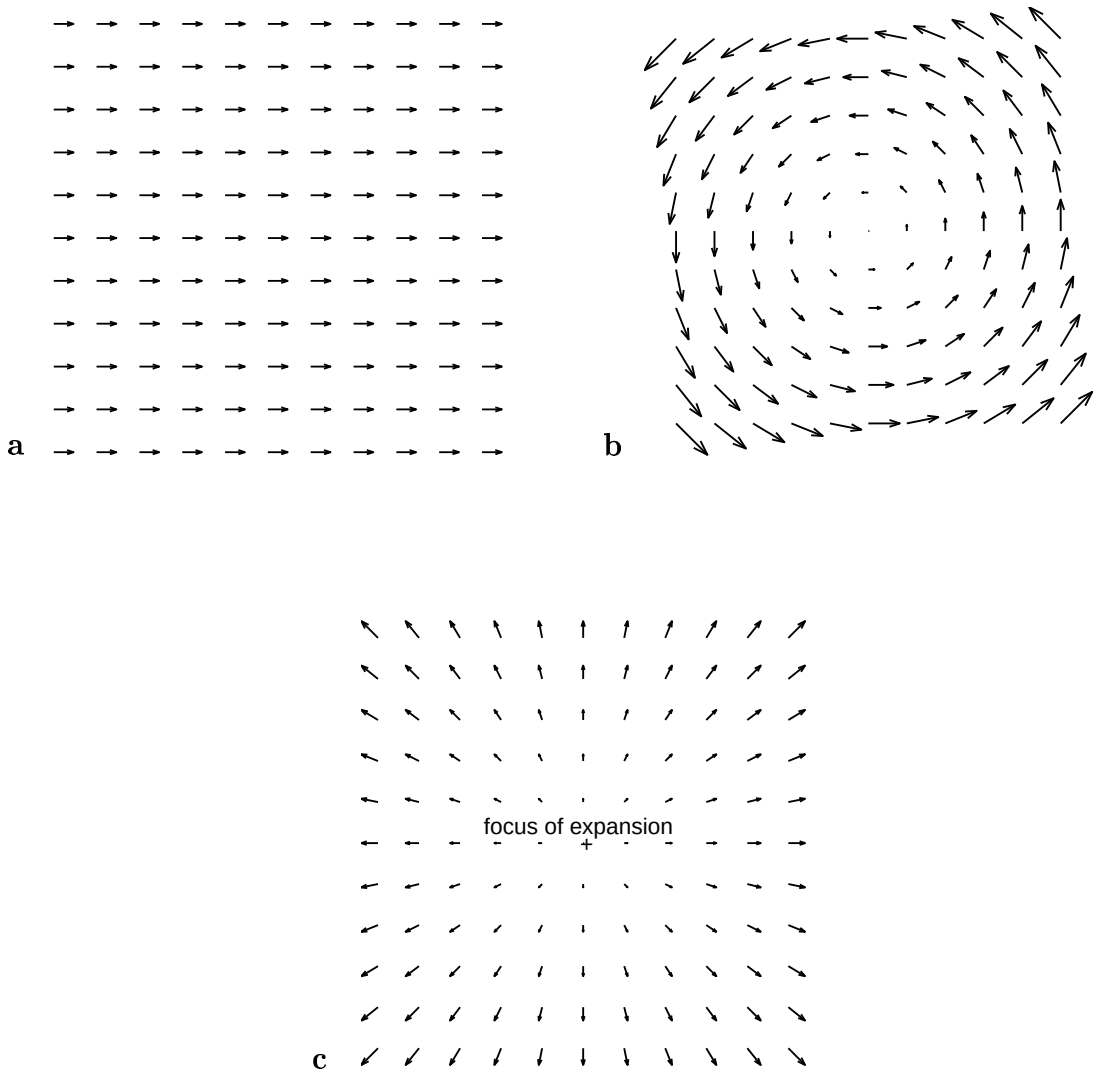


Figure 2.1: *Basic global flow patterns.* (a) A pure translational (a), rotational (b) and radial (c) flow field.

- Pure rotational motion between the observer and the background at constant distance will induce a *rotational* flow field as shown in Figure 2.1b. Again, the flow field has no source but now rotation is present. Clock-wise or counter-clock-wise rotation can be described by the sign of vector $\mathbf{c} = \mathrm{rot}(\boldsymbol{v})$ pointing either perpendicularly in or out of the image plane. Since $\mathbf{c}$ is constant in the complete image space, a single parameter is sufficient to describe the flow field.

- Approaching or receding motion will result in a *radial* flow field (Figure 2.1c) that contains a single motion source or sink respectively. An appropriate model of such a field would be to require the divergence to be constant. For approaching motion, the origin of the source is called **focus-of-expansion** and signals the heading direction of the observer.Again, a single parameter $c_0 = \mathrm{div}(\boldsymbol{v})$ is sufficient to describe the flow field whereas its sign signals approaching or receding motion and its value is a measure for speed.

Many more complicated models of affine, projective or more general polynomial type have been proposed to account for motion of tilted and non-rigid objects (for a review see [Stiller and Konrad 1999]). Furthermore, there are models that also consider the temporal changes of the flow field such as acceleration [Chahine and Konrad 1995]. However, it is clear that only in very special cases can the complete flow field be accurately described with a single model of few parameters. In general, the models will oversimplify the visual motion. Nevertheless, a single model approach can provide a very sparse representation of global motion and it provides a robust measure because it accounts for the whole image space, thus its **region-of-support** is maximal.

There is evidence that the visual system in primates does in fact extract global flow patterns for ego-motion perception [Bremmer and Lappe 1999]. Recent electro-physiological studies show that in regions beyond area MT of the macaque monkey neurons respond to global flow patterns [Duffy and Wurtz 1993]. In the medial superior temporal area (MST) neurons are selective to particular global flow patterns such as radial and rotational flow and combinations of it [Lappe et al. 1996]. Furthermore a large fraction of neurons in area MST receive vestibular input as well as input from MT [Duffy 2000] which suggest that MST is involved in the perception of ego-motion.

2.2.2 Local motion

We now consider the general case where several objects and the observer are moving relative to each other. Several moving objects induce several flow sources resulting in a complex flow pattern. It is not feasible to apply a global flow model to capture such arbitrary pattern. On the other extreme, a purely local measure is possible, but cannot be a very good estimate because of the earlier mentioned correspondence problem of which the

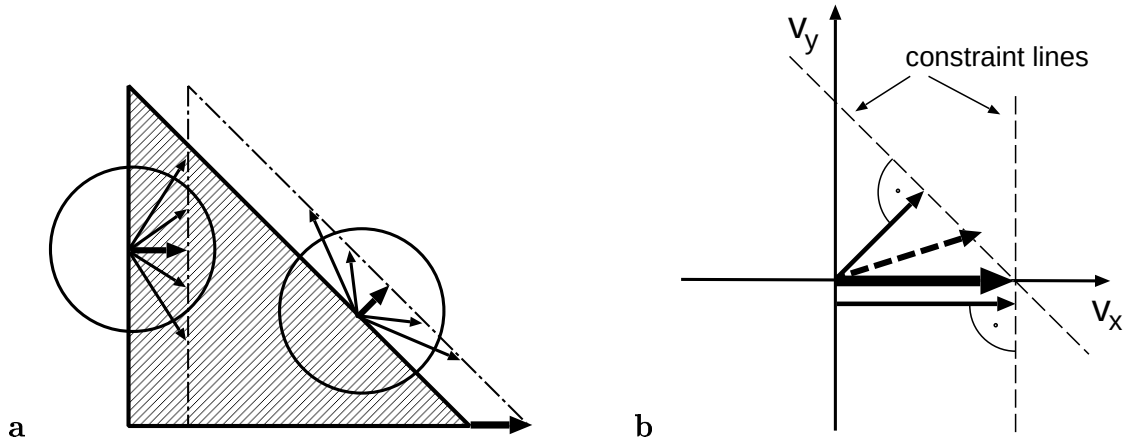


Figure 2.2: *The aperture problem.* (a) A rigid object moving translationally to the right induces local motion that is ambiguous within the two circular apertures. (b) Vector averaging of the normal flow field (dashed arrow) does not lead to the correct global motion. Instead, only the intersection of the constraint lines leads to the correct common object motion (bold solid arrow).

aperture problem is a well-known representative [Hildreth 1983]. To understand the implications of the aperture problem we consider Figure 2.2a: Each of the two circles represent the aperture of a local motion detector. In the present constellation, each detector observes only a local feature of a moving object which is in this case an brightness edge of a particular spatial orientation. The apparent motion perceived within each aperture is ambiguous because it could be elicited by an infinite number of possible displacements of the edge.

Nevertheless, we can obtain a pure local estimate of the optical flow when we apply a flow model that simply chooses the shortest one from the subset of all possible motion vectors (bold arrow in Figure 2.2a). Since the shortest flow vector points perpendicular to the edge orientation it is also called **normal flow** vector. Assuming all possible local motions to occur with equal probability, normal flow estimation is the optimal local motion estimation with respect to the accumulated error in direction because it represents the mean direction[4]. Normal flow can serve as a loose estimate of apparent motion which might be sufficient for some perceptual tasks [Huang and Aloimonos 1991]. However, object motion can only be directly reconstructed from the normal flow vectors for very particular object shapes. In general, the vector average of the normal flow along the object contour provides not a good estimate for apparent object motion (Figure 2.2b). Here, the correct motion is determined by the unique motion vector that is present in both subsets of possible flow vectors, represented by the two constraint lines.

[4]If we apply the dot-product as an error measure for direction *and* speed, the accumulated error remains constant, thus each of the possible flow vectors is an equally sub-optimal estimate.

In general, apparent motion cannot be correctly determined by the vector average of the local normal flow vectors along the object contour and body.

To avoid the complexity of a global model and to overcome the poor estimation performance of a purely local model we have to apply models to regions-of-support in the image space. A partition of the image space into sub-images of fixed size turns the modeling process into several global motion tasks. According to the model complexity and the partition sizes the apparent motion can be captured by a relatively small number of parameters. Several older video compression standards [Stiller and Konrad 1999] used such block-partitioning with simple translational flow models. Obviously a fixed partition scheme will fail for several flow sources present in one partition. In addition, depending on the partition size, the spatial resolution of motion information is coarse.

2.2.3 Motion discontinuities

A general property of biological visual systems is their ability to discard redundant information as early as possible in the processing stream and so reduce the huge information flow. Using antagonistic center-surround type receptive fields, spatiotemporal discontinuities in the visual feature space are transmitted preferably while regions of uniform visual input are hardly encoded. This is well-known for the peripheral visual nervous system, the retina that receives direct environmental input. However, it also holds within cortical visual areas such as V1 (*e.g.* [Hubel and Wiesel 1962]) and higher motion areas like MT/V5 in primates [Allman et al. 1985, Bradley and Andersen 1998]. Furthermore electro-physiological studies provide evidence that motion discontinuities are indeed separately encoded in early visual cortex of primates. Lamme et al. [1993] demonstrated on the awake behaving macaque monkey that motion boundary signals are present as early as in V1. In addition, studies on human subjects using fMRI have shown that motion discontinuities are represented by retinotopically arranged regions of increased neural activity, spreading from V1 up to MT/V5 [Reppas et al. 1997].

Such physiological data strengthens the argument that the special treatment of motion discontinuities plays a significant role already in very early visual stages such as V1. These findings suggest that the detection of motion discontinuities is vital even for the computation of very local visual motion. Since the above experiments use random-dot stimuli that induce unambiguous local motion, it is not clear whether the observed fMRI signals represent only low level motion discontinuities that could in principle be processed solely within V1. If this is not the case and the signals indicate much more object specific properties, the intrinsic ambiguities of apparent motion do require additional information from other higher level areas. It seems rather likely that recurrent, top-down connections do play a major role in the computation of motion discontinuities.

From a computational point of view, motion discontinuities are important for the

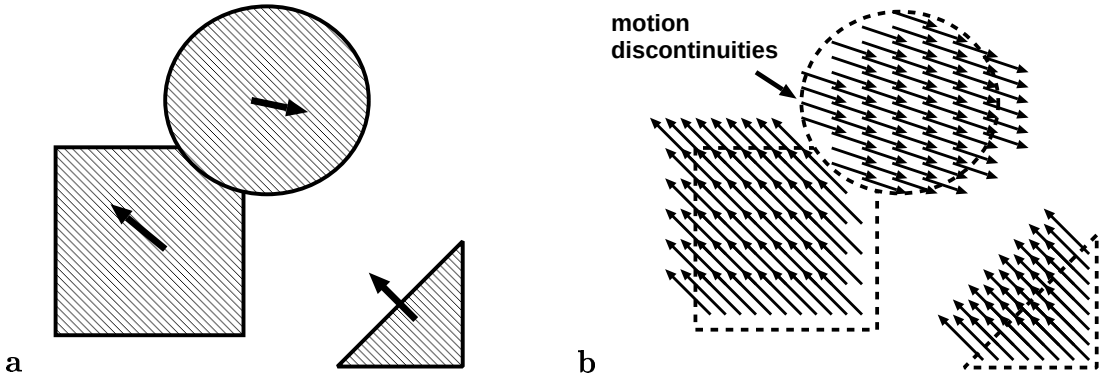


Figure 2.3: *Object Motion.* (a) Three rigid objects undergoing translational motion while the observer is not moving. (b) Ideally, a correct estimation for the optical flow field assumes three independent translational models, restricted to disconnected regions-of-support. The boundaries of these areas represent locations of discontinuous motion between objects as well as between objects and background.

correct extraction of the optical flow, which is illustrated by the example depicted in Figure 2.3. It shows a snapshot of an idealized visual scene of three moving objects taken by a static observer. The background is stationary and the three objects are moving independently. For the sake of simplicity they are assumed to undergo purely translational motion. Thus, the induced flow field of each object is preferably represented by a separate translational flow model. Ideally, the regions-of-support for each model coincide with the visible object outlines. In the case of translational flow, the outlines are the only regions where the flow field is discontinuous. Thus we are left with a substantial problem: how can we extract the object outlines according to the optical flow discontinuities, if the discontinuities are *a priori* necessary for an accurate estimation of the optical flow[5]?

> *The correct (best possible) estimation of apparent motion of an object requires knowledge of object size and location which requires knowing already the motion.*

Some computational process is needed to solve the problem in which the estimate of the optical flow and the regions-of-support are computed in parallel but within strong mutual recurrence such that their estimates are interactively refined until they converge to some optimal solution. Given the flow model, this process will provide the estimate of the optical flow and thus the parameters of the model as well as the region boundaries for

[5]Note, that for more general flow models, motion discontinuities do not necessarily coincide with the regions-of-support and thus the object boundaries. However, in any case the dilemma remains because the flow estimate defines the regions-of-support and vice versa.

which the particular parameters hold. Such computational interplay naturally reflects the typical integrative and differential interactions found in psychophysical experiments known as *motion capture* and *motion contrast* respectively [Braddick 1993].

From the discussion above it is evident that an accurate estimation of optical flow is inseparable from the process of **motion segmentation**, where the region boundaries represent the segments of the visual scene of common motion sources and the label for each segment is given by the model parameter, thus the optical flow.

2.3 Optimization

The poorly conditioned visual input requires a combination of matching and flow models in order to extract feasible estimates of the optical flow. This is typical for perceptive processes because perception is the interpretation of the often ambiguous input. Interpretation means a classification of the input according to some internal models of what has to be extracted and what is expected. Such classification can be seen as an optimization problem of finding the interpretation of the input that is in best agreement with the model. By defining some appropriate error measure the problem can be formulated in a mathematically sound way. The optimization procedure very much depends on the complexity and the number of free parameters of the models.

An estimation process of visual motion involves optimization on different processing stages and of different complexity. The estimation of the optical flow field, given the visual input and the regions-of-support, usually involves a least-square estimation or a maximum search problem where the computational demands increase polynomially in the size of the regions-of-support N. On the other hand, the selection of the regions-of-support and thus the segmentation process remains a typical combinatorial problem where the number of possible solutions are exponential in N. It has been shown that the problem of motion segmentation is indeed NP-complete [Jagota 1995].

Most approaches of visual motion estimation in computer vision and almost all of their hardware implementations use simple models and/or predefined regions-of-support that do not need expensive optimization to keep the computational load low enough for real-time applications. For example, the estimation of normal flow does not require an optimization process because the region-of-support is only one pixel and the matching model is defined such that it is well-posed[6].

[6] see Appendix A

2.4 Computational Approaches

The following paragraph provides an overview over some selected algorithms and system approaches used for optical flow estimation. Their characteristics and specialties will be discussed. However, optical flow is more of a description than a measure. Thus the motivation for the different approaches is potentially very different. There is nothing like *the* correct optical flow field and therefore a quantitative comparison of any kind is rather delicate (see *e.g.* [Little and Verri 1989, Barron et al. 1994]). All models rely on the same input – the spatiotemporal brightness distribution – to produce a result of common quality, a good estimate of the 2D projection of 3D motion. Hence, it is not surprising that despite having conceptually different routes, some approaches turn out to be computationally the same [Van Santen and Sperling 1985]. An inherently satisfying classification is therefore not possible. However, in agreement with our considerations in Section 2.1, we will distinguish between the following approaches:

2.4.1 Explicit matching methods

Such methods can be thought of performing a matching process either in the temporal or the spatial domain. In computer vision, images are usually captured at fixed time-intervals. This means, matching has to be seen as a spatial search of features in image space. On the other hand, if the brightness information is sensed continuously, the time that features need to travel between two locations of fixed distance in the image may serve as an indirect measure for motion. The matching process itself consists either of a continuous measure, such as a correlation function, followed by a decision mechanism or of an exclusive logical operation applied to the Boolean representation of features.

An early algorithm with explicit matching in the spatial domain is described by Anandan [1989]. The proposed approach uses spatially band-pass filtered image patches. A coarse-to-fine resolution pyramid is applied where initial estimates on coarse levels are fed into the next levels of finer resolution. The matching criterion is the sum of the squared differences (SDD) between the pixel values of a local patch centered at a particular image location in one frame and all patches within a certain search space around the same location in the next frame. The estimate of the local flow vector is given by the displacement that minimizes the SDD. The hierarchical organization introduces a directed matching of the estimate and allows a reduction of the spatial search space on each level to a maximum of one pixel. In addition, it increases the range of detectable displacements. The approach refines its flow estimation such that it introduces a continuous confidence measure that prefers estimates at prominent image features such as edges and corners. Furthermore, to prevent an unreliable estimate in regions of low feature contrast, a translational flow model is introduced that fills in information from high-confidence locations and performs a smoothing of the flow field. The model is formulated such that smoothing is only dom-

inant in regions with low confidence measure and thus ideally, motion discontinuities are preserved. A very similar approach was proposed by Bülthoff et al. [1989]. The authors point out that such correlation methods can account for a variety of psychophysical experimental results and therefore might be a candidate model to explain human motion perception.

Since the timing information is inherently analog and thus of arbitrary resolution, biological systems can restrict spatial interaction to only nearest neighbors. For frame based implementations in computer vision, the range of detectable image velocities is highly restricted by the fixed spatial and temporal quantization. Moreover, it turns out that a linear increase in the extent of the search space of the matching area increases the computational load for cross-correlation quadratically. However, it only scales linearly leaving the spatial window constant and increasing the temporal resolution (frame rate). Such a *time-linear* optical flow estimation scheme, in which the spatial search space is restricted only to nearest neighbors, was proposed by Camus [1994]. However, to obtain a reasonable optical flow resolution the frame rate has to be high and the maximal detectable speed must remain within one pixel per frame.

Explicit matching algorithms ('shift and correlate') are the methods of choice in traditional robotics and industrial applications. The use of spatially extended patches supports such approaches with decent robustness to image noise. Furthermore, the decisive character of a winner-take-all mechanism masks the aperture problem (it does not solve it) and thus keeps it well-posed, except in very pathological cases where the cross-correlation matrix provides several exactly equivalent maxima. However, the computational load is quite high, depending on the required resolution and the expected image velocity.

2.4.2 Gradient methods

Gradient methods represent implicit matching methods. They assume a continuous relation between the changes in the image brightness and the image motion. Clearly, such methods can only account for short-range apparent motion [Braddick 1980].

An early attempt to use brightness gradients to recover image motion was done by Limb and Murphy [1975]. In their approach motion along the x and y axis is treated independently, leading to a simple and computationally well-posed method for the sake of accuracy.

The algorithm by Horn and Schunck [1981] was the first attempt to use the brightness constraint equation (2.3) to estimate a dense 2D optical flow field. In order to obtain a mathematically well-posed formulation, they introduced an additional smoothness constraint that requires the optical flow field to vary smoothly over image space. The *smoothness constraint* is formulated as such that the sum of the squared flow gradients vanishes. Yuille and Grzywacz [1988] introduced a more general formulation of the

smoothing constraints, considering also smoothness in higher order derivatives of the flow field. Since both the brightness constraint as well as the smoothness constraint lead to a simplified description of a general motion scene the algorithm uses the mathematical framework of standard *regularization* where the sum of the squared and weighted error to both constraints is minimized. The relative weight of each constraint is determined by an additional parameter that allows a broad range of computational behavior. The algorithm provides a dense estimate of optical flow. In the case there is a single motion source present within the image space it can solve the aperture problem. Furthermore, it has been shown that first order gradient based methods can be mapped to neuronal correlates in primate visual cortex [Koch et al. 1991]. A disadvantage of the approach is that there exist input conditions for which the computation is ill-posed and robustness is not guaranteed under real-world conditions[7]. In the original formulation of the algorithm, motion discontinuities are not preserved because the smoothing constraint is applied isotropically.

In contrast to isotropic smoothing, Lucas and Kanade [1981] introduced a fixed region-of-support for each image location where translational motion is assumed. The algorithm determines the local flow vector such that it minimizes the sum of the squared and weighted brightness constraints within the local region. The algorithm applies global motion estimation within overlapping regions of fixed size. It can be computed in closed form, except where the aperture problem holds. By introducing a conditioning threshold, motion is only estimated where it is computationally well-conditioned. As a consequence, the estimated optical flow field remains sparse. The closed form optimization allows a relatively fast processing. However, the introduction of a threshold measure requires an external controller that ideally adjusts the threshold according to the spatiotemporal structure of the input.

In the derivation of the brightness constraint equation we neglected the higher order terms in the Taylor approximation (2.2). There are several approaches that take second order terms into account [Verri et al. 1990]. The expansion of second order terms leads to two equations for the two unknown velocity components. Thus, the local flow field should be directly computable. However, this is only possible if the Hessian[8] of the brightness distribution $E(x, y, t)$ is non-singular. Since the computation of second order gradients is very noise sensitive and unreliable in practice, a conditioning threshold for the Hessian must be introduced. Again, this approach leads to a sparse optical flow representation.

Since gradient based methods rely on the continuous and differentiable nature of the spatiotemporal brightness distribution, implementations that allow a continuous sensing of the brightness gradients are necessary. Hence, in frame-based techniques presmoothing with spatial and temporal filters is usually applied.

[7] as will be shown in Chapter 4

[8] see definition in Appendix A

2.4.3 Spatiotemporal filter methods

Models of this class are believed to be most plausible for biological systems. The **correlation based** approaches [Hassenstein and Reichardt 1956, Barlow and Levick 1965, Van Santen and Sperling 1985] apply mainly to motion perception in insects. These approaches did not attract much attention in computer vision because the computational demands of implementing the spatial and temporal filters are large in relation to the quality of the motion estimates obtained.

The general class of **motion energy models** seems to be more interesting because, especially in two spatial dimensions, they account for various perceptive effects. Although first proposed by Adelson and Bergen [1985], the approach is based on work of Watson and Ahumada [1985] who showed that non-separable spatiotemporally oriented filters can be constructed from separable ones. This means that spatially and temporally filtering can be applied sequentially on the image input. Such a scheme seems to be supported by physiological studies in primary visual cortex because it models very well *e.g.* the response of complex cells in V1. To provide a phase insensitive response of the filters, Adelson and Bergen introduced so-called quadrature pairs where the squared responses of two iso-oriented, but 90° phase shifted, spatiotemporal filters are summed. To obtain image velocity, they propose a population encoding scheme where velocity is expressed as the ratio between the responses of motion energy filters tuned to different velocities. Since linearity is assumed, the response of each filter will change with different contrasts but their ratio will remain constant.

Heeger [1987a] and later Grzywacz and Yuille [1990] extended the approach to two-dimensional visual space and proposed a different method to extract local image velocity from motion energy filters. A common local image velocity is supported by the response of motion energy filters with their optimal frequencies lying on a plane in the spatiotemporal frequency domain that intersects the origin. The explicit velocity is given by the slope of the plane with respect to each spatial frequency axis. In image regions where the aperture problem holds the filter responses are supporting several possible velocity planes. Summation of the filter responses along particular velocity planes reveals finally the velocity of common belief, similar to the intersection of constraints solution in gradient based methods.

In contrast to the previous methods, Fleet and Jepson [1990] use the phase relation of the velocity-tuned filters to derive the local motion estimate. In fact, their method resembles a gradient-based approach applied to the phase contours of the output of velocity-tuned filters rather than to the image brightness.

Motion energy models are plausible descriptions of what is happening in visual cortex. Elaborate systems show many of the psychophysical effects and responses to visual illusions known [Nowlan and Sejnowski 1994]. However, for real-time applications in traditional computer vision these approaches are computationally much too expensive.

2.4.4 Motion segmentation

The approaches mentioned so far do not show a segmentation behavior in the sense that they explicitly extract regions of common motion sources. There exist several approaches that do address the problem. A first class of methods proposes a feed-forward procedure where local motion information is extracted first and then segmentation is performed on this information according to some statistical classifier [Schunck 1989, Wu and Kittler 1993] or by applying a probabilistic model using Markov Random Fields [Murray and Buxton 1987]. As expected, they perform well for highly textured objects but fail for scenes containing larger and mildly textured objects.

A second class of algorithms combine static brightness cues with visual motion information to improve segmentation performance in regions of low contrast. An early approach by Thompson [1980] proposes a two-step region-merging process between local motion information and gray level intensities of the image scene. Local motion information is derived using a gradient method proposed by Fennema and Thompson [1979]. Unfortunately, a quantitative description of the method is missing, so that a clear understanding is impossible. Another recent attempt was proposed by Cesmeli and Wang [2000] where the brightness segmentation is not only combined with the local motion estimate but recurrently refines this estimate. Their presented results consist mainly of artificial and real image sequences with objects of high contrasts and suggest that the brightness differences play a dominant role in the segmentation process.

The third class covers approaches that estimate optical flow and perform motion segmentation simultaneously. To separate regions of common motion sources, several authors use the concept of **line processes**, first introduced by Geman and Geman [1984], adding a variable that controls the integration among neighboring image locations. Line processes have been suggested for many early visual tasks that need smooth spatial interaction while preserving discontinuities (see *e.g.* [Marroquin 1985, Terzopolous 1986, Marroquin et al. 1987, Gamble and Poggio 1987, Murray and Buxton 1987]). Typically, line processes require non-deterministic or annealing methods for optimization. Many of the stochastic approaches use a Bayes formalism to describe the probability distribution in Markov Random Fields [Chang et al. 1997]. When searching for the *maximum a posteriori* probability of a solution, this approach reduces to the minimization of a cost function. To avoid stochastic optimization, Memin and Perez [1998] apply an annealing approach where the optical flow estimation is refined during a multi-grid procedure. In addition, they expand their cost function by a term that forces closed contours. In general, such multi-scale approaches show good performance [Stiller and Konrad 1999].

Of special interest are approaches using analog electronic networks to perform combined estimation of optical flow and segmentation [Hutchinson et al. 1988, Koch et al. 1989]. Such networks undergo a simple gradient descent behavior to find near-optimal solutions. Although simulation results showed promising results there is no known attempt

of a physical implementation yet.

There is a common property to all deterministic approaches: Since the cost function of the optimization problem typically exhibits many local minima, the system has to be reset to some initial state in order to process new visual input. Otherwise, multi-stability occurs such that a once chosen segmentation will remain even when the visual scene changes.

2.5 Neuromorphic Implementations

Real-time performance of visual motion systems can hardly be achieved without any dedicated hardware solutions. The tremendous computational efficiency of biological systems encouraged scientists and engineers very early to build electronic circuits that emulate the principles of such systems. Historically, the first attempts were to implement the *outer plexiform layer (OPL)* of a typical mammalian retina. As early as in the 70's, [Fukushima et al. 1970] presented such a model using discrete (!) electronic elements. As access to **analog Very Large Scale Integrated (aVLSI)** technology became standardized and affordable, Carver Mead began to create computationally efficient and powerful electronic devices using the insights neurobiologists provided from biological neural computation. Since then many of these so-called **neuromorphic** [Mead 1990] or **neuro-inspired** [Vittoz 1989] systems for solving perceptive as well as purely computational tasks[9] were engineered.

Several approaches apply **explicit matching in the time-domain**: They compute the time-of-travel for a feature passing from one detector to its neighbor. In [Sarpeshkar et al. 1993] and later more elaborated [Kramer et al. 1997], the authors propose two circuits in which the matching features are temporal edges. In one of their circuits (facilitate-and-sample) a temporal brightness edge triggers a pulse that logarithmically decays in time until a second pulse occurs that indicates the arrival of the edge at the neighboring cell. The decayed voltage is sampled and represents the logarithmically encoded local image velocity. The reported results for 1D arrays exhibit a relatively accurate velocity estimation over many orders of magnitude. In the second scheme, the temporal edge elicits a pulse of fixed amplitude and length at the measuring detector as well as at its neighbor. The overlapping pulse width is an inversely linear measure for image velocity. In this scheme, since encoding is linear, the usable range of operation is limited. Either small velocities are not detectable because the pulses become non-overlapping or the resolution for high velocities decreases. An additional variation [Kramer 1996] encodes image velocity as inversely proportional to the length of a binary pulse. Such local motion circuits were also used successfully for the localization of the focus-of-expansion [Indiveri et al. 1996, Higgins and Koch 1999] or the detection of 1D motion discontinuities [Kramer

[9] However, as stated earlier, perception is always the result of a computational process.

et al. 1996]. Other similar implementations based on the time-of-travel of temporal edges are reported [Higgins and Koch 1997].

Benson and Delbruck [1992] present a direction selective retina circuit that is inspired by the mechanism for directional selectivity found in the rabbit retina [Barlow and Levick 1965]. It also represents a explicit matching method in the time-domain and is similar to the approach of Kramer [1996]. A temporal edge detected at a local motion cell elicit a pulse which is shut down by inhibitory input from the neighboring motion cells in the preferred direction. Thus the length of the output pulse is inversely proportional to the speed. In the null direction, inhibition suppresses the output of the neighboring detector. The reported chip consists of an array of 41x47 direction selective cells, all having the same directional tuning.

In [Horiuchi et al. 1991], the velocity estimation task is transformed into a spatial correlation task which is inspired by the process of auditory localization in the barn owl: The occurrence of a feature at a detector and its neighbor triggers two pulses traveling along oppositely directed delay lines. Correlators watching these lines detect where pulses pass each other. Active correlators detected by a winner-take-all circuit serve as the spatial correlate for the observed image velocity. Slow speeds in either direction will be detected at the end of the correlator array whereas fast speed lead to winners close to the middle.

Another aVLSI implementation of explicit matching in the time domain uses spatial edges as matching features [Etienne-Cummings et al. 1993]. Spatial edge detection is achieved by approximating a difference-of-Gaussian operation. Local brightness values are compared to the smoothed image provided by a resistive mesh using well implants or polysilicon lines as resistive elements [Etienne-Cummings 1993]. The presence and the location of spatial edges are coded as binary signals that are obtained by thresholding the edge enhanced image. Velocity is measured as the time-of-travel for the binary edge signals.

Moini et al. [1997] report a 1D motion chip called 'bugeye' that emulates the typical insect visual motion detection system. The implementation simplifies the biological system insofar that it quantizes the temporal changes in the visual input. It then uses these quantized temporal changes as labeled events to apply template matching in the time-domain in order to extract visual motion information. Matching is performed off-chip to allow to test different template models of different complexity levels. The chip therefore represents only the front end of a visual motion system. Special attention was paid to a multiplicative noise cancellation mechanism that reduces the sensitivity of the chip to 120 Hz flicker noise.

One of the few examples of a successful commercial application of integrated motion-sensing circuits is Logitech's Trackman Marble[10], a trackball pointing device for personal

[10] www.logitech.com

computers [Arreguit et al. 1996]. So far, several million units were sold world-wide. The core of the trackball consists of an 2D optical sensor-array of 75 motion processing units which uses spatiotemporal sampling to extract the global motion of the trackball. The implementation is insofar different from previous examples because it uses precisely controlled and pulsed lighting conditions as well as a given random-dot pattern stimulus that is directly imprinted onto the trackball itself. Obviously, these conditions simplify the motion extraction task significantly. Each pixel consists of a simple edge-detector that compares the differences in sensed intensity between neighboring photoreceptors to a given threshold. The global motion of the trackball is then directly computed by dividing the number of edges that moved between two time-steps in each of two orthogonal directions by the total number of edges present. The higher the total number of edges seen by the sensor is, the higher is the resolution of the motion estimate. Using a particular random dot pattern, the spatial average of the number of edges can be well-controlled and kept rather constant. The temporal sampling has to be fast enough such that the maximal on-chip displacement is guaranteed to be below one pixel-spacing, otherwise temporal aliasing will occur.

In general, the circuits presented above are compact, show a robust behavior and work over a wide range of image motion. However, there is a fundamental disadvantage related to approaches relying on the time-of-travel of features: Such implementations can only report local image motion *after* the moving feature has passed the local detector, thus being undetermined in situations where no motion is present. This means, these circuits cannot detect zero motion per definition. For example, if a moving object stops abruptly, this stop cannot be detected.

There are several implementations that use a **gradient based** motion estimation scheme. One of the earliest integrated motion chip was proposed by Tanner and Mead [1986]. The 8x8 motion array consists of identical cells that compute a common global motion estimate in a cooperative/competitive manner. The two components of the global motion vector are commonly provided to each cell as two voltages on global wires. Each cell tries to adjust this global estimate according to its *locally* measured image gradients such that the sum of the squared errors of the constraint equations (Equation 2.3) is minimal for the whole array. Thus, the chip implements the computational approach proposed by Lucas and Kanade [1981] with uniform weighting and a window size equivalent to the whole imaging array. In principle, the chip is able to solve the aperture problem for singular motion and thus to compute real 2D motion. However, as widely witnessed[11] the actual chip never worked robustly under real-world conditions. The feedback computation of the local error turned out to be difficult to implement correctly and sensitive to mismatch and noise in the gradient computation. A second attempt to improve the implementation also failed [Moore and Koch 1991].

[11] personal communication, Delbrück, T. and Harris, J. and Koch, C.

A gradient-based 1D implementation has been reported by Deutschmann and Koch [1998a]. In one spatial dimension, image motion can directly be computed as $v = -\partial E_t / \partial E_x$. To avoid the division by zero problem, a small constant current is added to the denominator such that it cannot become zero in any case. The circuit shows robustness and appropriate linearity. The use of wide linear-range circuits allows to span two orders of magnitude of velocity range. Compact gradient-based motion sensors were also presented [Moore and Koch 1991, Horiuchi et al. 1994] and later in two spatial dimensions [Deutschmann and Koch 1998b] where the resulting motion signal is basically the product of the spatial and temporal brightness gradients. Such a motion signal obviously strongly depends on the contrast and spatial frequency of the visual scene. Nevertheless, these compact implementations provide at least the correct direction of motion. This has been proven to be sufficient for visual-motor tasks such as saccadic and smooth pursuit tracking [Horiuchi et al. 1994].

In a third category we consider implementations that are very closely modeling **correlation based methods** observed in biological systems. There is a series of implementations based on the Reichardt scheme. An early representative [Andreou and Strohbehn 1990] consists of a 1D array of motion detectors where multiplication between the spatially filtered visual input and its delayed version of a neighboring detector cell serves as the non-linear correlation function. The correlation signals from local detectors are averaged to provide a global motion estimate. The first 2D implementation uses uni-directional delay lines as temporally tuned filters for moving edges [Delbruck 1993c]. The delay lines are oriented in three spatial directions within the hexagonal array of correlator cells. Photoreceptors sense changes in the intensity and couple their output into the delay lines where they propagate with a characteristic velocity in one direction. Whenever the velocity of a moving intensity edge matches the propagation speed along a particular delay line, then the signal on the delay line is reinforced. The circuit performs temporal integration of the motion signal along the orientations of the delay lines. Although the chip contains a 2D array of motion cells, local motion information is blurred by means of the temporal integration. Nevertheless, this approach utilizes collective computation since the motion estimate is the result of a spatiotemporal integration process along the elements of each delay line.

Recently, the fly visual motion system attracted some attention from neuromorphic engineers. One-dimensional silicon models of the fly's elementary motion detector have been presented that show similar tuning characteristics as physiological data from the insects [Harrison and Koch 1998, Liu 2000]. The estimate for visual motion is computed by multiplying the temporal high-pass filtered signal of the photoreceptor of a motion detector with the low-pass filtered signal of its neighbor. The fixed time constants of the low-pass and high-pass filter make the detector narrowly tuned to particular spatial and temporal frequencies of the input stimulus. Liu [2000] proposes a silicon model of

the spatial aggregation of the local motion responses observed in the fly which leads to a global motion output that is relatively insensitive to the contrast and size of the visual stimulus. Adaptation of the time constants of the low- and high-pass filters is also known to take place in the fly's visual motion system such that the sensitivity to changes in the perceived motion remains high. Such adaptation mechanisms have been implemented and show successful replication of physiological data [Liu 1996].

So far, the discussed correlation-based implementations are exclusively reporting visual motion along one spatial dimension. They report visual motion but not explicitly velocity. Motion signals from a correlation detector are inherently dependent on contrast, spatial and temporal frequency. Correlation based methods have been shown to be equivalent to **motion energy based** approaches. The estimation of two-dimensional motion requires large ensembles of spatiotemporally oriented filters. Furthermore, the extraction of apparent image velocity requires additional processes such as integration [Heeger 1987b] or normalization [Adelson and Bergen 1985]. However, there have been attempts to realize a motion energy based system using a special purpose hybrid hardware system [Etienne-Cummings et al. 1996, Etienne-Cummings et al. 1999]. Although the system only approximates two-dimensional spatial filtering, a single motion unit with filters of combinations of only three spatial and three temporal scales requires several hundred neuronal units with thousands of synapses. Obviously wiring and density constraints do not allow such a system to be implemented on a single or few-chip system.

Little is known about implementing aVLSI **motion segmentation** systems. Although several implementations have been reported that successfully demonstrate segmentation in one-dimensional feature space using so-called *resistive fuse circuits* [Harris and Koch 1989, Harris 1991, Liu and Harris 1992], there exists only one attempt by Kramer et al. [1996] to achieve segmentation by motion. Their implementation is based on the one-dimensional array of local motion elements described in Kramer et al. [1995], thus keeping the feature one-dimensional as well. A line process is performed where bump circuits [Delbruck 1993a] compare the output of two adjacent motion elements and control the conductance in a diffusion network accordingly. In regions of common motion, the output of the individual elements is averaged while regions of different motion are separated.

The majority of the discussed aVLSI motion systems (see Table 2.1) report visual motion only along single spatial dimensions thus avoiding the aperture problem and therefore complex integration schemes. Some of these circuits can be very compactly integrated. However, to estimate visual motion in two spatial dimensions, information about the two-dimensional structure of the image scene has to be available. Gradient based methods have the advantage that the spatial gradients are relatively easy to compute and are a compact representation of spatial orientation. Explicit matching methods would require a feature extraction stage that accounts for two-dimensional features. Implementations for the extraction of brightness corners have been reported [Pesavento and Koch 1999] but

Authors	motion resolution	motion dimension	explicit velocity	normal flow	object motion
Optical flow estimation					
feature matching					
[Horiuchi et al. 1991]	local	1D	yes	no	no
[Etienne-Cummings et al. 1993]	local	1D	yes	no	no
[Sarpeshkar et al. 1993]	local	1D	no	no	no
[Arreguit et al. 1996]	global	2D	yes	no	yes
[Kramer 1996]	local	1D	yes	no	no
[Higgins and Koch 1997]	local	1D	yes	no	no
[Moini et al. 1997]	local	1D	yes	no	no
gradient based					
[Tanner and Mead 1986]	global	2D	yes	yes	yes
[Moore and Koch 1991]	local	1D	no	no	no
[Horiuchi et al. 1994]	local	1D	no	no	no
[Deutschmann and Koch 1998a]	local	1D	yes	no	no
[Deutschmann and Koch 1998b]	local	2D	no	yes	no
correlation based					
[Andreou et al. 1991]	global	1D	no	no	no
[Benson and Delbruck 1992]	local	1D	no	no	no
[Delbruck 1993c]	global/local	3 x 1D	no	no	no
[Harrison and Koch 1998]	local	1D	no	no	no
[Liu 2000]	global	1D	no	no	no
motion energy based					
[Etienne-Cummings et al. 1999]	local	2 x 1D	yes	no	no
Motion segmentation					
[Kramer et al. 1996]	local	1D	yes	no	no

Table 2.1: Overview of existing neuromorphic aVLSI motion systems.

they require rather expanded circuitry. Motion energy based systems are probably the method-of-choice but turn out to be expensive to implement even in a simplified manner.

Chapter 3

Constraint Optimization

The extraction of optical flow and furthermore motion segmentation is substantially based on optimization processes that find the optimal solution with respect to the applied models and the visual data. According to the complexity of the models the level of optimization varies and different computational optimization approaches might be advantageous. The framework of **constraint optimization**[1] has some particular properties that makes it the method of choice for the purpose of this thesis. Constraint optimization requires a formalism where the optimization task is formulated (modeled) as different constraints that are imposed on the solution. These constraints might be partially contradicting. Mathematically, they are formulated and combined such that they form a **cost function** or energy, that has a minimum for the optimal solution. The influence of each constraint is thereby controlled by a weighting parameter. Since each constraint reflects partial properties of the applied model, the weighting parameter provides a means for continuous adaptation of the model. These parameters can be the output from another 'top-down' process, so providing the necessary dimensional reduction used in the control structure of complex systems.

The reduction of the optimization problem to the task of minimizing a cost function also allows the use of deterministic methods to find the (local) minima. Furthermore, applied to image processing or other problems with non-hierarchical, topographically arranged input space, such optimization tasks can be mapped to cellular network arrays of identical units that behave continuously in time according to some deterministic dynamics. Such parallel processing in locally connected analog networks exactly reveals the superiority of biological systems in solving perceptive tasks. And, it also matches the characteristics of neuromorphic aVLSI implementations.

This chapter gives an introduction to constraint optimization and how such opti-

[1] Not to be confounded with *constrained* optimization. Constraint optimization is actually an *unconstrained* optimization method because it implies no explicit additional constraints on the solution other than minimizing the cost function [Fletcher 1980, Fletcher 1981].

mization processes can be translated to the dynamics of simple processing units in non-hierarchical analog networks. Two simple examples are discussed that each represent an integrative and selective system, described in the framework of constraint optimization.

3.1 A Physical Description of Computation in Networks

Since the very first attempt by McCulloch and Pitts [1943] many different ways of describing the computational properties of neural networks were proposed. Among others, the concept of **content-addressable memory**, thus how neural networks can store information patterns and retrieve them from noisy or only partially present input, had a significant impact on the research community, starting in the early 70's. Although there were substantial contributions before [Little and Shaw 1975, Grossberg 1978], it was Hopfield [1982, Hopfield [1984] that clearly established the notion that memory can be represented by the locally stable states of a dynamical system consisting of a number of equivalent simple units. He showed that the stable states represent minima of an appropriate cost function and depend on the interaction strength between the single units. Furthermore by assigning such a cost function he pointed out the physical nature of computation.

Soon, it was discovered that the restoring properties of such a content-addressable memory could not only be useful to recall pre-learned patterns but also to find (local) solutions of problems that represent minima of a cost function, thus optimization. Hopfield and Tank [1985] demonstrated how the computationally hard *traveling salesman problem (TSP)* can be formulated as a combined cost function of several constraints and how a near-optimal solution can be found by a network of the appropriate architecture within a few characteristic time constants. Although this particular problem turned out to be not very well suited for such a network approach [Wilson and Pawley 1988, Kamgar-Parsi and Kamgar-Parsi 1990, Gee and Prager 1995] it clearly demonstrated its potential for solving optimization problems. Meanwhile, a variety of optimization problems including linear and quadratic programming [Tank and Hopfield 1986], and their generic network solutions have been addressed [Cichocki and Unbehauen 1993, Liang and Wang 2000].

3.2 Competitive and Cooperative Behavior of Constraint Optimization Networks

A successful application of constraint optimization for problem solving depends primarily on how well the problem is defined by the chosen constraints. Many computational problems can be formulated as constraint optimization problems; although those are more

suitable where a valid solution does not require all of the constraints to be exactly fulfilled in order to be valid.[2] In the following, two example problems are discussed.

3.2.1 Winner-take-all networks

Consider the typical winner-take-all (WTA) problem: Given a discrete set of positive input values $I_1, \ldots, I_N \subset \mathbb{R}^+$, find I_{max} such that

$$I_{max} \geq I_j \; \forall \; j \neq \max \; !$$

The problem can be reformulated as finding a network architecture of computational units that assigns to each given input vector $I_1, \ldots, I_N$ a binary output vector $V_1, \ldots, V_N$ where $V_{max} = 1$ and $V_{j \neq max} = 0$.

Once the problem is identified, we have to find suitable constraints and their appropriate measures to define a cost function. Below is a possible suggestion of such a cost function:

$$H(\boldsymbol{V}) = \underbrace{\frac{\alpha}{2} \sum_i \sum_{j \neq i} V_i V_j}_{\text{sparse activity}} + \underbrace{\frac{\beta}{2} (\sum_i V_i - 1)^2}_{\text{limited total activity}} - \underbrace{\gamma \sum_i V_i I_i}_{\text{biggest input wins}} \tag{3.1}$$

The cost function $H(\boldsymbol{V})$ is a combination of three constraints where their relative strengths are given by the weighting parameters α, β and γ. The first and the second constraint promote all the states where only one output unit is active and its activity level $V_{max} = 1$. We can consider these as *syntactic* constraints that ensure that the network performs a decision in any case whether it is the right one or not. The third constraint finally relates the data to the output such that it will be most negative if the winning unit is actually the one which receives largest input. For finite input values, the cost function is bounded from above and below where the lower bound is equal to $H_{min} = -\gamma I_{max}$. The WTA problem is now formulated as the optimization problem:

Given $I_1, \ldots, I_N$, find the output vector $V_1, \ldots, V_N$ such that $H(\boldsymbol{V})$ is minimal.

Once a cost function is assigned we want to derive a **local update rule** that we can repeatedly apply to each computational unit in the network such that from a given start condition the system ends up in a state of minimal costs. In the following, we define a simple rule such that we change the output state V_i of each unit if this lowers the total

[2] This was exactly the reason why the TSP could not be solved satisfactorily by the proposed constraint optimization approach. The syntactic constraints (every city once, only one city at a time) had to be perfectly matched in order to obtain a valid tour. With increasing number of cities, the ratio of valid to the total number of found solutions dropped to a inefficiently low value.

costs. The change in costs that a transition $V_i \rightarrow V_i^{'}$ of a single unit's output induces can be written as

$$\Delta H_{V_i \rightarrow V_i^{'}} = H(\boldsymbol{V'}) - H(\boldsymbol{V}). \tag{3.2}$$

Using a finite difference approximation we can compute $\Delta H_{V_i \rightarrow V_i^{'}}$ from Equation (3.1) as

$$\Delta H_{V_i \rightarrow V_i^{'}} = \Delta V_i \ (\alpha \sum_{j \neq i} V_j + \beta(\sum_i V_i - 1) - \gamma I_i). \tag{3.3}$$

Now we can formulate the update rule in terms of positive ($\Delta V_i > 0$) and negative ($\Delta V_i < 0$) transitions:

$$V_i^{'} \Rightarrow \begin{cases} \textit{apply transition} & \text{if } \ \Delta H_{V_i \rightarrow V_i^{'}} < 0 \\ \textit{no transition} & \text{otherwise} \end{cases} \tag{3.4}$$

A WTA network of units with such dynamics cannot serve as a model for an equivalent physical network because it does not consider any thermal noise to be present. A stochastic formulation of the dynamics with a finite temperature would be more appropriate but requires methods of statistical mechanics to describe the network behavior and find valid solutions. Rather then going into a detailed analysis of stochastic networks[3], we want to modify the model we have so far such that the output units are now allowed to take on **continuous values**. The characteristics of a single unit is thus described by a continuous activation function $g(u) : u \rightarrow V$, where u represents the input to unit and V its output. Figure 3.1 shows some examples of such non-linear activation functions.

A typical sigmoidal activation function is

$$g(u) = \frac{1}{2}(\tanh(u/u_0) + 1) \tag{3.5}$$

where the parameter $1/u_0$ characterizes the maximal gain of the unit's activation function. In the limiting case where $1/u_0 \rightarrow \infty$ we have the two-state unit as before. In general, the activation function does not necessarily have to be of sigmoidal nature. In the case of the WTA network we require only that the function is differentiable, monotonically increasing, and limited from below. In particular, the output range of the units does not have to be limited to the unity interval. Thus, linear-threshold units (see Figure 3.1a) can be also used and can be treated within the same analytical framework. The only relevant parameter is the activation gain. We see later, that for other optimization problems the units can also have purely linear activation functions.

The constraints imposed on the WTA system remain the same in the case of continuous units, although the output of the winning unit is not guaranteed to be exactly one. Also the cost function (3.1) is the same; except that it must now include an additional term

[3] which can be found in textbooks such as *e.g.* in [Hertz et al. 1991]

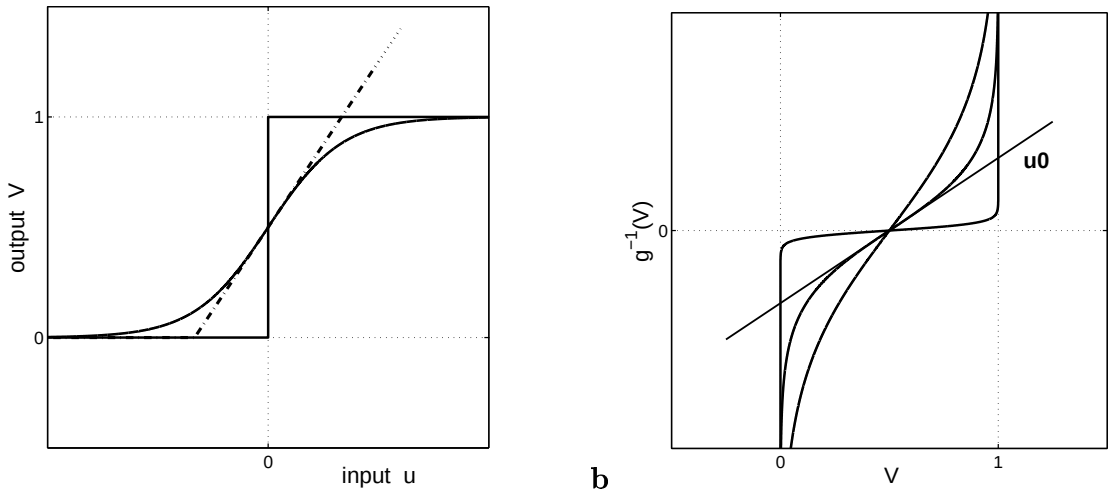


Figure 3.1: *Non-linear activation functions.* (a) A sigmoidal, a hard threshold and a linear threshold activation function (dash-dotted line) are shown. (b) The inverse of typical sigmoidal activation functions (3.5) with different slope parameters u_0 are plotted. The total integral under each curve represents the activation energy needed to activate the unit and decreases with decreasing u_0.

which represents the total **activation energy** needed to keep the units in states of either high or low activity. We can consider the activation energy as an additional constraint imposed by the computational machinery rather than by the problem itself. The cost function (3.1) hereby modifies to

$$\tilde{H}(\boldsymbol{V}) = \frac{\alpha}{2}\sum_i \sum_{j\neq i} V_i V_j + \frac{\beta}{2}(\sum_i V_i - 1)^2) - \gamma \sum_i V_i I_i + 1/R \sum_i \int_{1/2}^{V_i} g^{-1}(\xi)\, \mathrm{d}\xi \qquad (3.6)$$

where $1/R$ is the additional weighting parameter. We see later that in the context of an equivalent electronic network implementation the parameter $1/R$ can be interpreted as the non-zero leak conductance of each unit. As Figure 3.1b illustrates, the integral over the inverse activation functions and thus the activation energy decreases for an increasing activation gain $1/u_0$. Thus in the high-gain limit ($u_0 \to 0$) the cost function (3.6) approaches the initially proposed form (3.1).

$\tilde{H}(\boldsymbol{V})$ is sufficiently regular and at least twice continuously differentiable because $\boldsymbol{V}$ is continuous. We propose a new update strategy for the activity changes of the network to lower the global cost function. A simple **gradient descent** rule is applied where the output V_i of each unit changes proportionally to the negative partial gradient of the cost function, thus

$$\dot{V}_i \propto \frac{\partial \tilde{H}}{\partial V_i} \qquad (3.7)$$

until a steady state of minimal cost is reached where $\partial \tilde{H}/\partial V_i = 0 \;\forall\; i$.

Since V_i is monotonic in u_i, we define the following equations of motion:

$$\begin{aligned} \dot{u}_i &= -\frac{1}{C}\frac{\partial \tilde{H}}{\partial V_i} \\ \text{and expanded} \qquad \dot{u}_i &= -\frac{1}{C}\left[\frac{u_i}{R} + \alpha \sum_{j \neq i} V_j + \beta(\sum_i V_i - 1) - \gamma I_i\right] \end{aligned} \tag{3.8}$$

The constant C determines the time constant of the units. Its physical interpretation becomes obvious if we compare the dynamics (3.8) with a simplified electrical circuit as sketched in Figure 3.2: The dynamics exactly describe the current equilibrium at the summation node of the circuit where α, β and γ are conductances. The input and output values are represented by voltages. A typical sigmoidal activation function can be implemented by an transconductance amplifier [Hopfield 1984], but also other functions are possible.

We still have to test if the proposed dynamics actually guarantee a stable system such that it always converges to some asymptotically stable fixed point. To prove this, we show that under arbitrary starting conditions, the cost function (3.6) always decreases [Hopfield 1984]. Differentiating $\tilde{H}(\boldsymbol{V})$ with respect to time leads to

$$\begin{aligned} \frac{\mathrm{d}\tilde{H}}{\mathrm{d}t} &= \sum_i \frac{\partial \tilde{H}}{\partial V_i}\frac{\mathrm{d}V_i}{\mathrm{d}t} \\ &= \alpha \sum_i \sum_{j \neq i} V_j \frac{\mathrm{d}V_i}{\mathrm{d}t} + \beta(\sum_i V_i - 1)\sum_i \frac{\mathrm{d}V_i}{\mathrm{d}t} - \gamma \sum_i \frac{\mathrm{d}V_i}{\mathrm{d}t} I_i + 1/R \sum_i g^{-1}(V_i)\frac{\mathrm{d}V_i}{\mathrm{d}t} \\ &= \sum_i \frac{\mathrm{d}V_i}{\mathrm{d}t}\left[\sum_j (\alpha + \beta)V_j - \alpha V_i - \beta - \gamma I_i + \frac{u_i}{R}\right] \\ &= -\frac{1}{C}\sum_i \frac{\mathrm{d}V_i}{\mathrm{d}t}\frac{\mathrm{d}u_i}{\mathrm{d}t} \qquad \text{(substitution with (3.8))} \\ &= -\frac{1}{C}\sum_i \left(\frac{\mathrm{d}u_i}{\mathrm{d}t}\right)^2 g'(u_i) \leq 0 \; . \end{aligned} \tag{3.9}$$

Thus, $\tilde{H}(\boldsymbol{V})$ is never increasing and because it is bounded from below and has the dynamics (3.8), it will always converge to a asymptotically stable fixed point. Therefore, $\tilde{H}(\boldsymbol{V})$ represents a *Lyapunov function* of the system. As we see, this holds for many all differentiable activation functions that are monotonically increasing, *i.e.* $g'(u) \geq 0$.

WTA network architecture

So far, we proposed a network containing N simple computational units with a particular, non-linear activation function. We formulated the winner-take-all task as an constraint optimization problem and defined the dynamics of the units in order to guarantee that

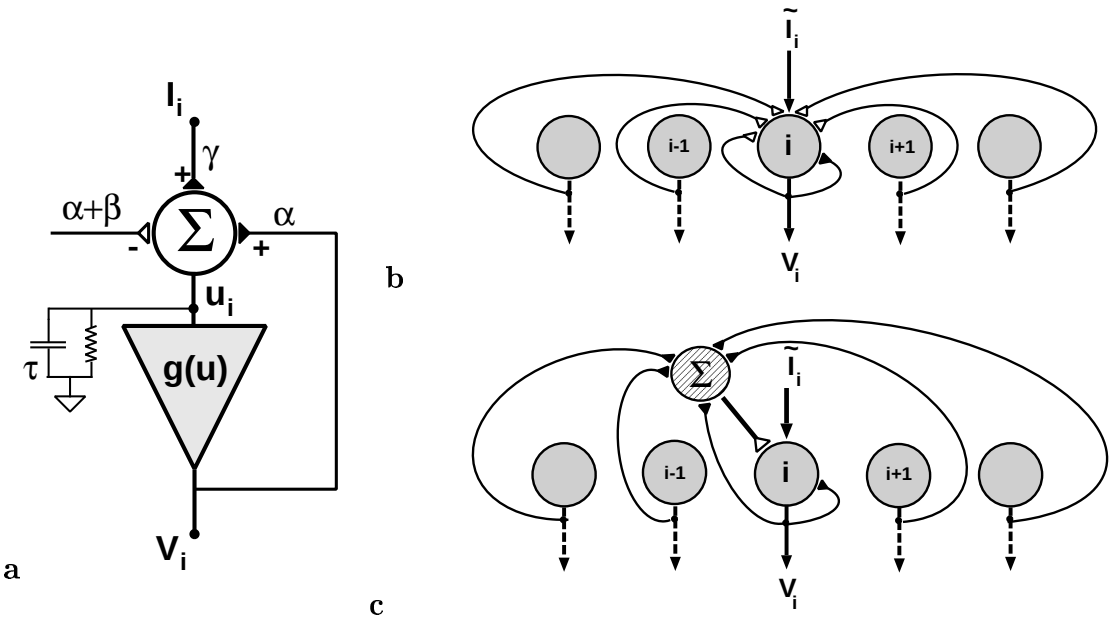


Figure 3.2: *WTA network architectures.* (a) Single unit of the WTA network. Excitatory connections are shown as filled triangle and inhibitory ones as open triangle. (b) WTA network with homogeneous structure (only inhibitory connections to unit i are shown). (c) WTA network with global inhibitory unit.

the system has asymptotically stable states representing the wanted solutions. However we did not consider yet what the structure of such a network will be. If we add a constant term to the external input ($\tilde{I}_i = I_i + \beta/\gamma$) and include the capacitance C in the weight parameters α, β and γ, we can rewrite (3.8) in the following way:

$$\dot{u}_i = -\frac{u_i}{\tau} + \underbrace{\alpha V_i}_{\text{self-excitation}} - \underbrace{(\alpha+\beta)\sum_i V_i}_{\text{global inhibition}} + \gamma \tilde{I}_i. \tag{3.10}$$

The dynamics (3.10) describe a typical WTA network that exhibits self-excitation and global inhibition [Grossberg 1978].

There are basically two interpretations for the network architecture: It can be thought of as a network of homogeneous units where each unit has inhibitory connections to all the other units and one excitatory connection to itself. Alternatively, the network has a single additional inhibitory unit[4] that sums the equally weighted output from all N units and feeds it back to each unit as a global inhibitory input. Both cases are illustrated in Figure 3.2. The additional unit and the loss of homogeneity are compensated by the massive reduction of connectivity. Applying a global inhibitory inter-neuron reduces the

[4] or inter-neuron, see [Kaski and Kohonen 1994]

connectivity to the fraction $3/(N+1)$ of the fully connected architecture, a substantial reduction to a few percent for networks consisting in the order of hundreds of units. Nevertheless, to be computationally equivalent we require the time constant of the inhibitory inter-neuron to be small enough such that asymptotic stability holds and thus limit cycles do not occur. In addition, we have to assume the activation function of the inhibitory unit as being strictly linear. Other possible characteristics like a sigmoidal function, would alter the cost function (3.6) such that α and β become dependent on the absolute level of activity. However, the qualitative behavior of the network would remain the same.

Global convergence, gain and multi-stage WTA amplifiers

Although we showed that the system always converges to an asymptotically stable state we did not consider whether it is globally convergent, *i.e.* always converges to the global minimum or not. In fact, the applied gradient descent method always gets stuck in local minima and there is no deterministic method to avoid so. The only way to ensure global convergence is to make sure that there are no local minima. In other words, under all conditions for which $\tilde{H}$ is shown to be **convex**[5], there is a global minimum and therefore the system globally converges.

However, before going deeper into this, we first have to decide if global convergence is actually required. Recalling our task, it is sensible to require the system always to recognize the winner independently of its starting conditions or, when changing the input distribution, of its previous state. Otherwise if local minima are present, a **hysteretic** or multi-stable behavior might occur and the unit winning the competition does not necessarily have to be the 'real' winner. Nevertheless, such a behavior can be of interest under some computational aspects. In any case, it is interesting to know what the conditions for hysteresis are.

We notice that convexity is a too strong but easily testable criterion for a global minimum. There are obviously non-convex functions that show the same property. The convexity of $\tilde{H}$ is guaranteed if its *Hessian* $\mathsf{J}_{\tilde{H}}$ is positive semi-definite[5]. We find $\mathsf{J}_{\tilde{H}}$ to be symmetric and only dependent on the weighting parameters. In order to be positive semi-definite its eigenvalues have to be real and non-negative. There are two distinct eigenvalues for $\mathsf{J}_{\tilde{H}}$

$$\lambda_1 = \beta + \frac{u_0}{R} + (N-1)(\alpha+\beta) \quad \text{and} \quad \lambda_2 = -\alpha + \frac{u_0}{R} \; . \tag{3.11}$$

Since we require the weighting parameters α, β, γ and R to be positive, λ_1 is always positive whereas λ_2 is only non-negative for

$$\alpha \leq \frac{u_0}{R}. \qquad \text{self-excitation gain limit} \tag{3.12}$$

[5] see Appendix A

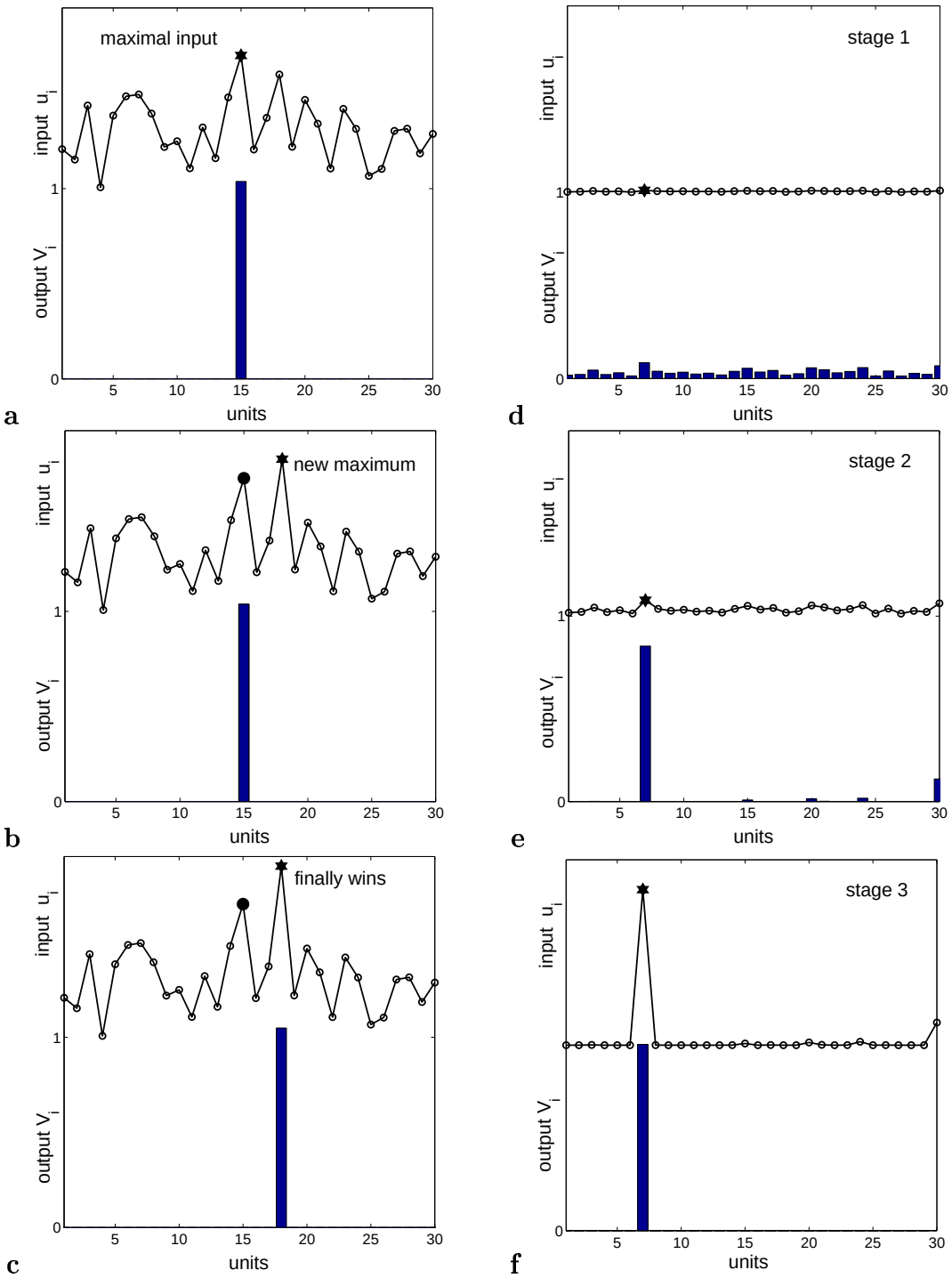


Figure 3.3: *Hysteresis.* (a,b,c): A winner is selected and kept although the input changes and actually a new unit receives the largest input. It needs a substantially larger input to overcome the local minimum and switch to the 'correct' winner. *Multi-stage WTA amplifier.* (d,e,f): For small input differences ($< 1\%$) obeying the self-excitation gain limit (3.12), the gain is too small to clearly select a winner. Successive stages multiply the gain and lead to a clear winner solution without hysteretic behavior.

For a given leak conductance $1/R$, there is obviously a trade-off between strong self-excitation (α) and a high gain activation function ($1/u_0$). In other words, in order to approximate a 'pure' implementation of our constraint problem (3.1) one has to accept a possible hysteretic behavior of the network. Figures 3.3a-c show the effect of hysteresis for a WTA network where the self-excitation gain limit does not hold. Only the winning unit is active. Once the winning unit is assigned, it remains in its winning position although the input changes and some other unit now receives the largest input. It needs a substantial input difference to overcome the local minimum and let the network switch to the 'real' winning input.

On the other hand, if the self-excitation gain limit is obeyed, the network might not be able to assign a clear winner, which is especially true for small input differences. In this case, a cascade of WTA networks can overcome these limitations: The output of one network serves as the input to the next network in the cascade (see Figure 3.3 d-f). Such a **multi-stage WTA** system can multiply the gain almost arbitrarily, and provides a hard WTA behavior while avoiding hysteresis.

3.2.2 Resistive networks for smooth data interpolation

Above, we have shown how a problem can be formulated in the framework of constraint optimization and how an appropriate network architecture can be constructed such that it solves the problem. In the present example, we would like to proceed the opposite direction: We consider a network of simple computational units which resembles a well-known electrical structure and we want to investigate what task it solves and how it can be formulated as a constraint optimization problem.

Consider the resistive (or diffusion) network shown in Figure 3.4. The voltage distribution U is the input to the network. The second voltage distribution V represents the output of the network. Each input node is connected to the output by some input conductance λ_1 and each output node is connected to its nearest neighbors by the lateral conductance λ_2. Applying Kirchhoff's current law to the node V_{ij} we find the following dynamics for a single node in the network:

$$\dot{V}_{ij} = \frac{1}{C}[\lambda_1(U_{ij} - V_{ij}) + \lambda_2(V_{i,j+1} + V_{i,j-1} + V_{i+1,j} + V_{i-1,j} - 4V_{ij})] \tag{3.13}$$

From the preceding example we know that if the system is asymptotically stable these dynamics can be interpreted as performing a gradient descent on a global cost function. If we consider the right-hand side of (3.13) as the negative discrete partial derivative of such a cost function, we simply have to find ways to partially integrate it in order to reconstruct the cost function. In general, there is no straight forward procedure to do so. However, in this simple case we can find heuristically an appropriate cost function. We recognize that the right-hand (λ_2) term in (3.13) is a discrete five-point approximation of

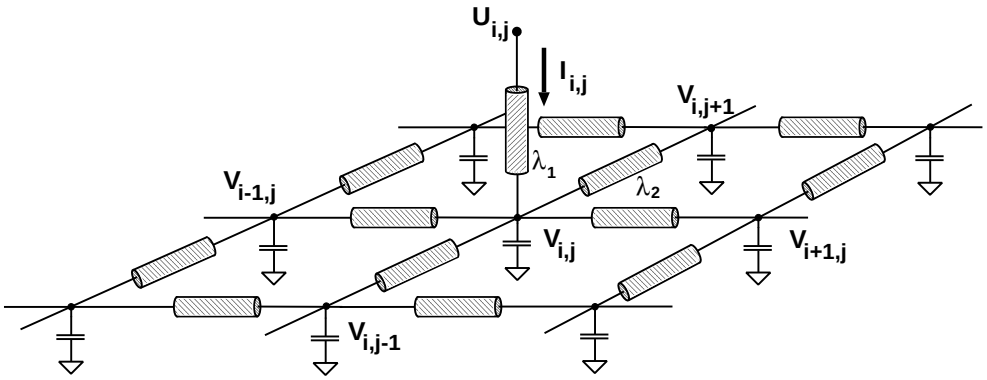


Figure 3.4: *Patch of a resistive network.* For readability, only a single input node $U_{i,j}$ is shown.

the Laplacian $\nabla^2 V_{ij}$ on a rectangular grid with bin-size one. Now, it is straight forward to show that

$$H(V) = \frac{\lambda_1}{2} \sum_{ij} (U_{ij} - V_{ij})^2 + \frac{\lambda_2}{2} \sum_{ij} ((\Delta_i V_{ij})^2 + (\Delta_j V_{ij})^2) \tag{3.14}$$

is a valid cost function for the resistive network. ΔV_{ij} represents the discrete approximation of the first order partial derivatives of V_{ij}.

The cost function (3.14) is the combination of two constraints: the first constraint restricts the output V_{ij} not to be too far from the input U_{ij} (data term) while the second one requires the spatial variation of the output units to be small (smoothness term). The relative weighting of both constraints is determined by the parameters λ_1, λ_2 that represent the uniform input and lateral conductances in the resistive network. If we assume the lateral conductances to be zero, for example, each unit is disconnected from its neighbors. As a consequence, the smoothness constraint is disabled and the output of the network follows perfectly the input.

The resistive network is the simplest embodiment of **standard regularization**. Regularization techniques usually impose a data and a smoothness constraint on the solution. They have been widely proposed as a common framework to solve problems in early vision [Poggio et al. 1985, Bertero et al. 1987, Yuille 1989, Szeliski 1996] which usually require smooth spatial interaction. We note that in terms of its connectivity, the resistive network is a very efficient implementation of a convolution engine with simple exponential kernels.

Chapter 4

Motion Perception Networks

'But it is remarkable, on the other hand, that a sure guide is found in physical interpretation: an analytic problem always being correctly set ('well-posed'), in our use of the phrase, when it is the translation of some mechanical or physical question.' (Jacques Hadamard in: Lectures on Cauchy's Problem in Linear Partial Differential Equations; Yale University Press, 1923)

4.1 Estimation of Optical Flow

A novel system approach for the estimation of two-dimensional optical flow is presented. The approach is **gradient based** and it extends an existing algorithm [Horn and Schunck 1981] such that well-posedness and robustness are guaranteed independently of the visual input applied. In particular, I will demonstrate how the computational processes of this approach can be mapped to the physical dynamics of an analog electronic network which offers computationally very efficient solutions for optical flow estimation. A hardware implementation of such network will be discussed in Chapter 5.

In the following, I will stress the notion of a **system** because a system represents an entity that dynamically defines the relations between the input it receives, and the output that it generates. Furthermore it also implies a sense of completeness: A system is defined preferably for the complete range of possible combinations of input, output and internal states. Operating under real-world conditions this is a crucial property because the environment is given and non-trivial and even for moderate levels of complexity it is usually impractical to predefine all possible states *a priori* such that exceptions might be identified for which the system does not work correctly. An ideal system therefore works and behaves reasonably under all conditions.

The **input** to the proposed system is always assumed to be the spatiotemporal gradients $E_x(x,y,t), E_y(x,y,t)$ and $E_t(x,y,t)$ of the brightness distribution of the visual scene rather than the brightness itself. It is evident that in a physical system the transduction of

the visual information and the extraction of the spatiotemporal gradients has to be carried out in a first processing stage that has a significant impact on the overall performance of the subsequent optical flow system. For the theoretical analysis within this chapter such input stage is not considered and the input is assumed to be given. However, in terms of a complete implementation this has to and will be discussed (Chapter 5). The systems **output** is, of course, the vector field $\boldsymbol{v}(x, y, t) = (u, v)$ representing the instantaneous estimate of the optical flow in image space.

According to the methodology introduced in the previous chapter, the system is characterized by different constraints that do express the applied models and assumptions. These constraints are weighted and combined such that they express an optimization problem that has to be solved in order to obtain the desired optical flow estimates.

4.1.1 The optical flow model: smooth and biased

The first constraint imposed is the **brightness constraint**[1]. The brightness constraint equation describes a line in velocity space that represents the set of all possible velocity estimates with respect to the input. The function

$$F(\boldsymbol{v}(x, y, t))^2 = (E_x u + E_y v + E_t)^2 \tag{4.1}$$

is a measure of how good the motion estimate satisfies the brightness constraint. Since the Euclidean distance e between the brightness constraint line and a point $P = (u_p, v_p)$ representing a motion estimate, is given by

$$e = \frac{|E_x u_p + E_y v_p + E_t|}{\sqrt{E_x^2 + E_y^2}} \,, \tag{4.2}$$

the measure (4.1) can be seen as the square of this distance times the absolute brightness gradient. This implies that if the intensity contrast and thus the signal-to-noise ratio is high, the input information is more reliable and therefore a deviation of the motion estimate from the constraint line is less tolerated than if the contrast is low. This is a feasible assumption. The brightness constraint alone can be sufficient to define a least-square solution for global motion.

There is an interesting relation to the flow dynamics in fluids: If the brightness distribution is identified with the density distribution of an incompressible fluid then the brightness constraint equation describes exactly the two-dimensional dynamics of the fluid under the assumption of *preserved total mass*. This analogy plausibly illustrates that the brightness constraint also holds for non-rigid objects as long as the total brightness (total mass) of the object stays constant. The restriction $\mathrm{div}(\boldsymbol{v}) = 0$ which excludes the presence of sources also follows directly from the assumption of constant mass.

[1] see Equation 2.2, Chapter 2

The assumption that objects with smooth contours introduce smooth flow fields is reasonable and has been shown to be immediately related to the rigidity of the objects observed [Ullman and Yuille 1987]. Smoothness assumes that locations in the image close to each other are more likely to belong to the same moving object and thus, are more likely to have the same visual motion. Smoothness can be imposed by requiring the spatial gradients of the flow field to be minimal. The **smoothness constraint**

$$S(\boldsymbol{v}'(x,y,t)) = u_x^2 + u_y^2 + v_x^2 + v_y^2 \tag{4.3}$$

defines a quadratic measure of how large the flow field varies across the image space. It is only defined for flow fields with continuous partial derivatives. For convenience, we require the flow field to be at least twice continuously differentiable. The algorithms by Horn and Schunck [1981] and related ones [Hildreth 1983] are using such a smoothness measure together with the brightness constraint to estimate optical flow. In fact, applying the smoothness constraint forces the solution to behave according to a physical *membrane model*. Since (4.3) constrains the derivatives rather than the flow field itself, there remain input conditions under which the optimization is ill-conditioned or even ill-posed.

Therefore, a third term is introduced which I call the **bias constraint**. Expressed as

$$B(\boldsymbol{v}(x,y,t)) = (u-u_0)^2 + (v-v_0)^2, \tag{4.4}$$

it measures how close the estimated flow vector is to some reference motion (u_0, v_0). The reference motion can be understood as the *a priori* expected motion in case the visual information content is unreliable or missing. This reference motion can undergo changes according to some long-term adaptation process such that it represents *e.g.* the statistical mean of the experienced visual motion. Such an adaptation process further increases the computational power of the approach.

Finally, the three constraints above are combined to express the following constraint optimization problem:

Given the input $E_x(x,y,t), E_y(x,y,t)$ and $E_t(x,y,t)$ on an image region $\Omega \subset \mathbb{R}^2$, find the optical flow field $\boldsymbol{v}(x,y,t) \in \mathbb{R}^2$ such that the cost function

$$H(\boldsymbol{v}(x,y,t);\rho,\sigma) \;=\; \int_\Omega \left(F^2 + \rho S + \sigma B\right)\; \mathrm{d}\Omega \;=\; \min\,! \tag{4.5}$$

$$\text{with} \quad \boldsymbol{v}'(x,y,t) \;=\; \mathbf{0} \quad \text{along} \quad \partial\Omega\,. \tag{4.6}$$

The weight of each constraint is defined by the parameters ρ and σ. In the first instance, these parameters are assumed to be constant. However, later we will see how

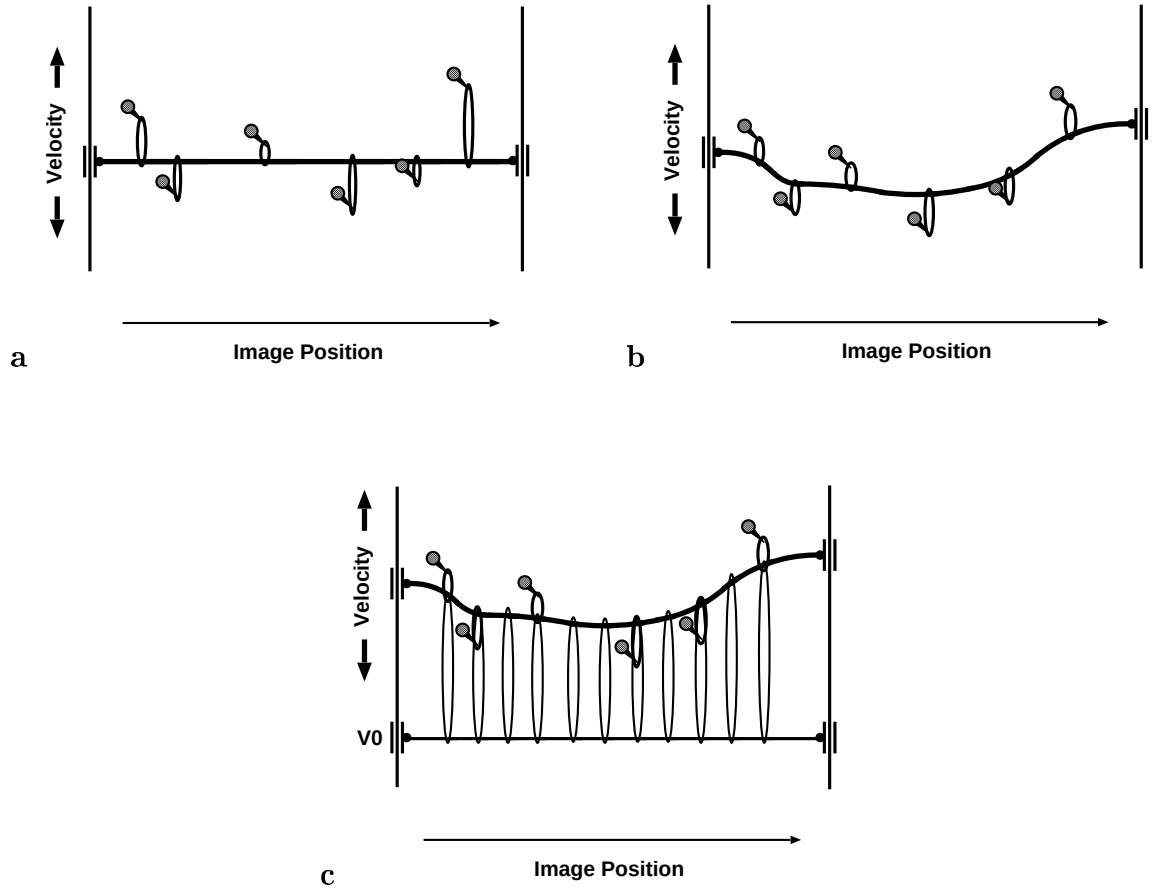


Figure 4.1: *String models.* One-dimensional mechanical equivalents for the optical flow models are shown where the motion estimate is represented by the position of a one-dimensional membrane (string) relative to some resting position under the influence of external forces applied by rubber bands. The binding conditions at the end points of the string are horizontally fixed but vertically free. (a) Global least-square solution to the sum of the brightness constraint equations (4.1). The string is totally stiff and the boundary conditions ensure its horizontal position. The elasticity constant of each individual rubber band is proportional to the local brightness gradient. (b) Introducing the smoothness constraint allows the string to bend and thus accounts for local solutions. The tension of the string is given by the weighting parameter ρ. Obviously, for no visual input, the rubber band strengths are zero and the string is floating which is an ill-posed condition. (c) Introducing the bias constraint is equivalent to introducing another set of rubber bands connecting the string to some reference level V_0. The elastic constant of these bands is proportional to the parameter σ. The position and shape of the string is always well-defined.

they can be a function of time, space and the current optical flow estimate and so permit the computational behavior to be altered significantly.

We assume the motion gradients to vanish at the boundary $\partial\Omega$, that is that the area outside Ω does not influence our motion estimate. This assumption is reasonable, because we do not have any information available outside our image space. In the mechanical equivalent (Figure 4.1) the binding of the string is such that it can freely move along the $\boldsymbol{v}$-axis. The tension of the string (proportional to ρ) ensures that the gradient vanishes at the boundaries.

A well-posed problem

The constraint optimization problem (4.5) can be solved using *variational calculus*. However, before we show an appropriate network model that can solve the problem, we want to make sure that it has always a solution which is unique and depends continuously on the input, thus it is said to be well-posed.

We prove the following theorem:

Theorem. *The estimation of optical flow formulated as the constraint optimization problem* (4.5) *is well-posed. In particular, this holds under all possible visual input conditions.*

Proof. Consider the integrand

$$L(\boldsymbol{v}(x,y,t), \boldsymbol{v}(x,y,t)', x, y; \rho, \sigma) = F^2 + \rho S + \sigma B \tag{4.7}$$

of the cost function (4.5). L is continuously defined for all $\boldsymbol{v}(x,y,t) \in \mathbb{R}^2$ that are sufficiently smooth. Hence, each twice continuously differentiable flow field that fulfills the boundary condition is a candidate for a solution. According to proposition 1 in Appendix A, the variational problem (4.5) has a unique solution $\boldsymbol{v}_0$, if the associated Euler-Lagrange equations have a solution, the integrand L is strictly convex and the natural boundary condition

$$L_{\boldsymbol{v}'}(\boldsymbol{v}_0(x,y,t), \boldsymbol{v}_0(x,y,t)', x, y) = 0 \quad \text{holds.} \tag{4.8}$$

For (4.5), the associated Euler-Lagrange equations are

$$\begin{aligned} \rho(u_{xx} + u_{yy}) &= E_x(E_x u + E_y v + E_t) + \sigma(u - u_0) \\ \rho(v_{xx} + v_{yy}) &= E_y(E_x u + E_y v + E_t) + \sigma(v - v_0). \end{aligned} \tag{4.9}$$

They represent a typical system of stationary, inhomogeneous *Poisson* equations. For the required *von Neumann* boundary conditions (4.6), existence of a solution is guaranteed [Bronstein and Smendjajew 1996].

Strict convexity of the integrand is given if the Hessian $\mathsf{J} = \nabla^2 L(\boldsymbol{v}, \boldsymbol{v}', x, y)$ is positive definite, thus all eigenvalues are real and positive[2]. We find the Hessian for the integrand L to be

$$\mathsf{J} = \begin{pmatrix} E_x^2 + \sigma & E_x E_y & 0 & 0 & 0 & 0 \\ E_y E_x & E_y^2 + \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho \end{pmatrix} \tag{4.10}$$

The matrix J is symmetric and three real and distinct eigenvalues are found:

$$\lambda_1 = \rho \quad , \quad \lambda_2 = \sigma \quad \text{and} \quad \lambda_3 = \sigma + E_x^2 + E_y^2 \quad . \tag{4.11}$$

Since we assume the weighting parameters ρ and σ to be positive, the eigenvalues are always positive independently of the visual input. Therefore, the integrand L is strictly convex. It is straight forward to show that the optical flow gradient vanishes at the image boundaries due to the natural boundary condition (4.8) since

$$L_{\boldsymbol{v}'} \;=\; 2\rho\; \boldsymbol{v}(x,y,t)' \;\stackrel{!}{=}\; \mathbf{0} \qquad \text{only for} \quad \boldsymbol{v}'(x,y,t) = \mathbf{0}. \tag{4.12}$$

Thus all conditions of proposition 1 are fulfilled and the optimization problem has a global minimum and therefore a unique solution. Again, this is truly independent of the brightness gradients and thus the visual input.

Finally, continuity of the solution on the input is guaranteed because L is a continuous function of $\boldsymbol{v}(x, y, t)$ and the spatiotemporal gradients $E_x(x, y, t)$, $E_y(x, y, t)$ and $E_t(x, y, t)$. Thus the proof is complete. □

Boundary conditions

If σ is set to zero and therefore the bias constraint looses its influence and is abandoned, the cost function (4.5) is equivalent to the smooth optical flow formulation of Horn and Schunck [1981]. Now, at least one eigenvalue in (4.11) is zero. The cost function is still convex but no longer strictly convex and thus multiple minima might coexist. Local minima occur when the local brightness gradients are not independent throughout the whole image. That means either they are zero and therefore no features are present or there are only single oriented edges perceived such that the aperture problem holds[3]. In these cases the problem is ill-posed if special assumptions are not made concerning the boundary conditions. Not surprisingly some authors [Poggio et al. 1985] suggest that the visual input space be restricted to guarantee well-posedness. However, such

[2] see Appendix A

[3] *i.e.* the strong aperture problem or the aperture problem in the large

restrictions prevent a continuous operation under real-world conditions and require a separate threshold operation.

Applying resistive networks in visual processing tasks [Horn 1988], it was recognized that in order to be well-posed, membrane models typically require a given value on its boundaries. Thus Koch et al. [1991] proposed an approach where the flow estimate $\boldsymbol{v}(x, y, t)$ is set to zero along the image boundary $\partial\Omega$. Clearly, there is no rational reason for such an assumption other than to prevent ill-posedness. The assumption of a fixed motion value at the image boundary cannot be legitimated by any reasonable model. It heavily influences the estimation of optical flow near the image boundary and implies an unnecessarily strong bias.

The here introduced bias constraint, however, ensures well-posedness of the problem while allowing natural boundary conditions. The formulation within the constraint optimization framework allows to express a *tendency* rather than a commitment to some reference motion. The strength or impact of this tendency not only depends on a continuous parameter σ but also on the visual input such that it only affects the estimate significantly when the visual input is ambiguous.

4.1.2 Network architecture

The minimization problem (4.5) so far is formulated continuously on the image space $\Omega \subset \mathbb{R}^2$. Since the visual input is usually provided by some imaging device with finite spatial resolution, however, it is feasible to integrate the cost function only over locations where the input is accessible. We discretize the image space on an orthogonal basis and label each node by two integer values $i \in [1 \ldots n]$ and $j \in [1 \ldots m]$. The nodes define a quadratic grid with space constant h. The estimation of optical flow is therefore reduced to a finite number of locations which allows to consider each individual optical flow vector as the resulting output of a single processing unit in a network.

The discrete form of the cost function (4.5) is

$$\begin{aligned} H(\boldsymbol{v}; \rho, \sigma) = \sum_{i=1}^{n} \sum_{j=1}^{m} \big[& (E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}})^2 + \\ & \rho((\Delta u_{ij})^2 + (\Delta v_{ij})^2) + \sigma((u_{ij} - u_0)^2 + (v_{ij} - v_0)^2) \big] \end{aligned} \tag{4.13}$$

where the partial derivatives of the flow vectors in the smoothness term are replaced by a symmetric difference operator[4] Δ such that

$$(\Delta x_{i,j})^2 = \left(\frac{x_{i+1,j} - x_{i-1,j}}{2h} \right)^2 + \left(\frac{x_{i,j+1} - x_{i,j-1}}{2h} \right)^2 .$$

[4] not to be confused with the Laplace operator ∇^2

We have previously shown that our optimization problem has only a global minimum. Therefore, it follows directly that

$$H'(\boldsymbol{v_0}; \rho, \sigma) = 0 \tag{4.14}$$

is a sufficient condition for the unique solution $\boldsymbol{v_0}$. Thus partial differentiation of the discretized cost function (4.13) leads to a linear system of $2n \times m$ equations. Instead of solving this linear system analytically we proceed differently. We create a dynamical system that performs steepest gradient descent on the cost function (4.13): the estimated optical flow components $u_{i,j}$ and $v_{i,j}$ are altered negatively proportional to the partial gradients of the cost function thus,

$$\dot{u}_{ij} \propto -\frac{\partial H(\boldsymbol{v}; \rho, \sigma)}{\partial u_{ij}} \qquad \text{and} \qquad \dot{v}_{ij} \propto -\frac{\partial H(\boldsymbol{v}; \rho, \sigma)}{\partial v_{ij}} \tag{4.15}$$

until steady state is reached. This leads to the following system of $2n \times m$ linear partial differential equations

$$\begin{aligned}
\dot{u}_{ij} = -\frac{1}{C}\big[&E_{x_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) - \\
&\rho(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}) + \sigma(u_{ij} - u_0)\big], \\
\dot{v}_{ij} = -\frac{1}{C}\big[&E_{y_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) - \\
&\rho(v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{ij}) + \sigma(v_{ij} - v_0)\big]
\end{aligned} \tag{4.16}$$

for all $i \in [1 \ldots n],\ j \in [1 \ldots m]$ and where C is a constant.

Why transform a system of linear equations into a system of linear differential equations?

Because, employing such dynamics might allow us to identify some equivalent physical system onto which the optimization problem can be mapped. And in fact, the dynamical system (4.16) should appear familiar to us: The second and third term in the right-hand side describe the dynamics of the resistive network as shown in the previous chapter[5]. It follows that it is relatively straight-forward to construct an equivalent analog electronic network for the purpose of optical flow estimation. We identify the optical flow components each with the potential on a node in a resistive network with capacitance C. The right-hand of (4.16) in square brackets represents a sum of currents that charges up or down these nodes until equilibrium is reached.

The total *dissipated power* in the resistive network is actually the physical equivalent of the total costs induced by the bias and smoothness constraint[6]. For constant conductances ρ and σ the behavior of the resistive network is ohmic, and the dissipated power is

[5] see Equation (3.13)

[6] Note, that using the expression *energy* instead of *cost function* would have been confusing here in terms of its physical meaning.

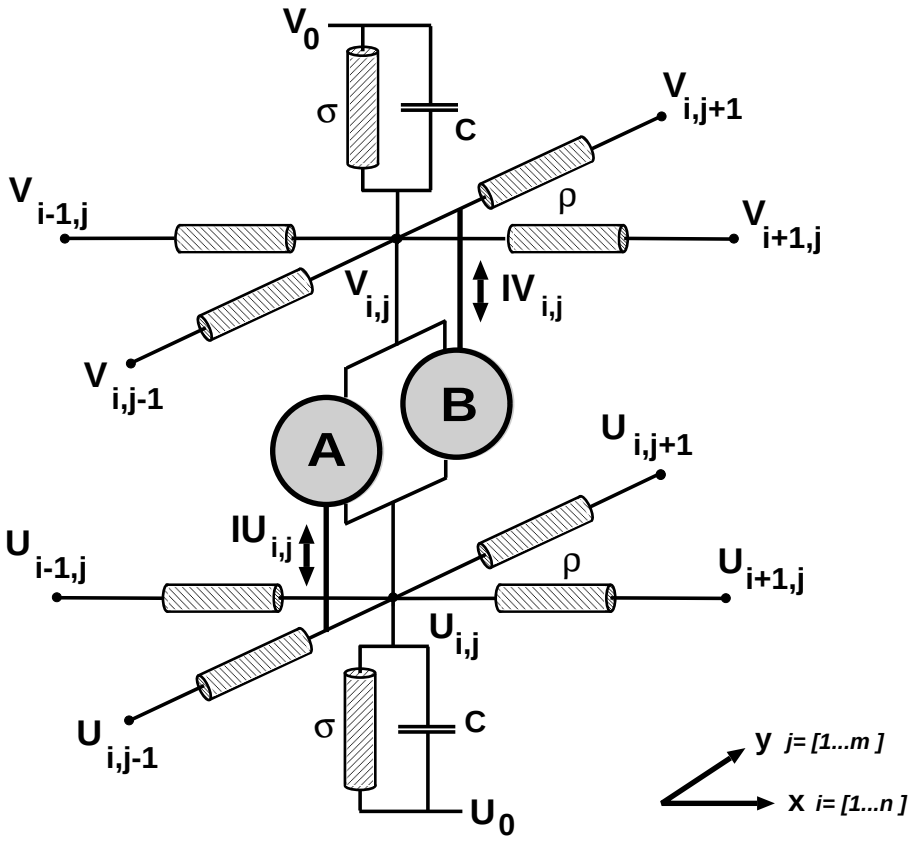


Figure 4.2: *Optical flow network architecture.* A single unit of the optical flow network is shown. The smoothness constraint is embodied in two resistive networks composed of the conductances ρ. The components of the local optical flow vector are encoded as the voltages $U_{i,j}$ and $V_{i,j}$ at each node of these networks. The vertical conductance σ accounts for the bias of the optical flow estimate towards some reference motion represented by the potentials U_0 and V_0. A and B constitute active circuitry that acts as a current source or sink depending on the deviation of the optical flow estimate from the local brightness constraint.

equivalent to the current flowing through the conductances times the potential difference across a node. For non-constant conductances, the dissipated power is given by the total resistive co-content which is the integral of the current over the potential difference [Harris and Koch 1989].

The schematics in Figure 4.2 sketch a single unit of the optical flow network. We can clearly recognize that each constraint of the cost function (4.13) has a physical counterpart: smoothness is enforced by the resistive networks with lateral conductances proportional to the weighting parameter ρ. The bias constraint is implemented as a leak conductance σ to some reference potentials V_0 and U_0. Only the realization of the brightness constraint needs some active circuitry represented by the boxes A and B. At each node these boxes compare the instantaneous local estimate of optical flow with the visual input and inject or sink the following currents $IU_{i,j}$ and $IV_{i,j}$ such that

$$\begin{aligned} IU_{i,j} &\propto -E_{x_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) \\ IV_{i,j} &\propto -E_{y_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}). \end{aligned} \tag{4.17}$$

The circuitry within these boxes includes several multiplication and summing operations. Since the components of the optical flow vector can take on negative values, the actual encoding has to be seen as the differences of the voltages in the resistive network to some null-potential.

4.1.3 A dynamical minimization problem

We have assumed that the gradient descent dynamics decrease the cost function (4.13) along all its possible trajectories and the system globally converges to a single solution as $t \to \infty$. If this is true, $H(\boldsymbol{v}; \rho, \sigma)$ is a Lyapunov function of the system, *i.e.*

$$\frac{\mathrm{d}H}{\mathrm{d}t} \leq 0 \qquad \text{with} \qquad \frac{\mathrm{d}H}{\mathrm{d}t} = 0 \quad \text{only for} \quad \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = 0. \tag{4.18}$$

With (4.16) we can show that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{ij}\Big[\frac{\partial H}{\partial u_{ij}}\frac{\mathrm{d}u_{ij}}{\mathrm{d}t} + \frac{\partial H}{\partial v_{ij}}\frac{\mathrm{d}v_{ij}}{\mathrm{d}t}\Big] = -\frac{1}{C}\sum_{ij}\Big[\big(\frac{\partial H}{\partial u_{ij}}\big)^2 + \big(\frac{\partial H}{\partial v_{ij}}\big)^2\Big] \leq 0 \tag{4.19}$$

which is obviously true for all possible trajectories of $\boldsymbol{v}$. However, the underlying assumption for this analysis was that the input, thus the spatiotemporal brightness gradients, is constant over time. This is of course not very sensible *per se* because our optical flow system particularly requires information of the temporal structure of the image brightness.

So far we have considered that the system always moves down the steepest gradient on the surface of a static cost function until it reaches the global minimum. Now we must modify this picture to allow the cost function to change as time evolves. The problem is

no longer static but rather a dynamical optimization problem. Taking into account the temporal structure of the input, we must rewrite the total temporal derivative of the cost function as

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{ij} \left[\frac{\partial H}{\partial u_{ij}} \frac{\mathrm{d}u_{ij}}{\mathrm{d}t} + \frac{\partial H}{\partial v_{ij}} \frac{\mathrm{d}v_{ij}}{\mathrm{d}t} + \frac{\partial H}{\partial E_{x_{ij}}} \frac{\mathrm{d}E_{x_{ij}}}{\mathrm{d}t} + \frac{\partial H}{\partial E_{y_{ij}}} \frac{\mathrm{d}E_{y_{ij}}}{\mathrm{d}t} + \frac{\partial H}{\partial E_{t_{ij}}} \frac{\mathrm{d}E_{t_{ij}}}{\mathrm{d}t} \right] \quad (4.20)$$

Although the first two terms are less than or equal to zero as shown before (4.19), this is not necessarily the case for the total expression. The cost function is not a Lyapunov function in its strict definition. However, if we can ensure that the dynamics of our network are much faster than the input dynamics (thus *e.g.* $\dot{u}_{ij} \gg \mathrm{d}E_{x_{ij}} / \mathrm{d}t$), then the last three terms in (4.20) can be neglected. Hence, we can treat the optimization problem to be quasi-stationary and safely assume that the estimated flow field is always at the momentary global minimum of the cost function – or at least very close to it. Consider again the mechanical analogy of the string as shown in Figure 4.1, and assume now that the position of the pins, and thus the external forces applied by the rubber bands, change. Then, the shape of the string immediately adapts to the new configuration such that it will always represent the state of minimal costs. A critical limit is reached when the changes of the external forces happen on such a short time-scale that the mass of the string becomes relevant for its dynamics, and thus the kinematic energy of the system must be considered.

The dynamics in our optical flow system are mainly determined by the network capacitance C and the time constant τ_{rec} of the active recurrent circuitry represented by the boxes A and B in Figure 4.2. Those parameters are to a large extent under active control of the designer of the system, and therefore can be kept as small as possible[7].

4.1.4 Computational behavior

The computational characteristics of the system mainly depend on the relative impact of the three different constraints and thus on the values of ρ and σ. According to the values of these parameters, the computational behavior of the system changes significantly and exhibits different models of visual motion estimation such as normal flow, smooth local flow or global flow. Figure 4.3 schematically illustrates the different operational regimes the system exhibits depending on the two parameters. The textured oval region represents the parameter space in which the system provides robust and sensible optical flow estimation. The limits of sensible operation are determined by the bias conductance σ. For values approaching zero, computation becomes ill-posed whereas for high values, the behavior is trivial because the bias constraint dominates such that the system constantly reports the reference values independently of the visual input.

[7] see discussion on processing speed in Section 5.3.4

The notion of computational complexity of the system is interesting. In Figure 4.3 a contour plot illustrates the different levels of computational complexity. As we will see later, the computational behavior of the system in the regimes of purely normal or global flow can be described by mathematically closed-form expressions. Using traditional sequential computers such feed forward computation needs much less effort and thus time than the highly iterative process required for smooth local flow estimation. If we consider the execution time (computational load) to be a fair measure of computational complexity, then the estimation of smooth local flow is computationally most complex. However, in the context of the analog network proposed here it is not obvious how to define an equivalent measure of the computational load. Clearly, the active processing in each unit of the network does not change for different values of σ and ρ. So the processing load for each unit remains the same independently of whether it is estimating normal flow or contributing to a global flow estimate. The only difference in behavior is induced by the parameter ρ. Thus, what an observer of the system finally recognizes as a computational behavior of different complexity reduces to the amount of interaction within the network!

In the following, simulation results are presented that illustrate the different computational behaviors in relation to the parameters ρ and σ. The artificial image sequence used as input consists of a triangle of uniform brightness moving on a stationary lightly structured or unstructured background. The triangle undergoes horizontal translational motion to the right with a constant speed. Any additional information on the simulation methods and the stimulus sequence can be found in Appendix B.

The lateral coupling conductance ρ

The conductance ρ is an important network parameter because it controls the degree of communication between single units. If $\rho = 0$, the lateral coupling amongst the units in the network is disabled. In this case, the network performs an optical flow estimate where each unit reports the flow vector that is on its local brightness constraint line and best matches the reference motion $\boldsymbol{v}_0$ enforced by the bias constraint. By the prohibition of lateral interaction, the network actually breaks up into its single processing units. In this case, the output of each unit can be computed analytically by applying the necessary condition $H'(\boldsymbol{v}; \rho, \sigma) = 0$, thus

$$\begin{aligned} E_{x_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) + \sigma(u_{ij} - u_0) &= 0 \\ E_{y_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) + \sigma(v_{ij} - v_0) &= 0\ . \end{aligned} \tag{4.21}$$

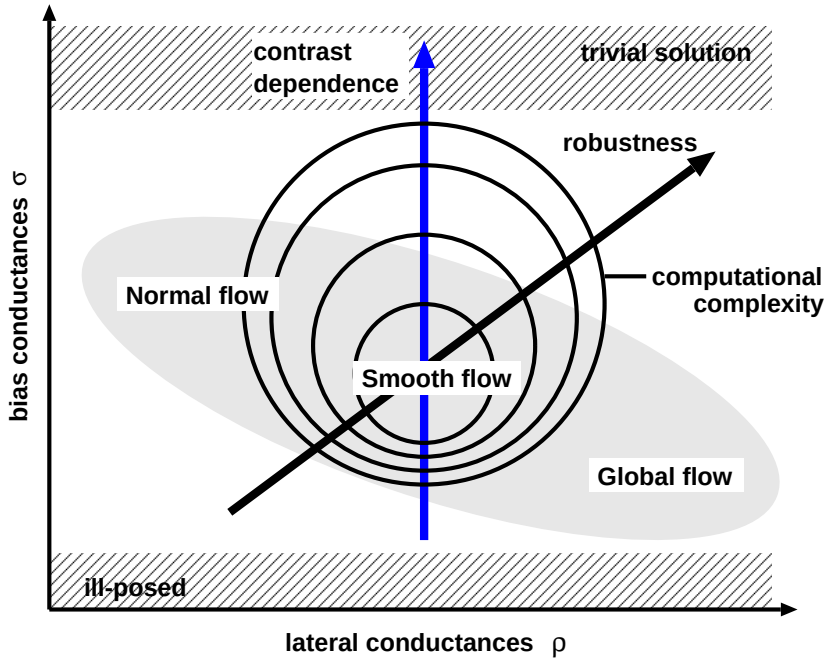


Figure 4.3: *Computational complexity dependent on ρ and σ.* The computational characteristics of the optical flow system is shown in dependence of the parameters ρ and σ. None or very weak lateral coupling amongst the optical flow units combined with a finite value of σ leads to a normal flow estimate. A large lateral conductance on the other hand allows a global estimate of the visual motion. Robustness typically increases with larger integration areas (large ρ). Larger values of σ also increase robustness but are not desired because they strengthen the contrast dependence of the motion estimate significantly. The system shows its computationally richest and most complex behavior for small σ's and intermediate ρ's. This is indicated by the symbolic contour plot showing the different levels of computational complexity.

Solving for the local flow vector $\boldsymbol{v}_{ij}$ finally leads to

$$\begin{aligned} u_{ij} &= -\frac{E_{t_{ij}} E_{x_{ij}}}{\sigma + E_{x_{ij}}^2 + E_{y_{ij}}^2} + \frac{u_0(\sigma + E_{y_{ij}}^2) + v_0(E_{x_{ij}} E_{t_{ij}})}{\sigma + E_{x_{ij}}^2 + E_{y_{ij}}^2} \\ v_{ij} &= -\frac{E_{t_{ij}} E_{y_{ij}}}{\sigma + E_{x_{ij}}^2 + E_{y_{ij}}^2} + \frac{v_0(\sigma + E_{x_{ij}}^2) + u_0(E_{y_{ij}} E_{t_{ij}})}{\sigma + E_{x_{ij}}^2 + E_{y_{ij}}^2} \end{aligned} \quad . \tag{4.22}$$

If we apply a typical bias for slow motion, *i.e.* $\boldsymbol{v}_0 = \mathbf{0}$, the second term on the right-hand side of Equations (4.22) vanishes. With $\sigma \to 0$, Equation (4.22) provides an infinitely close approximation of a *normal flow estimate* as shown in Figure 4.4b. Only local information is processed and therefore the aperture problem holds dominantly: The resulting optical flow is not identical to the correct object motion. Note that the vector average of the normal flow does not resolve the object motion either. The bias conductance σ is required

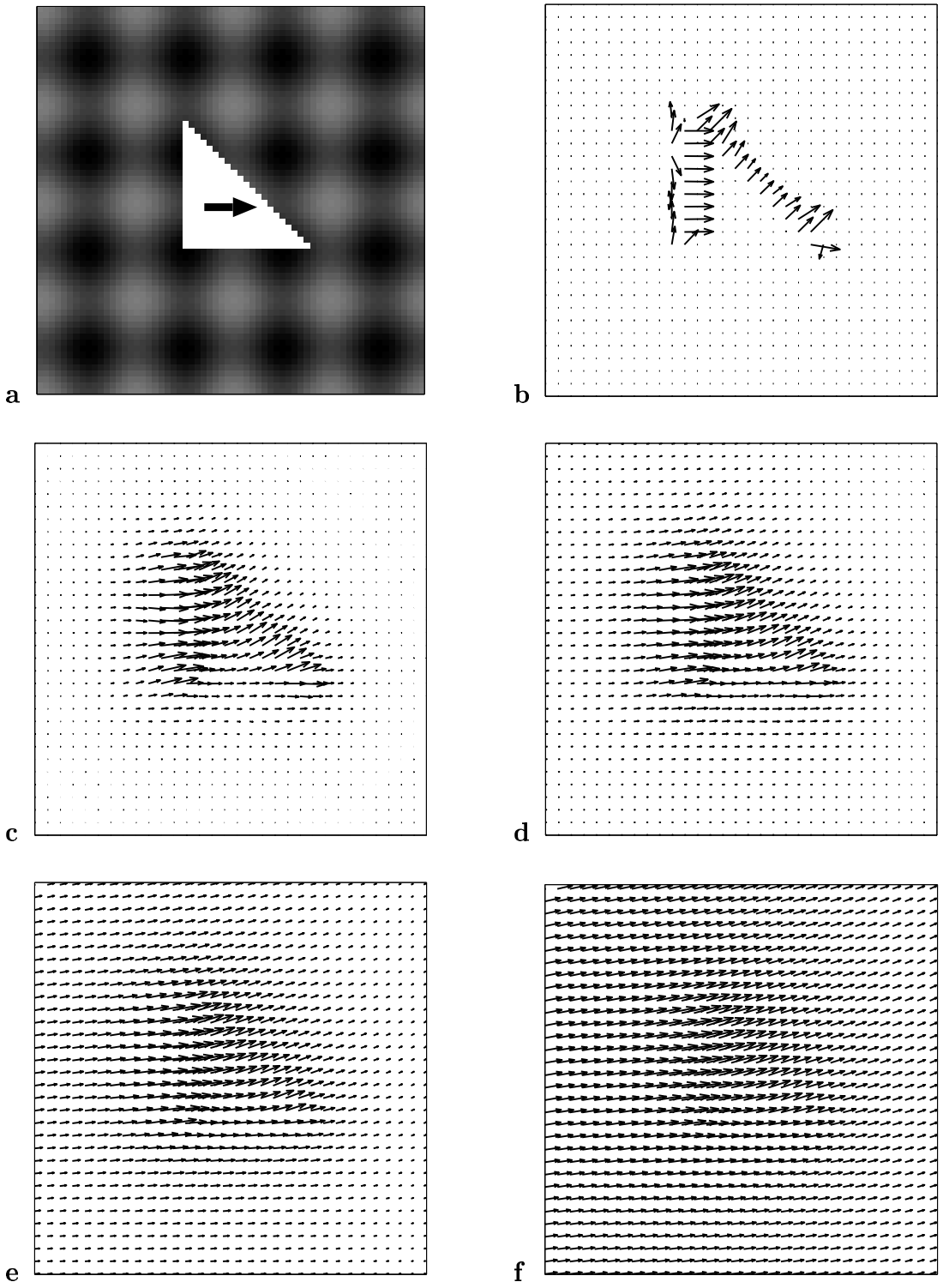


Figure 4.4: *Varying the smoothness strength ρ.* The network shows a wide range of computational behaviors. (b)-(f) Keeping $\sigma = 0.001$ constant and increasing $\rho = [0, 0.05, 0.15, 0.5, 1.5]$ exhibits a continuous transition from normal flow to global flow estimation.

to be non-zero to avoid division by zero. The problem is well-posed even with no visual input present. In this case the output simply follows the reference motion.

For $\rho > 0$ the lateral coupling between the units in the network is enabled and information and computation is spread amongst the units in the network. Now, the network provides estimates of *smooth optical flow*. Increasing values of ρ also increase the region-of-support and lead to smoother flow fields (Figure 4.4b-f). Obviously, there is a clear trade-off between obtaining local motion information and the possibility to solve the aperture problem: A large region-of-support is required to solve the aperture problem but the isotropic character of the resistive network cannot preserve the important optical flow discontinuities at *e.g.* object boundaries.

As $\rho \to \infty$ the smoothness constraint dominates such that it does not allow the slightest spatial deviation in the flow field. Thus the flow field is uniform within the whole image range and can be represented by a single *global flow* vector $\boldsymbol{v_g} = (u_g, v_g)$. Now, the linear system (4.14) becomes an over-complete one for which we can find the following analytical solution

$$u_g = \frac{\sum_{ij}(E_{y_{ij}}^2 + \sigma) \cdot \sum_{ij}(\sigma u_0 - E_{x_{ij}} E_{t_{ij}}) - \sum_{ij} E_{x_{ij}} E_{y_{ij}} \cdot \sum_{ij}(\sigma v_0 - E_{y_{ij}} E_{t_{ij}})}{\sum_{ij}(E_{y_{ij}}^2 + \sigma) \cdot \sum_{ij}(E_{x_{ij}}^2 + \sigma) - (\sum_{ij} E_{x_{ij}} E_{y_{ij}})^2}$$

and

$$v_g = \frac{\sum_{ij}(E_{x_{ij}}^2 + \sigma) \cdot \sum_{ij}(\sigma v_0 - E_{y_{ij}} E_{t_{ij}}) - \sum_{ij} E_{x_{ij}} E_{y_{ij}} \cdot \sum_{ij}(\sigma u_0 - E_{x_{ij}} E_{t_{ij}})}{\sum_{ij}(E_{y_{ij}}^2 + \sigma) \cdot \sum_{ij}(E_{x_{ij}}^2 + \sigma) - (\sum_{ij} E_{x_{ij}} E_{y_{ij}})^2} \tag{4.23}$$

respectively. These two equations represent the least-square solution of the optimization problem (4.5) assuming a common global motion model.

An attempt to quantitatively measure the effect of the lateral conductances on the performance of the optical flow estimation is summarized in Table 4.1, using the same artificial image sequence as in Figure 4.4. Results for the simulation examples (Figure 4.4b-f and Figure 4.5c) are shown as well as two results for global motion estimation ($\rho = 100$). Two different error measures were applied to quantify the error between the estimated and the exact flow field. In a first measure, the angle and the speed differences are considered independently. As can be noted, the angular error decreases for increasing values of ρ because a larger region-of-support gradually solves the aperture problem. On the other hand, the average speed difference increases as well since the smoothing assigns also motion to the background that is actually stationary. For a global motion estimate, the mean angular error grows again because the region-of-support extends to the complete image space and includes now two motion sources: the stationary background and the moving object. The resulting estimate represents now some sort of an average estimate between the background motion and the object motion. The zero motion in the background biases the global estimate towards the mean of the normal flow vectors inducing a directional tilt due to the 45°edge of the triangle. As can be seen, this effect is hardly affected by

regime		angle Φ		speed $\|v\|$		combined Φ_c	
ρ	σ	mean	std	mean	std	mean	std
0	0.001	47.4°	29.1°	0.07	0.25	3.3°	11.3°
0.05	0.001	15.4°	9.6°	0.11	0.19	5.4°	8.9°
0.15	0.001	12.8°	7.3°	0.16	0.18	8.2°	8.5°
0.5	0.001	12.1°	5.1°	0.25	0.15	13.2°	7.2°
1.5	0.001	12.5°	3.3°	0.33	0.11	17.9°	5.3°
100	0.001	20.1°	0.3°	0.22	0.13	13.5°	4.8°
0.05	0	9.6°	5.7°	0.14	0.19	7.4°	9.1°
100	0	20.1°	0.3°	0.26	0.13	14.0°	4.6°

Table 4.1: *The deviation of the estimated flow fields from the correct motion.* The error between the estimated and the correct motion of the 'triangle' stimulus is shown in dependence on the lateral conductance ρ. The independent angular and speed errors are calculated as well as a combined measure (see text for explanation).

the finite σ. A global estimate only reveals the object motion if the background does not provide any spatiotemporal structure (see also Figure 4.6). The speed error obviously grows with increasing smoothing because a non-zero velocity estimate is assigned to more and more units in the background.

The absolute angular error is a measure that does not consider the strength of the motion and thus is very sensitive to additive noise for small motion. We therefore introduce a combined measure that considers the angular as well as the speed error. To do so, we consider the flow vectors as the 2D projection of the three-dimensional space-time direction vector $\boldsymbol{v} = (u, v, 1)$. The combined error measure is defined then as the angular error in space-time direction between the estimated and the correct motion. Thus we define the combined error

$$\Phi_c = \arccos\left(\frac{\boldsymbol{v}_{estimated} \cdot \boldsymbol{v}_{correct}}{|\boldsymbol{v}_{estimated}|\ |\boldsymbol{v}_{correct}|}\right). \tag{4.24}$$

The same measure has been used previously to compare different optical flow methods [Barron et al. 1994] and allows us later to compare the performance of the present network with algorithms reported in the literature – at least for one particular image sequence. The combined measure favors a correctly located motion estimate, and is lowest in a pure normal flow regime. However, taking into account both error measures a modest value of $\rho \approx 0.1$ provides the best results which seems to be in good agreement with the visual impression of the flow field (Figure 4.4c,d). Note, that such quantitative error measures have to be carefully interpreted. The performance of the network for a particular parameter setting depends significantly on the properties of the visual input.

The influence of the bias conductance σ

The bias constraint is crucial for the computationally well-posed nature of the system. We noticed that it keeps the system well-posed even for completely separated units in the network which allows the estimation of normal flow. Now we want to consider its influence on smooth local flow estimates where sufficient visual input is present such that well-posedness is guaranteed even without the bias constraint.

The two columns of Figure 4.5 show the simulated responses to a visual sequence consisting of an identical object undergoing the same movement on a stationary background. The only difference is the non-textured background of the second sequence. Figures 4.5c-d show the estimated flow field for parameter values $\rho = 0.1$ and $\sigma = 0$, thus disabling the bias constraint which is equivalent to the approach of Horn and Schunck [1981]. For comparison, Figures 4.5 e-f show results for the same parameter settings but the bias constraint enabled ($\sigma = 0.005$, with $\boldsymbol{v}_0 = \mathbf{0}$). Obviously, if the bias constraint is not active, the effective smoothing kernel substantially depends on the visual input and not only on the lateral conductance ρ.

To understand this effect we consider the effective **diffusion length** in resistive networks. The diffusion length L is the distance from a spatial position x_0 at which the response to a voltage-source applied at x_0 has decayed by one e-fold. Because in our network architecture, it is typically the case that $\sigma/\rho \ll 1$, we can use the continuous approximation, in which the diffusion length $L = \sqrt{\rho/\sigma}$ [Mead 1989]. For $\sigma \to 0$, the diffusion length becomes infinite, *i.e.* a single voltage applied at some location spreads everywhere in the network. Superposition holds because the resistive network is linear. If $\sigma = 0$, the network actually performs a first order moment computation for a given voltage distribution. Returning to the example in Figure 4.5 d: Since the background does not provide any visual input, the units are not active and the infinite diffusion length causes a spreading to the image boundaries.

One can argue that spreading available information into regions of no information is a sensible process [Horn and Schunck 1981]. Clearly, under some circumstances this might be the case, *e.g.* in a scene containing a single moving object. In general however, there is not much reason to assume that for regions receiving no or ambiguous visual input, the motion information from some distant image location is a better and more likely estimation than some other, arbitrary value. Instead, using the bias constraint provides the means to assign *a priori* information. It allows us to do better than chance by forcing the motion estimate to take on a **most probable reference value** in case the input is ambiguous. As mentioned, this reference value can be seen as the result of some long term adaptive process that includes information of the history and the environmental conditions the system has to deal with.

Also, we have encountered the trade-off between the size of the region-of-support and the possibility to obtain local motion estimates. It is desirable to be able to define this

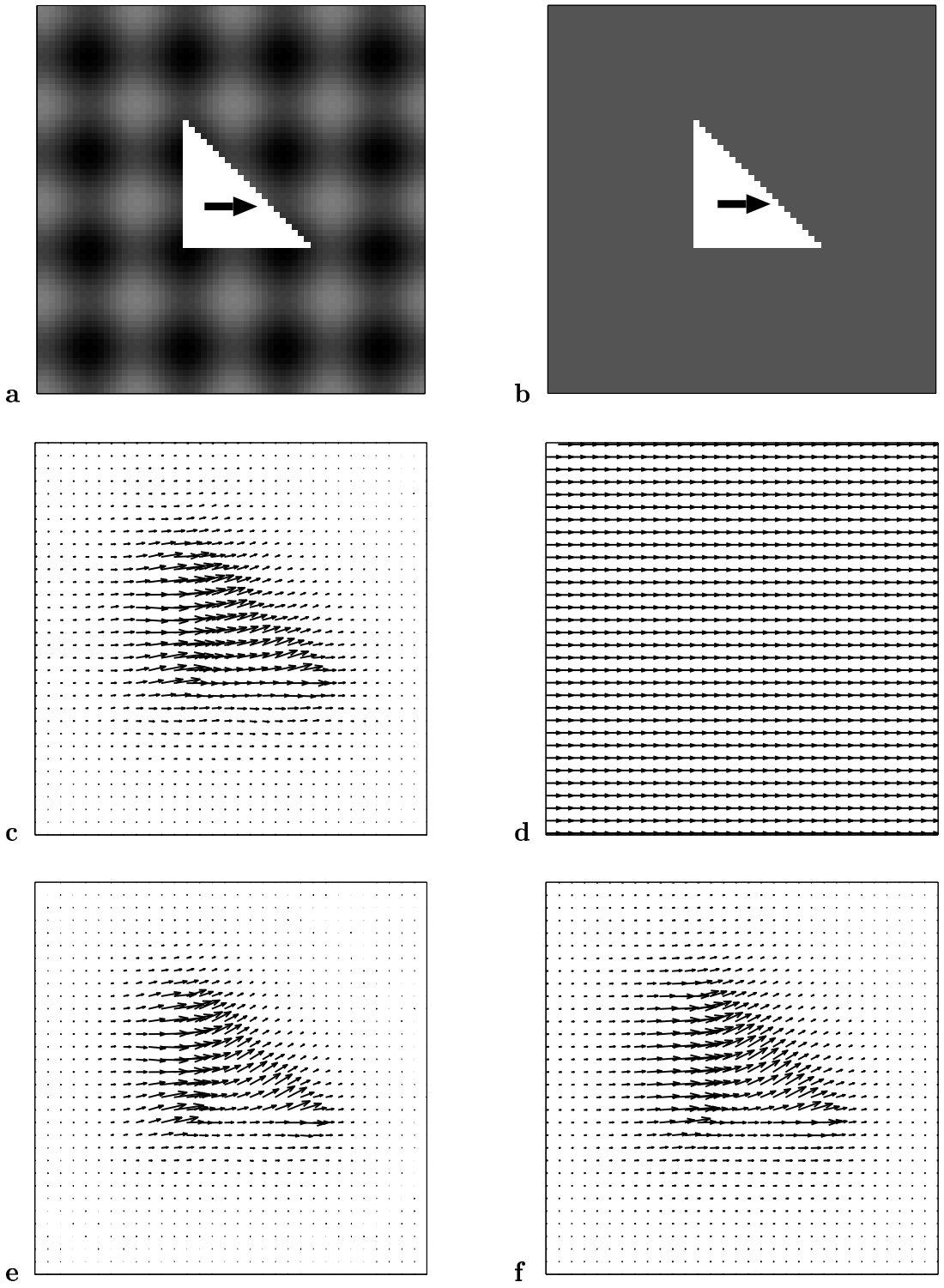


Figure 4.5: *Diffusion length dependence.* Same object, same motion but different background texture: Disabling the bias constraint ($\rho = 0.05, \sigma = 0$) leads to a drastically changed behavior (c and d). Enabling the bias constraint (e and f) ensures that the intrinsic diffusion length is finite ($\rho = 0.05, \sigma = 0.001$).

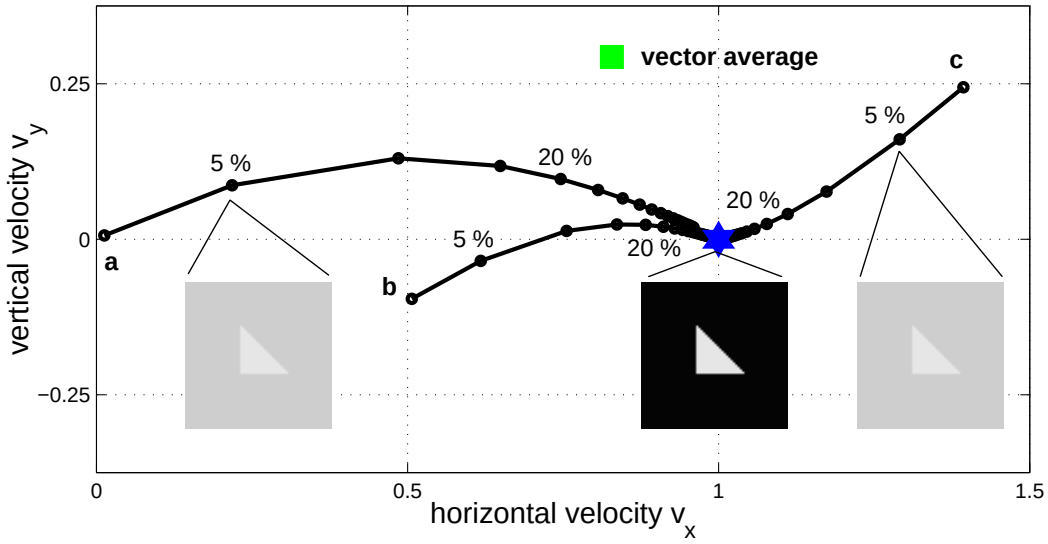


Figure 4.6: *Dependence of the perceived motion on contrast.* The global motion of the triangle stimulus is estimated for varying contrast. The correct IOC motion of the triangle is $\boldsymbol{v} = (1, 0)$. The three trajectories represent the estimates for decreasing figure-ground contrast $[95\%, \ldots, 1\%]$ for three different reference motion $\boldsymbol{v_0} = [(0, 0); (0.5, -0.1); (1.4, 0.25)]$. At high contrast, the estimate is in any case close to the true velocity $\boldsymbol{v} = (1, 0)$. Only for low contrast, the bias constraint ($\sigma = 0.001$) has a significant influence and biases the estimate towards the applied reference motion. Clearly, the direction of the perceived motion is affected also. In the particular case of $\boldsymbol{v_0} = (0, 0)$, the perceived direction of motion can continuously shift from the correct motion towards a vector average estimate depending on the contrast.

trade-off by adjusting the diffusion length without being dependent on the visual input. We will discuss later methods on how the diffusion length can be locally adapted to enhance the estimation of optical flow.

The bias constraint also affects the integration process necessary to estimate object motion. For a given σ, the strength of the effect depends mainly on the contrast and the reference motion $\boldsymbol{v_0}$ but also on the shape of the object. In general, the resulting motion estimate deviates not only in speed but also in direction from the correct intersection-of-constraint (IOC) solution. Figure 4.6 illustrates the changes of the perceived motion for the same triangle stimulus with non-textured background used before. The three trajectories represent the estimated object motion (4.23) for decreasing figure-ground contrast. Each trajectory is processed with a different reference motion. Given a reasonably low weight to the bias constraint ($\sigma = 0.001$), the object motion estimate is close to the true motion for sufficiently high contrast. For low contrast, it degrades towards the particular reference motion. For a typical slow motion preference (*e.g.* trajectory a), however,

the direction of motion shifts from the correct IOC motion towards a normal flow vector average estimate. Thus for object shapes for which the correct motion and the vector average estimate do not have the same direction, the perceived direction of motion varies with contrast.

4.1.5 Simulation results using realistic image sequences

Some qualitative behaviors of the optical flow network are investigated using more realistic and natural image sequences. The network is simulated according to the derived dynamics (4.16) where the values of the conductances ρ and σ are set as indicated. More technical details about the simulation methods and the applied image sequences are listed in Appendix B.

The 'tape-rolls' image sequence

The first example consists of an indoor image sequence where two tape-rolls are rolling on a table in opposite directions and with different relative distances to a stationary observer. Figure 4.7 shows every second frame of such sequence and the estimated and twice subsampled optical flow field. A rather moderate smoothness strength was chosen ($\rho = 0.075$) while the bias constraint ($\sigma = 0.0025$) was slightly increased compared to the previous artificial image sequences to account for the higher signal-to-noise ratio of real image sequences. The 'tape-rolls' sequence has many properties that makes it a real challenge for visual motion processing systems. First of all, the spatial and temporal sampling rate is low (60x60 pixels, 15 frames/sec) which increases the number of sampling artifacts. The low temporal sampling also leads to a maximal interframe displacement of > 2 pixels. Large interframe displacements are hard to resolve. Furthermore, the sequence contains two identical objects appearing at different spatial scales. Although the tape-rolls roll on the table, the apparent motion is largely translational. Still, especially for the larger roll, rotational components might influence the homogeneity of the flow sources. During several frames, occlusion between the two tape-rolls occurs which certainly violates the brightness constraint. Also contrast conditions are partially non-ideal and in conjunction with the present shadows lead to almost smooth object boundaries.

In general, the network produces an optical flow estimate that seems appropriate for the visual motion perceived. Although the tape-rolls are actually rotating the estimate primarily consists of the translational component of their motions because distinguishable visual features are not visible on the surface of the tape-rolls. Nevertheless, a rotational component can be observed also, especially at the left side of the roll in front (Figure 4.7e,f). As a consequence of the smoothness constraint, the regions where optical flow is reported, overlap the actual the size of the two moving tape-rolls in the image. Because of this, the part of the table that is visible through the hole of the tape is assigned to move

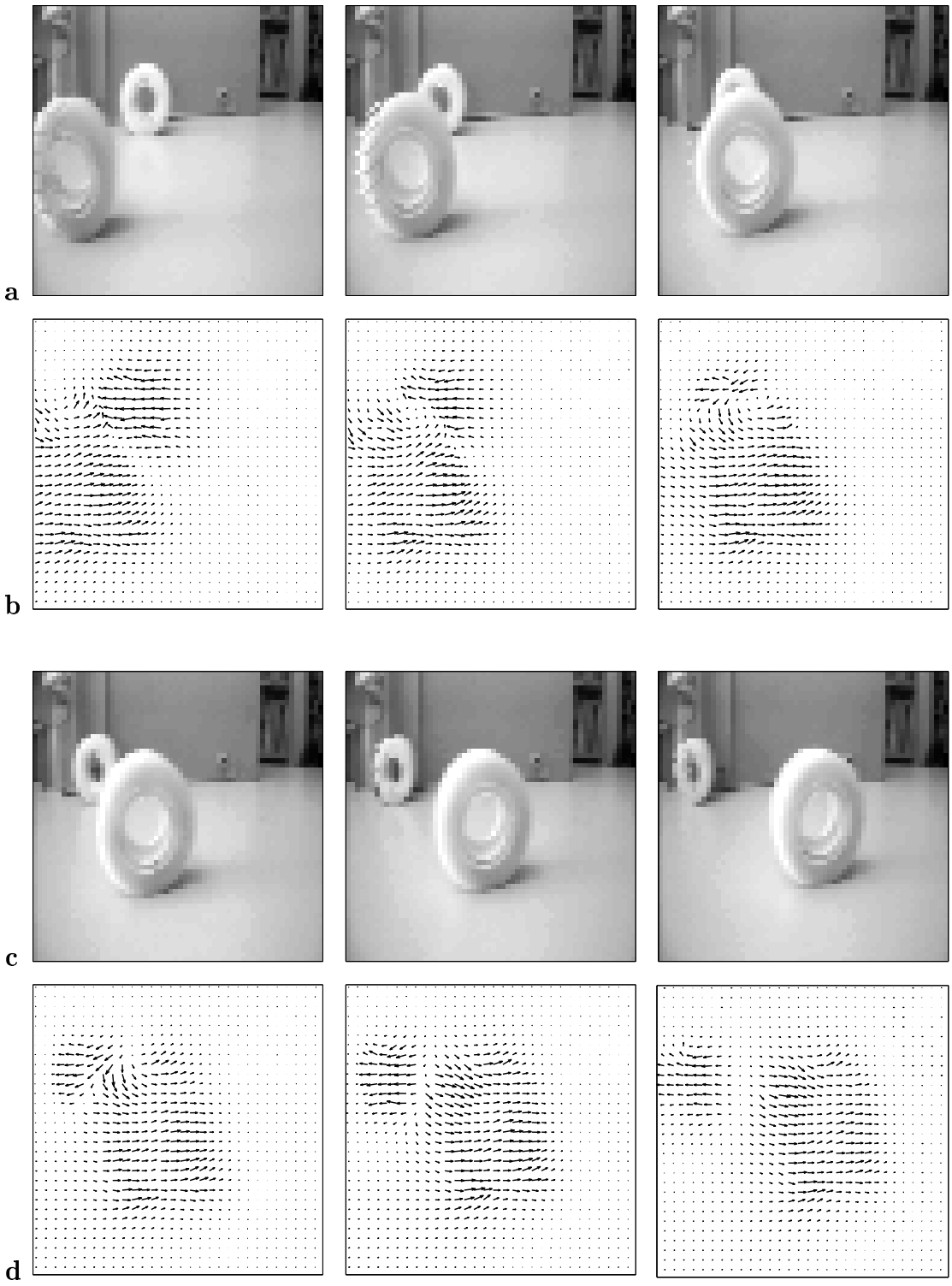


Figure 4.7: *The 'tape-rolls' sequence.* The sequence (a,c) shows two tape-rolls that are rolling in opposite direction on a table. Only every second frame is shown. The estimated flow field (b,d) is down-sampled by a factor of 2. The conductances were $\rho = 0.075$ and $\sigma = 0.0025$.

with the tape-roll also. Furthermore, shadows associated with the rolls induce a visual motion perception as well, *e.g.* the optical flow area at the bottom of the frontal row is elongated in the direction of the shadow. The limitations of the isotropic smoothness constraint become obvious in the case where occlusion occurs and the two flow sources are present within a close neighborhood. As can be seen in Figure 4.7b-d, the partial occlusion of the distant tape roll results in a highly disturbed flow field because the assumption of having locally only a single flow source present is violated.

Performance of the optical flow network

There are online databases that provide image sequences that can be used to test optical flow algorithms[8]. Some of these sequences have become quasi-standards in the literature, and allow at least some qualitative comparison among different approaches. Three of the most frequently used examples were applied to the optical flow network. For each sequence, a particular frame and the associated optical flow estimate is shown (Figure 4.8 and Figure 4.9, respectively). While two of the sequences are natural image recordings, the third one is an artificially rendered image sequence where the projected image velocities are known exactly. Thus, the sequence is often used as a quantitative benchmark for different optical flow algorithms (see *e.g.* [Little and Verri 1989, Barron et al. 1994]). Although we will provide a quantitative error measure for reasons of completeness, again we have to recall that the significance of such numbers is fairly limited. For the three simulation examples, the weight of the bias constraint is kept constant ($\sigma = 0.001$) while the strength of the smoothness constraint is set as indicated.

The first test sequence shows the well-known 'Rubic's cube' placed on a platform that rotates counter-clockwise. The optical flow field reflects correctly the large rotational motion. The moderate smoothing ($\rho = 0.15$) still allows to resolve the local motion ambiguities at the grid pattern of the cube. This example shows that the brightness constraint model, although it accounts only for translational motion, may be suited to estimate large-field rotational motions.

The second sequence is called the 'Hamburg taxi' sequence. It is a typical traffic scene and includes four different motion sources: a dark car crossing from left to right, the taxi turning around the corner, a van passing from right to left behind a tree and finally a pedestrian in the upper left corner walking to the left on the side-walk. Figures 4.8c,d show frame #19 and the corresponding estimate of the optical flow field respectively. As can be seen, the network resolves the different motion sources clearly although the flow field induced by the pedestrian is hardly visible due to subsampling and the low walking speed. The smoothness strength was kept moderate ($\rho = 0.15$). Interestingly, the flow field of the taxi seems to be more homogeneous and correct than the one of the black

[8] see Appendix B for more details

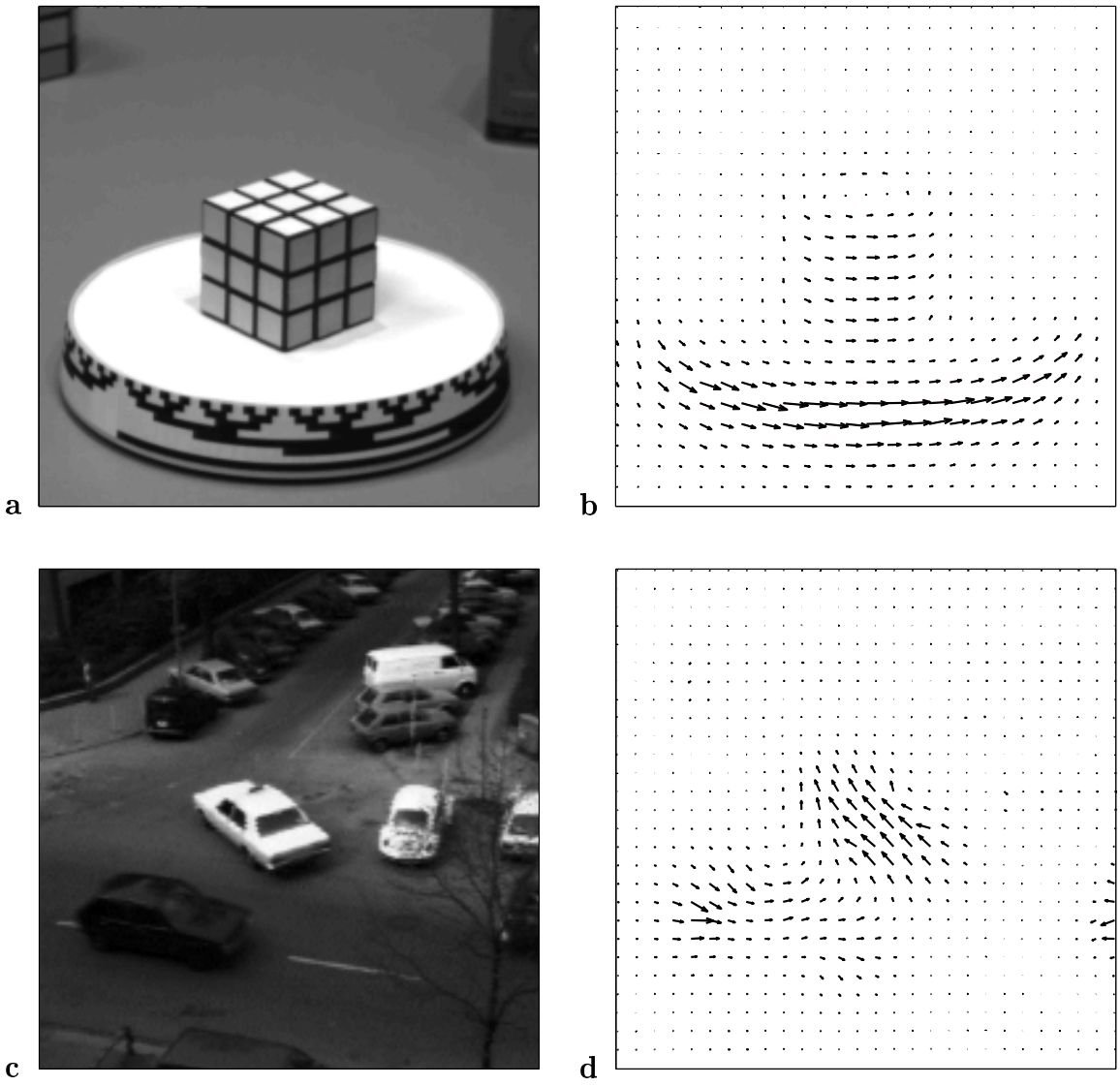


Figure 4.8: *Simulation results on natural test sequences.* (a,b) The 5th frame of the 'Rubic's cube' sequence and the associated optical flow estimate are shown. The result shows that rotational motion on the large is well captured with the translational model. (c,d) The last frame of the 'Hamburg taxi' sequence. Note that the low contrast within the lower car induces the tendency towards a more normal flow estimate. Both flow fields are subsampled by a factor 10.

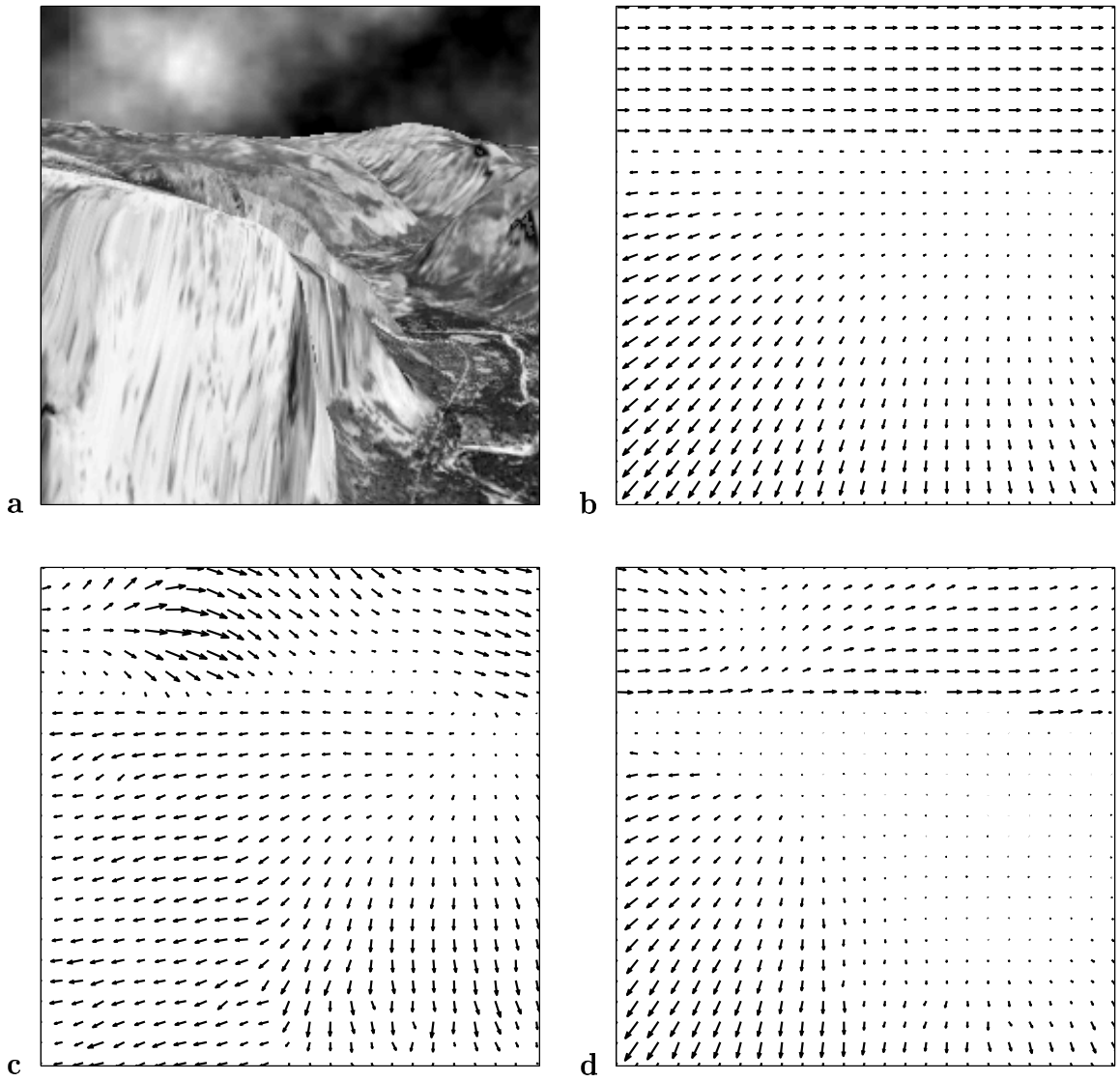


Figure 4.9: *Simulation results for the 'Yosemite' sequence.* (a) Frame #10 of the artificially rendered flight sequence over the Yosemite valley (CA, U.S.A.). (b) The correct underlying 2D image motion $\boldsymbol{v}_c$. (c) The estimated optical flow field $\boldsymbol{v}_e$. (d) The error vector field $\boldsymbol{v}_c - \boldsymbol{v}_e$. Note that especially the estimated flow on the mountain peak in the foreground is affected by the aperture problem in the large and thus provides a substantial directional error. Again, all above flow fields are subsampled by a factor 10.

car in the front. This is mainly due to the low visual contrast condition within the black car. It increases the influence of the bias constraint and results in a tendency of the flow vectors to point towards their normal direction.

The 'Yosemite' sequence (Figure 4.9) represents, in addition to its artificial nature, a different motion pattern. It reflects the visual perception of ego-motion of a moving observer in a mostly static environment. The clouds in the sky undergo translational motion to the right. To best account for the large field motion although it is not translational, the smoothness constraint was assumed to be important ($\rho = 0.4$). Figure 4.9d shows the difference vectors between the correct 2D projection of the simulated motion and the estimated flow field. As can be seen, the largest deviations occur around the mountain peak in the foreground and the sky in the background. The estimate on the area of the mountain peak suffers from the dominantly vertically-oriented texture, thus reflecting the aperture problem in the large. The flow field is biased towards the normal flow orientation which is horizontal. By contrast, the landscape in the foreground on the right-hand side shows a rich and randomly oriented texture and allows a better estimation of the flow field. The second area of large estimation errors is the cloudy sky. Right at the horizon, the error is large because the smoothness constraint does not allow to preserve the motion discontinuity that occurs between the clouds and the mountains. Furthermore, the whole sky area shows regions of large error. This is not surprising, because the brightness constraint is highly violated due to the successive formation and extinction of the cloud patterns.

A quantitative error comparison with a few selected methods reported in the literature shows that the proposed optical flow network is adequate in its performance (Table 4.2). In order to provide a fair comparison, only methods that provide a 100% dense flow field were considered. Any threshold applied to the flow field facilitates the estimation task because it neglects regions of ambiguous or noisy visual motion where an estimate is difficult to achieve and thus the error is likely to be high. Furthermore, for methods with sparse flow fields the reported error measure in the literature [Barron et al. 1994] are only taken over the sparse subset naturally leading to better performance measures.

As expected, the measured error is close to the one for the approach by Horn and Schunck [1981] because the bias constraint is set rather weak and the smoothness parameter $\rho = 0.5$ is equivalent to the value used by Barron et al. [1994]. In addition, large regions of uniform brightness as *e.g.* seen in Figure 4.5d where the bias constraint would be most effective are not present. Nevertheless, the finite σ seems to improve slightly the estimation accuracy by reducing the noise sensitivity in regions of low contrast. The other two methods are explicit matching methods and provided little more accurate estimates for this particular sequence. Again, these numbers have to be interpreted cautiously.

To summarize: The proposed optical flow network provides good results for realistic image sequences. The influence of the smoothness constraint has to be adjusted according

method	$\mathbf{\Phi_c}$ mean	$\mathbf{\Phi_c}$ std
Horn and Schunck [1981]	31.7°	31.2°
Anandan [1989]	13.4°	15.7°
Singh [1991]	10.4°	13.9°
The optical flow network	23.2°	17.3°

Table 4.2: *Quantitative error comparison for the 'Yosemite' sequence.* The mean and the standard deviation for the combined error of the flow estimates are given for different approaches reported in the literature. For reasons of fairness, only methods that provide a 100% dense flow field are considered. The referred error values are taken from Barron et al. [1994] and the applied error measure is defined according to (4.24).

to the expected visual motion. For rather small moving objects in front of a stationary observer, ρ is preferably chosen to be small in order to obtain a reasonable spatial resolution. For large visual motions, in particular induced by ego-motion, an increased ρ results in a more robust and correct estimate.

4.1.6 Related methods

The **motion coherence theory** introduced by Yuille and Grzywacz [1988] (see also [Yuille and Grzywacz 1989, Grzywacz and Yuille 1991] and re-formulated within a Bayesian framework [Weiss and Adelson 1998]) reveals some strong compliance with the minimization problem (4.5) though it introduces a more general description of the smoothness constraint. Yuille and Grzywacz propose the minimization of the following cost function in order to obtain an optimal estimate of the optical flow:

$$H_{GY}(\boldsymbol{v}(\boldsymbol{r}), \boldsymbol{U_i}) = \sum_i (M(\boldsymbol{v}(\boldsymbol{r_i})) - M(\boldsymbol{U_i}))^2 + \lambda \int \mathrm{d}\boldsymbol{r} \sum_{m=0}^{\infty} c_m (D^m \boldsymbol{v})^2. \tag{4.25}$$

According to the original description in Yuille and Grzywacz [1988], the first term of the cost function represents the difference between the local measurement of visual motion $M(\boldsymbol{U_i})$, where U_i is the *true* image motion, and the local estimate of the constructed smooth flow field $M(\boldsymbol{v}(\boldsymbol{r_i}))$, summed over all locations r_i where visual input is available. This formulation follows strongly the formalism of standard regularization, where the regularized solution is required to be close to the data. However, as noted before, there is nothing like a true image motion that can be measured locally in visual motion processing. Rather, we have to understand this first term of (4.25) as a total measure of how well the estimated smooth optical flow fits the matching model applied at each location. For example, using an implicit matching model, the brightness constraint (4.1) would represent such a measure.

The second term is a description of a general smoothness constraint that requires the derivatives ($D^m \boldsymbol{v}$, $m = [0 \ldots \infty]$) of the optical flow field of any order m to be small. This is interesting because it includes the zero order derivatives which can be seen as the formulation of the bias constraint (4.4) with the particular reference motion $(u_0, v_0) = \mathbf{0}$. One might therefore conclude that motion coherence theory basically covers the optimization method proposed in this chapter. Nevertheless, it constitutes a much broader and more general smoothness constraint framework for optical flow estimation and gives reason to the following critical objections:

One theorem of the motion coherence theory claims that *"A necessary and sufficient condition (to be well-posed) is that derivatives of higher than first order exist in the smoothing operator."*(Grzywacz and Yuille [1991], page 243, theorem 4). We have shown that this theorem is incorrect as it stands. Actually higher order derivatives in the smoothing constraint **alone** are not a sufficient and necessary condition for well-posedness. Part of the reason for this strong statement might be the fact that Yuille and Grzywacz did not prove the theorem themselves but rather refer to the literature [Duchon 1977]. It is remarkable however, that Duchon neither proves directly the theorem nor does he relate to the general smoothness term. Duchon's statement on which Yuille and Grzywacz base their theorem refers to second order or so-called *thin plate models* in higher spatial dimension $\mathbb{R}^n$ for which he requires the cost function to include derivatives of order m, where $m > n/2$ in order to apply his proposed spline methods. Duchon does not consider first order models such as the membrane model.

Higher than first order derivatives (in conjunction with the zero order terms) are necessary to guarantee a unique solution if the visual motion information is only available at sparse locations[9]. However, under the assumption of a continuously differentiable brightness distribution $E(x, y, t)$ the constraint problem (4.5) is guaranteed to be well-posed. As we have seen, this is true even when the smoothness constraint is completely neglected ($\rho = 0$). Subsequently, the claim that the approach of Horn and Schunck [1981] is not well-posed due to the lack of higher order derivatives in the smoothing term is wrong. In fact, the Horn and Schunck approach is ill-posed but only because it does not contain the bias constraint.

Although the formulation of a general smoothness constraint is mathematically elegant, it is much too broad with respect to reasonable simple models of the flow field. For almost all known models, there is no physical relevance to account for smoothness in higher than second order derivatives of the optical flow field. Not surprisingly, Yuille and Grzywacz did not find any significance for higher than second order derivatives in their simulation results. In particular, pure translational models only require smoothness in the first order derivatives per definition[10]. On one hand, the general smoothness constraint seems to

[9] personal communication, A. Yuille

[10] see Chapter 2

be too general in terms of higher order derivatives, it is too limited for the zero order derivatives. In contrast to the bias constraint (4.4) it assumes *a priori* zero reference motion. Thus, it discards the possibility to apply a bias to the flow field towards another, non-zero reference motion that can be the result of some adaptive process.

Last, Yuille and Grzywacz did not show how to implement their method in a network architecture. It seems that although they proposed a general mathematical formalism of smoothness in optical flow estimation, they were not completely aware of the physical nature and the relevance of such a constraint. By contrast, the smoothness and bias constraints introduced here are founded on the physics of the applied matching model and the functional needs of a simple and efficient network architecture for optical flow estimation in an unconditioned real-world environment.

4.2 Control of the Local Conductance Pattern

There is a strong dependence of the optical flow estimate on the values of the parameters ρ and σ. As yet, we assumed these parameters to be uniform and constant over the entire network. The resulting isotropic diffusion process leads to uniform neighborhood interactions that are not directed by any local information of possible object boundaries. However, if such information is available, we can significantly improve the optical flow estimation by making local changes to the lateral connectivity strength ρ such that motion integration takes place preferably over regions containing single flow sources. The trade-off between loosing information about the locality of motion and gaining more accurate and robust estimation results through spatial integration can be circumvented. Especially, smoothing over motion discontinuities as exhibited in the previous simulation examples can be suppressed.

Ideally, we prefer to partition the image scene into disjoint regions of common (translational) motion (see also Figure 2.3) by introducing **line processes** that literally cut ($\rho = 0$) the lateral conductances in the resistive network around these areas of single flow sources. We call such process **motion segmentation** if it is mainly a function of the present visual motion information. The resulting common motion within each partition can thereby be seen as the individual labels of the segments. Line processes allow an efficient optical flow estimation using a single flow model: The optical flow network serves as the common computational basis, whereas the effective collective computation is performed simultaneously for several different motion sources, each located within an isolated region of the visual space. The connectivity pattern serves thereby two-fold: It improves the optical flow estimate, but at the same time it represents the location of motion boundaries that could be used as input for other visual or cognitive systems.

Furthermore, also σ can be changed locally. A high σ literally clamps the potentials in the resistive networks to the reference motion $\boldsymbol{v}_0$ by shunting the currents at each node to

the reference potential. It suppresses the optical flow response in selected image regions and therefore provides a means of **selective attention**. Such attentional processes have been shown to play a significant role in cortical motion processing of primates. For example, already in the early stages of the motion pathway of the macaque monkey such as MT and MST, it has been shown that spatial attention results in an increased neural response to stimuli at the attended location and a reduced response to stimuli elsewhere [Treue and Maunsell 1996, Treue and Trujillo 1999]. Selective attention is believed to be a mechanism to reduce the huge amount of visual information; and it accounts for the limited processing capacity of cortical areas observed in humans that were tested for their visual awareness of objects [Kastner and Ungerleider 2000].

We now rewrite the optimization problem (4.5) in a more general form

$$H(\boldsymbol{v}; \rho(x,y,t), \sigma(x,y,t)) = \int_\Omega F^2 + \rho(x,y,t)S + \sigma(x,y,t)B \ \mathrm{d}\Omega = \min\,! \tag{4.26}$$

There is a unique solution along trajectories $\rho(x,y,t)$ and $\sigma(x,y,t)$ for which the integrand is strictly convex at any time t. However, global asymptotic stability is not guaranteed because the global minimum for every *given* $\rho(x,y,t) \geq 0$ and $\sigma(x,y,t) > 0$ depends on the history of $\rho(x,y,t)$ and $\sigma(x,y,t)$. This can lead to hysteretic behavior of the motion estimate because motion is *per se* a process in time and has a history, of course.

In the following we discuss possibilities of how to control the weighting parameter ρ and σ locally in order to enhance optical flow estimation and perform higher level visual motion tasks such as motion segmentation and object recognition by motion. In particular, we present two elementary network architectures that contain additional retinotopically organized maps of computational units that receive input from the optical flow network and recurrently control the weighting parameters $\rho(x,y,t)$ and $\sigma(x,y,t)$. These networks demonstrate the potential and flexibility of the optical flow network approach as a basic computational machinery for visual motion perception. Nevertheless, we have to bear in mind that higher level visual motion processes as performed by *e.g.* humans are based on rich and rather complex cortical mechanisms that are not at all understood yet. The presented network architectures therefore are not meant to reflect biophysical models of visual motion processing.

4.2.1 Passive non-linear conductances

Previously, we have identified the weighting parameters ρ and σ as vertical and lateral conductances in the resistive networks of the optical flow network architecture. In a first instance, we consider now these conductances as typical two-terminal devices that control the amount of current flowing through as a function of the local potential differences V across them. This way, the general and explicit space- and time-dependent formulation

in (4.26) reduces to an implicit one, thus $\rho(V)$, $\sigma(V)$ respectively. These *passive conductances* remain (non-linear) functions of only local measures which allows us to derive local rules and architectures for their implementation.

Applying passive conductances, the smoothness constraint (4.13) in the discrete case transforms to

$$S(\Delta u_{ij}, \Delta v_{ij}) = \int_0^{\Delta u_{ij}^x} I(V)\,\mathrm{d}V + \int_0^{\Delta u_{ij}^y} I(V)\,\mathrm{d}V + \int_0^{\Delta v_{ij}^x} I(V)\,\mathrm{d}V + \int_0^{\Delta u_{ij}^y} I(V)\,\mathrm{d}V, \qquad (4.27)$$

where $I(V)$ is the current flowing at node ij in both orthogonal directions in each of the two resistive networks. The integrals over the currents are equivalent to the co-content and their sum represents the total dissipated power in the networks.

Global asymptotic stability is guaranteed if each integral term in the smoothness constraint (4.27) is convex with respect to V. Let $I(V) = Vg(V)$ be the current through the conductance $g(V)$. Then, convexity is given if the second derivative of the co-content, called the *incremental conductance* g^* is non-zero, thus

$$g^* = \frac{\mathrm{d}I(V)}{\mathrm{d}V} = g(V) + V\frac{dg(V)}{dV} \neq 0. \qquad (4.28)$$

Global asymptotic stability of the optical flow network is ensured if the incremental conductances ρ^* and σ^* at each node ij are non-negative or non-zero respectively.

The lateral conductance $\rho(V)$

What would be good functions $\rho(V)$ and $\sigma(V)$ in order to increase the performance of the optical flow estimation? The lateral conductances preferably should be high at places where the potential difference V and thus the variation of the flow field is small and low if it is large. That way, small variations in the optical flow field are smoothed out whereas large velocity gradients at motion discontinuities are preserved by limiting the current. Or in other words, the smoothness constraint is modified such that it penalizes large velocity gradients less than small ones.

We distinguish two classes of implementations of such conductances. The first class describes so-called **saturating resistances** for which the *Horizontal Resistor (HRes)* [Mead 1989, Mahowald and Mead 1991] and the *Tiny-Tanh Resistor* [Harris 1991] represent successful examples of appropriate electronic circuits. The current is typically a sigmoidal function of the potential V across, thus for example given as

$$I(V) = I_0 \tanh(\alpha V)\ . \qquad (4.29)$$

I_0 is the maximal saturation current of the conductance and $\alpha = \rho_0/I_0$ a gain factor, where $1/\rho_0$ represents the maximal possible conductance, thus the maximal slope of (4.29).

Figure 4.10 shows the I-V curve, the co-content, the incremental and the effective conductances of saturating resistances. We note that the co-content, and thus the smoothness constraint, transforms from a quadratic to a linear function of the voltage gradients. The transition point depends on the parameter ρ_0 for a given I_0: For high values, as in the tiny-tanh circuit, saturation occurs for very small voltages, thus the current approximates a step function, going from $-I_0$ for $V < 0$ to I_0 for $V > 0$ and the co-content turns into an absolute value function $|I_0 V|$. Absolute value functions have been suggested for solving *exact constraint*[11] problems in image processing [Platt 1989]. Using saturating resistances, global asymptotic stability is given because the incremental conductance $\rho^* \geq 0$.

The second class of implementations contains so-called **resistive fuse** circuits [Harris and Koch 1989, Yu et al. 1992, Sawaji et al. 1998]. As for the saturating resistances, these circuits allow smoothing in regions of small voltage variations across their terminals. Above some voltage difference, however, the effective conductance decreases and finally goes to zero, ending up in a complete electrical separation of the two terminals. Such behavior serves well to implement line processes necessary to separate distinct regions of common motion *completely*. The incremental conductance of any resistive fuse is partially negative (see Figure 4.10), meaning that networks of such conductances are multi-stable. In fact, multi-stability cannot be avoided in any networks performing line processes.

Figure 4.11 allows a comparison between the behavior of the optical flow network using ohmic or non-linear passive conductances. Using saturating resistances or resistive fuses reduces significantly the smoothness of the resulting optical flow field at motion discontinuities. It prevents an extensive blurring of the motion contours but still allows the interaction within areas of common motion.

However, there is a general problem of applying non-linear two-terminal conductances in the optical flow network: Optical flow is a two-dimensional feature encoded in two resistive networks, one for each component of the flow vectors. Obviously, such conductances would regulate the connectivity in each network independently, according to local differences in the components. A difference measure in each of the components can only approximate an appropriate measure for the complete features such as *e.g.* the difference in absolute vector length. Also, we assume that the conductance pattern reflects the physical properties of the moving scene such as boundaries of moving objects. Since the physical properties are only dependent on the image location they should be spatially coherent within both networks which is not guaranteed with passive non-linear conductances.

[11] *Exact* because the linear dependence of the co-content on V penalizes small deviations more strongly than in the quadratic case and thus forces the solution to be piece-wise smooth and very close to the data.

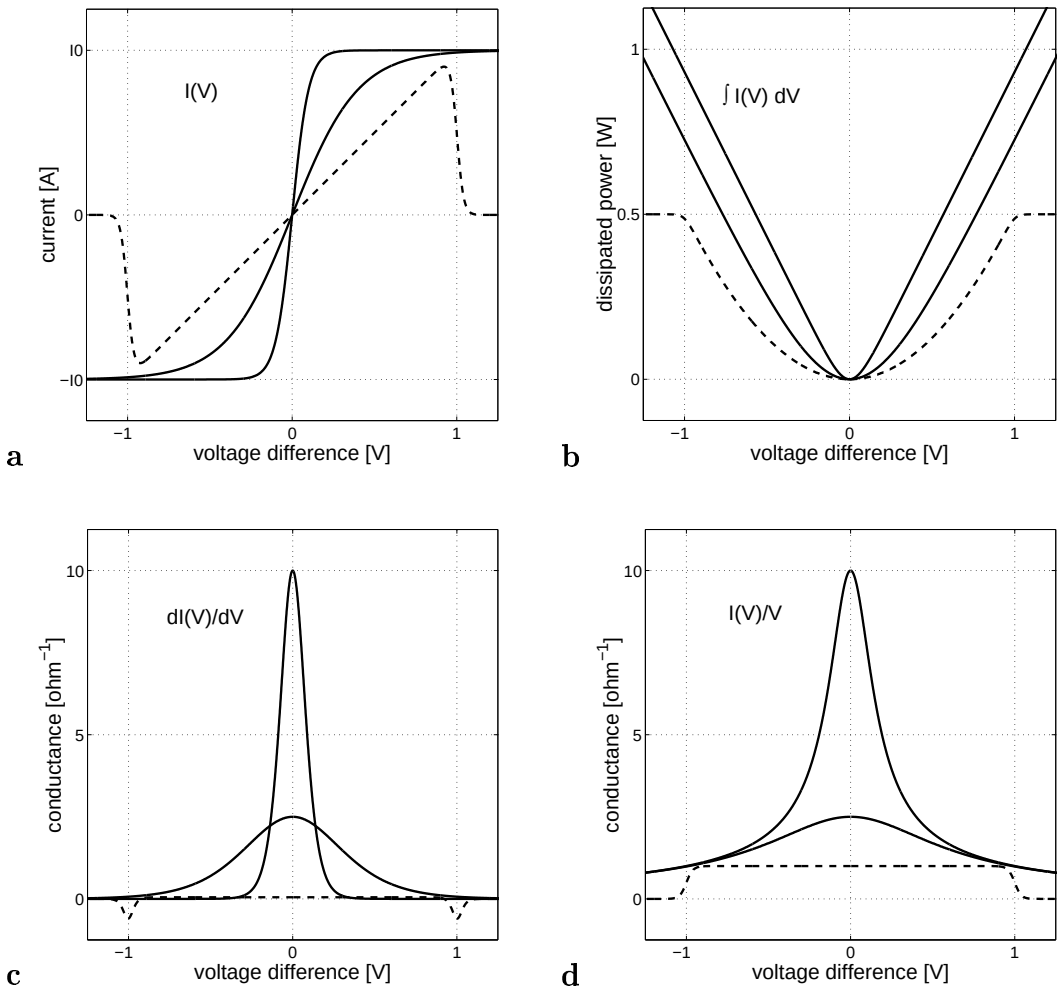


Figure 4.10: *Characteristics of passive conductances.* The characteristic curves for different saturating resistances ($\rho_0 = [2.5, 10]$ Ω) and a typical analog resistive fuse (dashed lines) are shown. (a) The current through the saturating resistances is clearly limited for high voltage differences. Furthermore, the resistive fuse electrically separates the two terminal for high voltage differences completely. The current for this resistive fuse model is described as $I(V) = \lambda V[1 + \exp(\beta(\lambda V^2 - \alpha))]^{-1}$ with $\lambda = 1\ \Omega^{-1}$, $\alpha = 1$ VA and the free temperature dependent parameter $\beta = 25$ [Harris et al. 1990]. (b) The associated co-contents for the saturating resistances show clearly the transition between the quadratic and linear characteristics as the voltage difference increases. For large ρ_0 the co-contents approximate the absolute-value function. The co-content for the resistive fuse is non-convex which is reflected also by the partially negative incremental conductances (c) in the regions of ± 1 V. (d) Nevertheless, the effective conductances are strictly positive for both classes.

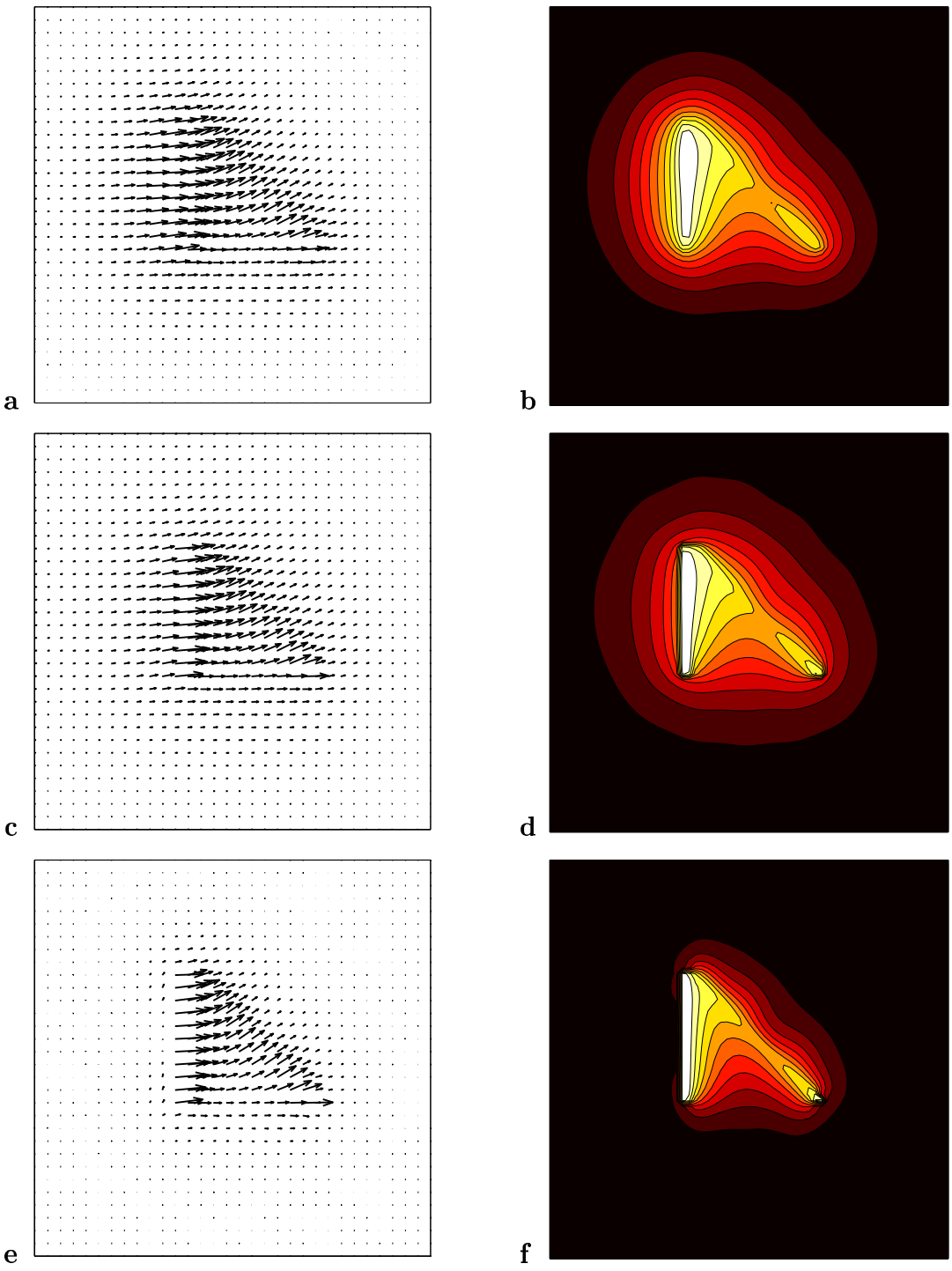


Figure 4.11: *Optical flow estimation using passive non-linear conductances.* (a) A linear smoothness conductance (equivalent to Figure 4.4d, with $\rho = 0.15, \sigma = 0.001$) generates a smooth flow field which becomes particularly well visible in the contour plot (b) of the absolute velocities. (c,d) Saturating resistances ($\rho_0 = 0.15, I_0 = 0.01$) partially prevent smoothing into the background. (e,f) Resistive fuses ($\lambda = 0.15\ \Omega^{-1}, \beta = 200$ and $\alpha = 0.025$ VA) clearly separate the units in the network at high motion gradients.

The bias conductance $\sigma(V)$

The quadratic measure of the bias constraint (4.4) results in a constant conductance σ and therefore an ohmic bias current in the resistive network. The optical flow estimate is penalized *proportionally* to the size of its components. However, the original purpose of the bias constraint was to ensure that the optimization problem is well-defined in cases where the visual input is ambiguous. It seems reasonable to strengthen the bias constraint locally according to the degree of ambiguity. Only, the degree of ambiguity is not at all reflected by the amplitude of the motion estimate which is the only information available for any potential passive conductance $\sigma(V)$. Therefore, it is most sensible to penalized the motion estimate *independently* of its size, which is achieved by applying a constant bias current. As shown earlier, such behavior can be implemented approximately using saturating resistances with $\rho_0 \gg 1$. Again, the bias constraint (4.4) must be replaced by the measure of the electrical co-content:

$$B(u_{ij}, v_{ij}; u_0, v_0) = \int_0^{|u_{ij}-u_0|} I(V)\,\mathrm{d}V + \int_0^{|v_{ij}-v_0|} I(V)\,\mathrm{d}V, \tag{4.30}$$

with $I(V)$ as in (4.29). For $\rho_0 \to \infty$, the bias constraint (4.30) reduces to the absolute-value function. Global asymptotic stability is guaranteed because the incremental conductance σ^* is non-zero for all finite ρ_0.

4.2.2 The motion segmentation network

An extended motion segmentation system is introduced that consists of two discontinuity networks recurrently connected to the optical flow network (see Figure 4.12). Each unit P_{ij} and Q_{ij} controls a motion discontinuity segment oriented in x- and y-direction respectively at node ij in the optical flow network. The whole system forms a recurrent feedback-loop where the optical flow network provides the motion input to the discontinuity networks which then control the lateral conductances of the former. In contrast to the passive conductances discussed previously, the connectivity strength in both resistive networks is now controlled identically. Since motion discontinuities strongly reflect physical properties of the underlying visual scene, clearly they either exist or not. Thus, the units in the discontinuity networks ideally approximate a binary behavior to reflect the decisive character of detecting discontinuities, but being 'sufficiently' analog to allow a deterministic implementation. One can also think of switches that enable or disable the connection between neighboring units in the optical flow network. Each open switch thereby represents a line segment in the overall line process which is perpendicularly oriented to the pair of connected units.

Again, the desired behavior of the discontinuity networks is formulated as a constraint optimization problem. Since both networks are identical except in orientation we restrict

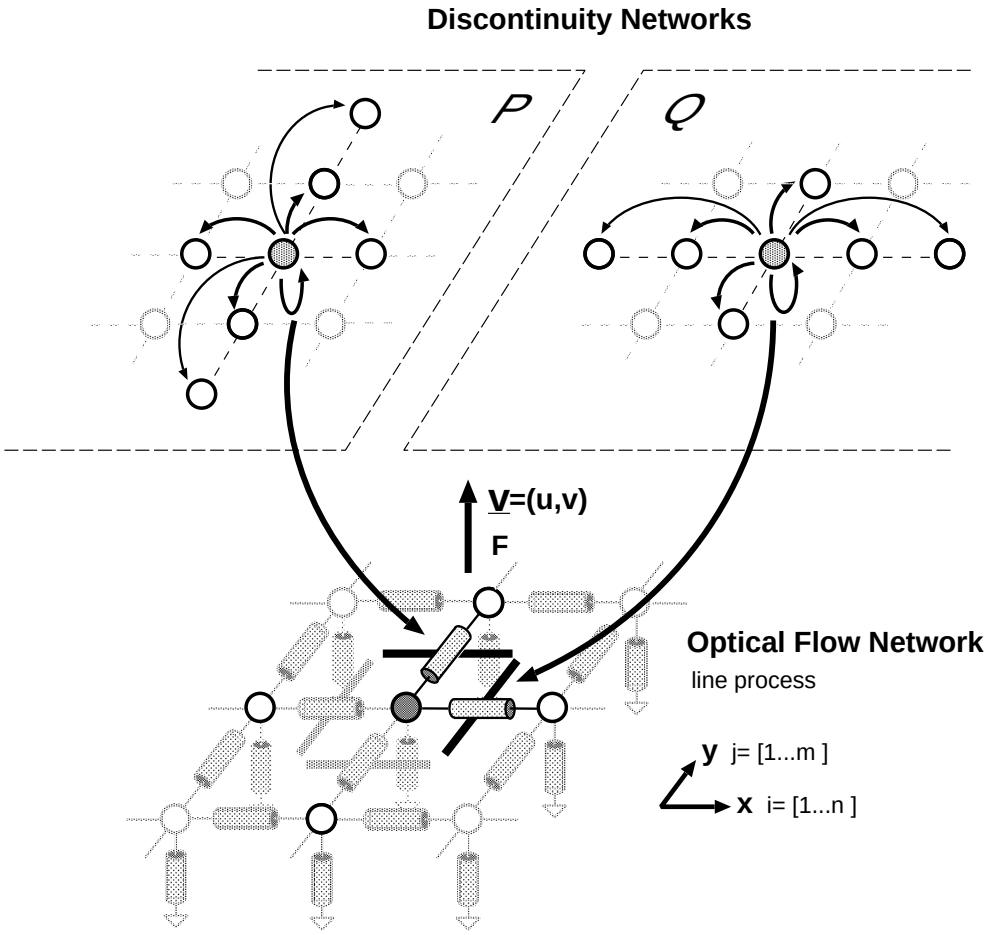


Figure 4.12: *Schematics of the motion segmentation network.* Two additional networks control the lateral conductances in the optical flow network in x- and y-direction respectively. The units in each network excite their neighbors along the orientation of the corresponding line segment to force a continuation of the line process. The soft-WTA competition supports the continuation by suppressing activity in the units orthogonal to the line orientation. Furthermore, it adapts the threshold of discontinuity detection by forcing a winner also in areas of low input contrast. For clarity, the optical flow network is shown only as a single layer network.

the analytical discussion to the 'P-unit' network and silently address the 'Q-unit' network as well. The following cost function is proposed:

$$H_P = \sum_{ij}\Big(\alpha P_{ij} + \beta(1-P_{ij})(\Delta \boldsymbol{v}^y_{ij})^2 + \gamma(1-P_{ij})(\Delta F^y_{ij})^2 - \frac{\delta}{2}P_{ij}(P_{i+1,j}+P_{i-1,j}) + \frac{\epsilon}{2}P_{ij}\sum_{w_0 \notin W} P_W + \frac{\phi}{2}(\sum_W P_W - 1)^2 + 1/R\int_{1/2}^{P_{ij}} g^{-1}(\xi)\,\mathrm{d}\xi\Big). \quad (4.31)$$

The first three terms represent a typical threshold function where the unit P_{ij} is turned on, if the sum of the weighted input is larger than a constant α. The input consists of the square of the local optical flow gradient $(\Delta \boldsymbol{v}^y_{ij})^2 \equiv (\Delta u^y_{ij})^2 + (\Delta v^y_{ij})^2$ and the *brightness constraint* gradient in y-direction $(\Delta F^y_{ij})^2$, where F is defined as in (4.1).

Where most known approaches for motion segmentation use the flow field gradient as the measure of motion coherence[12], using the local gradient of the brightness constraint in addition is very specific to the chosen optical flow network approach and has not been proposed before. The major advantage of this combined difference measure is its relative insensitivity to the strength of the smoothing conductance. Figure 4.13 characteristically shows the smooth optical flow field for weak and strong smoothing at an idealized motion discontinuity. Since the brightness constraint represents the weighted distance of the velocity estimate to the brightness constraint line, its local gradient is largest at motion discontinuities because there, two different motion sources collide and have to find a common flow estimate if smoothing takes place. However, if smoothing is weak or even disabled, the brightness constraint can be fulfilled much better on both sides of the discontinuity and thus its gradient is shallow. For the gradient in the flow field, the conditions are just opposite. Thus, whatever the smoothness conductance pattern in the optical flow field is, the information about potential motion discontinuities cannot 'get lost' using both gradient measures. Of course, the input of the two gradient measures must not – and cannot due to non-local interactions – be balanced perfectly in order to obtain a net input difference when line-segments change their state and thus the smoothness conductance changes rapidly. A net input is essential to perform the optimization task correctly because it relates the input to the state of the networks. Nevertheless, the combined gradient input is a useful way to reduce the expected hysteretic effects due to the performed line process.

The first three terms of the cost function (4.31) do not yet define a network structure because they do not imply any direct interaction between individual units. This situation changes with the next constraint (δ): It favors those units to be active that have already active nearest-neighbors in the same direction as the orientation of their line segments. The idea is to force the formation of a continuous discontinuity preferably along its orientation. The next two terms support this process by inducing a **soft-WTA** competition

[12] see Chapter 2

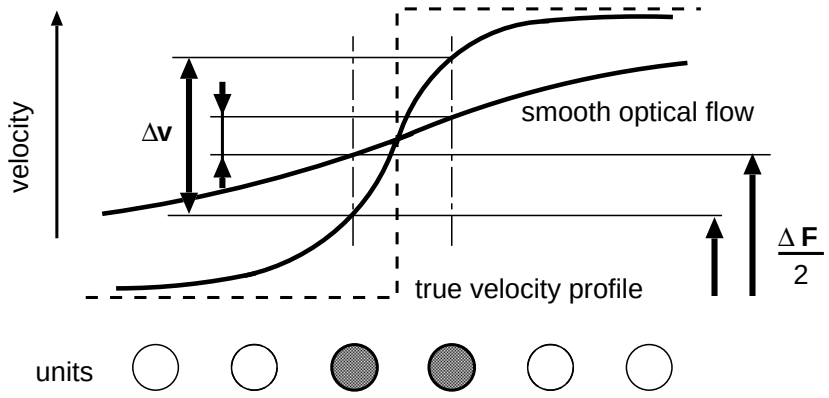


Figure 4.13: *Combined gradient measure for the detection of motion discontinuities.* An idealized motion discontinuity is shown with its true one-dimensional velocity profile (dashed line) and two examples of smooth flow estimates (bold curves). The local gradient of the brightness constraint signals how strong the local motion estimate differs from the set of possible motions defined by the observed spatiotemporal brightness pattern. It is naturally large at motion discontinuities and depends on the smoothness conductance. As indicated by the arrows, the brightness constraint gradient ΔF across the discontinuity is large if the flow field is smooth and small if it is not. However, the optical flow gradient $\Delta \boldsymbol{v}$ just shows the opposite behavior. A combined measure thus signals motion discontinuities independently of the smoothness conditions in the optical flow network and therefore independently of the conductances pattern.

along the units perpendicular to the orientation of the line segment (see Figure 4.12). It inhibits activity in a row of units perpendicular to the orientation of the unit's particular line segment which suppresses the formation of many, very close and parallel oriented line segments. The constraints for the soft-WTA behavior are formulated identically to the simple WTA example[13], except that competition is now spatially limited to the extent of the (one-dimensional) neighborhood kernel W and modulated by the local kernel coefficients. The coefficients are chosen to decay exponentially with distance and are normalized such that their sum is equal to one. The kernel is assumed to be symmetric where w_0 is its central coefficient. The soft-WTA competition also provides means for an *adaptive threshold* because it forces a winner also in regions of low input contrast. The tendency to force discontinuities even when no input is present is counter-balanced by the static threshold constraints.

The last constraint in (4.31) is imposed again by the non-zero leak conductance of

[13]see Equation (3.6), Chapter 3

the analog units with activation function $g : p_{ij} \rightarrow P_{ij} \in [0, 1]$. Although chosen to be a typical sigmoidal function (3.5) it can be any function that is non-linear, monotonically increasing and limited to $[0, 1]$.

The cost function (4.31) is bounded from below for any given and finite input. Thus a network that performs gradient descent on the cost function is given to be asymptotically stable. This leads to the following dynamics for the discontinuity networks:

$$\begin{aligned} \dot{p}_{ij} = \; - \; & \frac{1}{C} \Big[\underbrace{\frac{p_{ij}}{R} + \alpha - \beta(\Delta \boldsymbol{v}_y)^2 - \gamma(\Delta F_y)^2}_{\text{static threshold}} \\ & - \underbrace{\delta(P_{i+1,j} + P_{i-1,j}) + \epsilon \, \overline{P_{ij}} + \phi \, \overline{\overline{P_{ij}}} - \epsilon \, w_0 P_{ij} - \phi}_{\text{soft-WTA with cross-excitation}} \Big] , \end{aligned} \tag{4.32}$$

where $\overline{P_{ij}} = \sum_W w_W P_W$ stands for the weighted average of the output activity within the neighborhood W and $\overline{\overline{P_{ij}}} = \sum_W w_W \overline{P_W}$ for its average. The network architecture follows directly from the dynamics. Again, connectivity can be reduced significantly by introducing two layers of additional inhibitory units. The first layer receives weighted input from all units within a single neighborhood and thus provides $\overline{P_{ij}}$. The second layer does the same but receives now input from the first layer and therefore provides $\overline{\overline{P_{ij}}}$. Besides inhibitory input, each unit also gets excitatory input of different sources: a general and constant input ϕ, input from its neighbors in the orientation of its line segment and excitatory input from itself according to the parameter ϵ and w_0.

Global asymptotic stability for the discontinuity networks is desirable, because the motion input changes and does not favor a hysteretic behavior. It is the weight connection matrix of the network thus the parameters δ, ϵ, ϕ and W that decide its stability properties. Strict convexity of the cost function was easily be verified for the (simple) classical WTA network by requiring the Hessian of the cost function to be positive definite. However, such procedure becomes complex for the discontinuity networks, and may only be solved numerically. Nevertheless, to get some estimate of what the conditions for strict convexity and thus for global asymptotic stability may be, the eigenvector analysis was performed for two extreme cases of W:

- $W = [1]$
  The kernel includes only the local unit itself. The excitatory and inhibitory ϵ-terms in (4.32) cancel and the smallest eigenvalue leads to the condition

$$\frac{p_0}{R} + \phi - \sqrt{2}\delta \geq 0 \tag{4.33}$$

  to hold for global asymptotic stability where $1/p_0$ is the maximal gain of the activation function g. Clearly, since the excitatory and inhibitory ϵ-terms in (4.32) cancel, only the neighbor excitation term remains that has to be smaller than the inhibition imposed by the leak (driving force) p_0/R and the (self-)inhibition ϕ.

- $W = [w_0, \ldots, w_N]$, where $w_i = 1/N \quad \forall\, i \in [1, N]$
  The kernel now includes a complete row (or column) of N units with uniform weights which changes the characteristics to be a hard-WTA along that row or column respectively. Therefore, the eigenvalue condition

$$\frac{p_0}{R} - \frac{\epsilon}{N} - \sqrt{2}\delta \geq 0 \tag{4.34}$$

  is equivalent to the self-excitation gain limit (Equation 3.12 on page 32) plus it also includes the excitation from the nearest neighbors. The row length N appears in (4.34) since we required the weights to be normalized.

General kernels will lead to more complex conditions including every weight parameter. Numerical experiments however showed that condition (4.34) under the rigorous assumption $N = 1$ represents a conservative lower limit for which global asymptotic stability holds. Increasing the inhibitory weight parameter ϕ always exhibits a positive effect on this limit, thus allows stronger excitatory connections ϵ and δ. There are theorems on global asymptotic stability in additive networks based on simple conditions on the inhibitory and excitatory strengths [Hirsch 1989]. However, tested on the hard-WTA network, the theorems were not able to predict a reasonable stability criterion. Reasons for this failure are not investigated further here.

Closing the recurrent loop, we allow the discontinuity networks to control the smoothness conductances in the optical flow network. Hence, we replace the smoothness constraint in (4.13) by

$$S = \sum_{ij} \rho_0 \; \left[(1 - P_{ij}) \; ((\Delta u_{ij}^y)^2 + (\Delta v_{ij}^y)^2) + (1 - Q_{ij}) \; ((\Delta u_{ij}^x)^2 + (\Delta v_{ij}^x)^2)\right], \tag{4.35}$$

where $\Delta u_{ij}^x, \Delta v_{ij}^x$ and $\Delta u_{ij}^y, \Delta v_{ij}^y$ are the linearized gradients of the optical flow field in x- and y-direction respectively. The effective conductance is within the range $[0 \ldots \rho_0]$.

The complete system (discontinuity networks plus optical flow network) is asymptotically stable because each sub-network is. However, global asymptotic stability is difficult to investigate. It is clear, that the interaction between the optical flow and the discontinuity networks is non-linear due to the multiplicative interaction between the different state variables as seen in Equation (4.35). Any statements on additive networks are therefore not applicable. Qualitatively, the combined system is expected to be multi-stable because the feedback loop has certainly a positive component.

Simulation results

The behavior of the complete system was simulated and its suitability demonstrated using the 'tape-rolls' sequence. The parameters of the two discontinuity networks were set as follows: The static threshold $\alpha = 0.12$, the input weights $\beta = \gamma = 1$, neighbor excitation

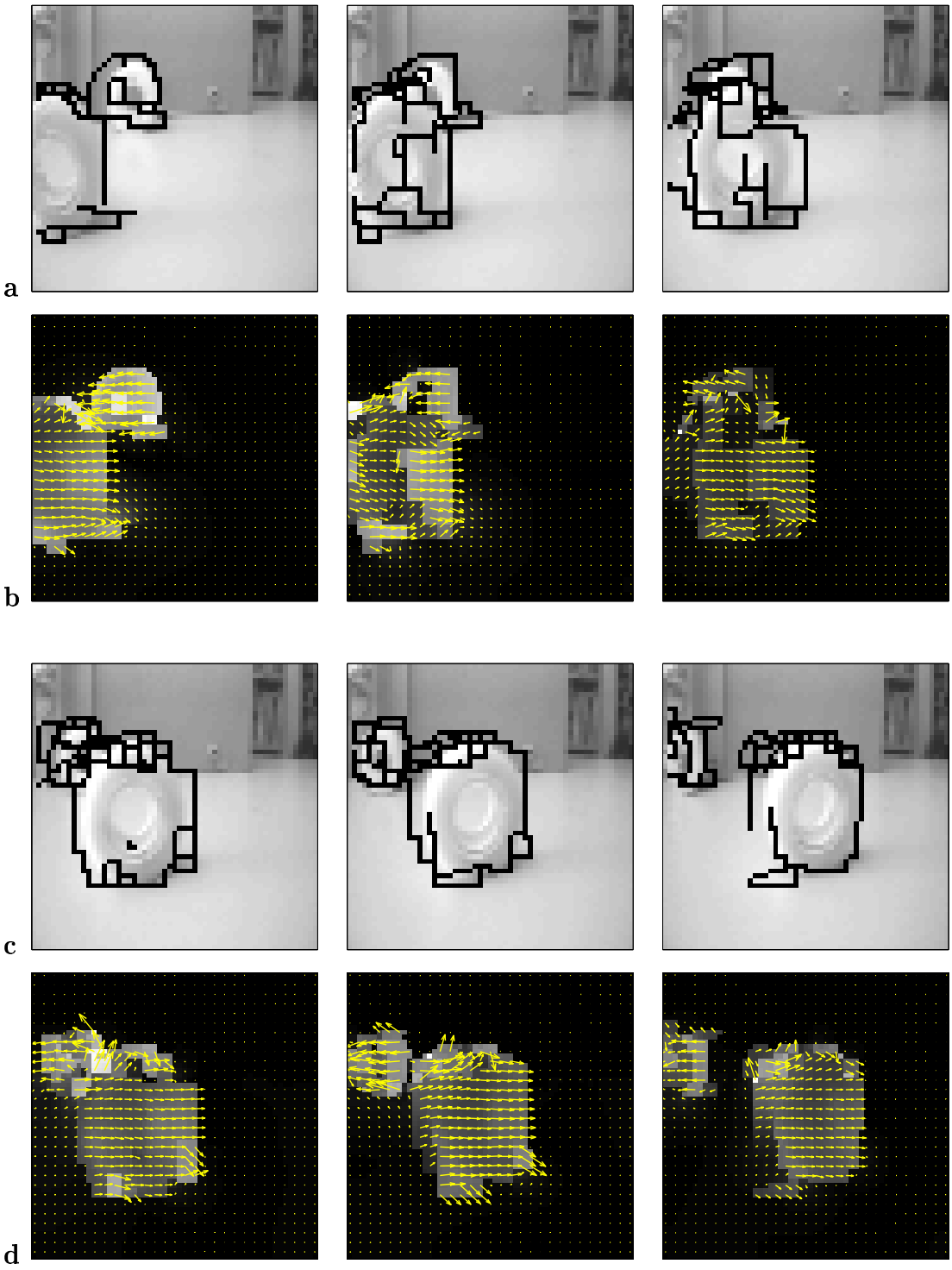


Figure 4.14: *Motion segmentation of the 'tape-rolls' sequence.* (a,c) Every second frame of the 'tape rolls' sequence is shown overlaid with the estimated motion discontinuities. (b,d) The associated optical flow is displayed as vector field and gray scale image of its absolute value. See text for further explanations.

$\delta = 0.002$, the soft-WTA parameter $\epsilon = 0.003$ and $\phi = 0.13$. Self-inhibition was set rather small with p_0/R, resp. $q_0/R = 0.01$. Clearly, the parameter values obey the excitation gain limit (4.34). The bias conductance and the reference motion of the optical flow network were kept as in previous simulations ($\sigma = 0.0025$, $\boldsymbol{v}_0 = \mathbf{0}$).

The 'tape-rolls' sequence is a true benchmark because it includes many features that are difficult to resolve[14]. Figure 4.14a,c shows every second frame (# 2, 4, 6, 8, 10) of the sequence, overlaid with the estimated motion discontinuities. The associated optical flow is shown as flow field superimposed on the gray scale representation of its absolute value (Figure 4.14b,d). We recall again, that the assessment of the simulation results has to remain on a qualitative and descriptive level because absolute benchmarks are not available. Any comparison with other approaches can therefore be, at most, suggestive.

The overall performance is reasonable especially given the fact that there was **no applied reset** of the network states in between different frame presentations. A reset clearly would help to improve the segmentation quality because it eludes the problem of hysteresis. Unfortunately, most of the proposed approaches in the literature do not report any simulation results of whole image sequences with continuous frame presentation [Hutchinson et al. 1988, Harris et al. 1990, Memin and Perez 1998, Aubert et al. 1999, Cesmeli and Wang 2000]. Thus, it seems rather likely that their presented results are based on single frame simulations with default initial conditions. It is therefore difficult to judge how strong these approaches are susceptible to hysteresis. It is clear that if an inter-frame reset is applied in the presented discontinuity networks, hysteresis is of no concern and the parameter values do not have to obey limits such as the self-excitation gain limit (4.34). Most likely, this would allow more optimal settings that lead to better single frame results than shown here.

It is certainly worth to consider an interframe reset of the network states for frame based, sequential segmentation systems. However, such periodic reset would severely contradict the continuous nature of the here proposed analog networks. At least a clock and an external control mechanism would be required that *decide* when the states in the network are valid and thus represent the desired estimate and when not (*e.g.* in the reset phase).

The presented motion segmentation system is designed such that it shows as little hysteresis as possible, although it cannot avoid it completely. Hysteresis becomes significantly noticeable *e.g.* by comparing the first frame in Figure 4.14 with all subsequent frames: Since the first frame is not burdened by any history, it exhibits a much cleaner discontinuity estimation that also matches the physical shape of the objects quite well. In later frames, the discontinuities appear to correlate less with the physical boundaries of the objects, which is especially obvious within the large tape-roll in frame #4 and #6. There, spurious discontinuities cause an inhomogeneous flow estimate within the large

[14] see Section 4.1.5 on page 56

tape-roll. Nevertheless, the coarse outline detection of the two objects is good and does not show much hysteretic artifacts. This surprises, regarding *e.g.* the low contrast conditions at the trailing edge of the tape-roll in front. Only in the last frame, part of the trailing boundary seems to get stuck, which causes the disrupted contour.

As expected from the limitations of the brightness constraint model, occlusion causes some trouble and leads to small partitions of unreliable motion estimates. It is interesting also to see that the system cannot distinguish between moving objects and dynamic shadows which can be observed *e.g.* in frame #2 and #6 at the lower bottom of the tape-roll in front.

4.2.3 The motion selective network

Local control of the bias conductances provides the possibility to suppress the response of a particular optical flow unit by clamping its output to some reference value. Thus, the perception of visual motion can actively be changed according to what is expected to be perceived and where.

In the following, the **motion selective network** is proposed which extends the optical flow network by a single additional network. As sketched in Figure 4.15, units of this network are retinotopically organized and receive input from the optical flow network according to some measure of how close the optical flow is to some expected motion $\boldsymbol{v}_{model}$. The connectivity of the motion selective network is chosen such that active units tend to appear in clusters of a preferred minimal size. In feed forward configuration, the activity of the network encodes roughly the spatial extent to which a motion source matches the expected motion. If the feedback connections to the optical flow network are enabled, then the active units in the network control the bias conductances of the related optical flow units: the conductance is increased significantly if the motion selective unit is not active and as a result, the optical flow estimate is suppressed locally.

We formalize the properties of the motion selective network once again as a constraint optimization problem and introduce the following cost function

$$\begin{aligned} H_A =& \frac{\alpha}{2} \sum_{ij} \sum_{kl \neq ij} A_{ij} A_{kl} + \frac{\beta}{2} \Big(\sum_{ij} A_{ij} - N_{max} \Big)^2 - \frac{\gamma}{2} \sum_{ij} \sum_{W} A_{ij} A_W \\ & - \delta \sum_{ij} A_{ij} (\boldsymbol{v}_{ij} \cdot \boldsymbol{v}_{model}) + 1/R \sum_{ij} \int_{1/2}^{A_{ij}} g^{-1}(\xi) \, \mathrm{d}\xi, \end{aligned} \tag{4.36}$$

where A_{ij} is the response of the motion selective unit at node ij and $g : a_{ij} \rightarrow A_{ij}$ is its non-linear activation function.

The first two terms are the equivalent constraints of the simple WTA network (3.6) with the small difference that the total activity is forced to be N_{max}. Choosing an appropriate activation function such a constraint will allow ideally N_{max} winners. We will call

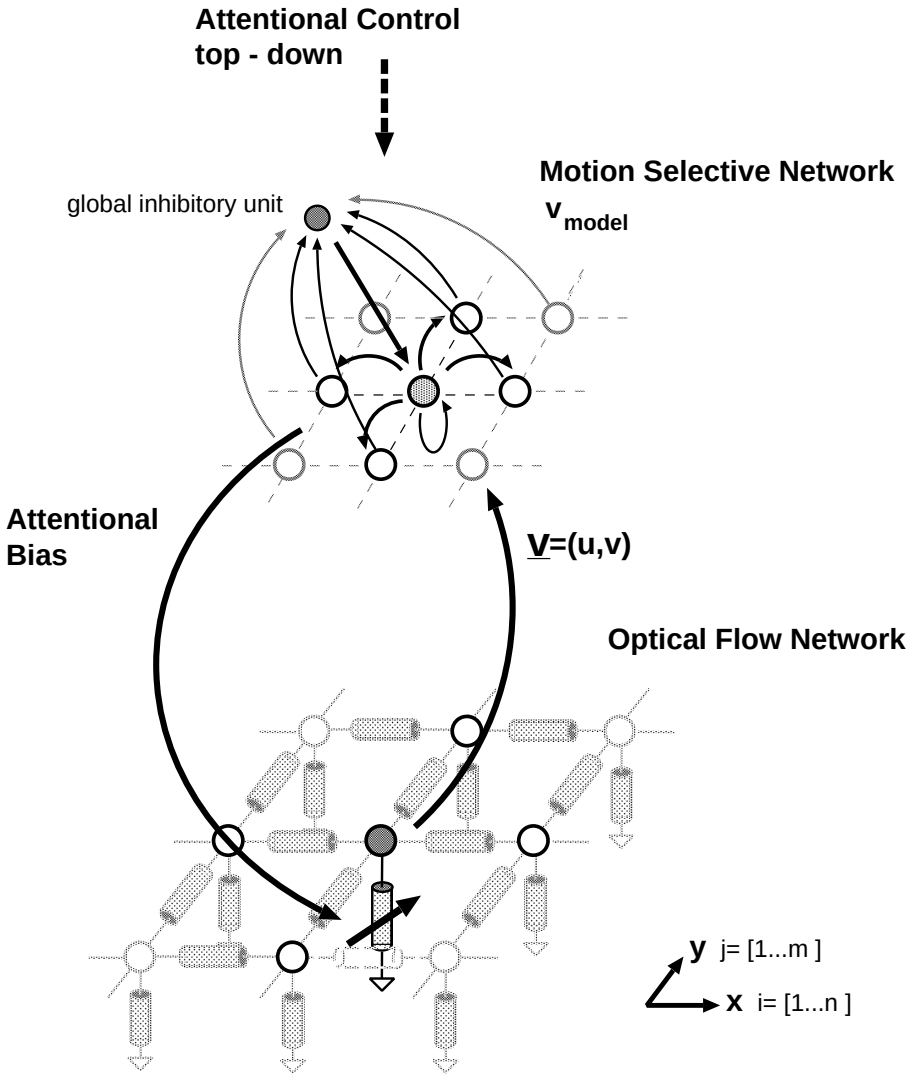


Figure 4.15: *Schematics of the motion selective network.* The optical flow network is combined with another retinotopic network that is selective to a particular type of motion. Its architecture is based on a WTA network that forces multiple winners. Additional excitatory connections to nearest-neighbors promote the clustering of active units. The output of the selective network is fed back to control the strength of the local bias conductances in a way that motion perception can be selectively restricted to only a particular kind of motion, specified by $\boldsymbol{v}_{model}$.

this network a **multi-winner-take-all** (mWTA) network. In contrast to the soft-WTA circuit in the motion discontinuity networks, every unit in the array participates in the *same* competition and has an equal chance to win. The fixed number of winners can be seen as the maximal amount of attentional resources for which the input has to compete for. Such a paradigm has been proposed as the underlying mechanism that accounts for attentional processes observed in primates [Kastner and Ungerleider 2000].

The activation function can be arbitrary as long as it is monotonically increasing, limited from below and – in contrast to the simple WTA network – saturates. Saturation is vital because otherwise, a single active unit with output $A_{winner} = N_{max}$ would correspond to the optimal solution of (4.36) and the desired multi-winner behavior would not occur. In our case, the activation function was chosen to be sigmoidal (3.5).

The third term favors units to be active within an active neighborhood W, hence leading to clusters of the network activity. The forth term represents the input to the network and prevents units to be active at locations where the optical flow estimate does not match the expected motion[15]. Finally, because we are using analog units with activation functions of finite slope the dissipated power in the unit must be taken into account, which is expressed in the last term.

Applying gradient descent to minimize (4.36) leads to the following dynamics of the network

$$\begin{aligned}\dot{a}_{ij} &= -\frac{1}{C}\Big[\frac{a_{ij}}{R} + \alpha \sum_{kl \neq ij} A_{kl} + \beta\big(\sum_{ij} A_{ij} - N_{max}\big) - \gamma \sum_{W} A_W - \delta(\boldsymbol{v}_{ij} \cdot \boldsymbol{v}_{model})\Big] \\ &= -\frac{1}{C}\Big[\frac{a_{ij}}{R} + (\alpha+\beta)\sum_{ij} A_{ij} - \alpha A_{ij} - \beta N_{max} - \gamma \sum_{W} A_W - \delta(\boldsymbol{v}_{ij} \cdot \boldsymbol{v}_{model})\Big]. \quad (4.37)\end{aligned}$$

The network architecture is given by the dynamics (4.37) and is very similar to the one of the simple WTA network[16]: Each unit receives global inhibitory input from all units in the network $(\alpha + \beta)$ but also excites itself (α). In addition, each unit has excitatory connections (γ) from and to units in a defined neighborhood W. Finally, a constant excitatory input which is proportional to N_{max} is imposed on all units. As proposed in Section 3.2.1 and shown in Figure 4.15, the connectivity is massively reduced using a global inhibitory unit that sums up the output of all units and inhibits each one accordingly.

The cost function (4.36) is bounded from below and above for any given and finite flow field $\boldsymbol{v}$ and is a Lyapunov function of the system (4.37). To assure global asymptotic stability within the mWTA network, the self-excitation gain limit has to be obeyed. Again, this limit depends on the exact form of the excitation kernel W. In the simple case of a 4-nearest-neighbor kernel with unity weights, we find that

$$\alpha + \gamma \leq \frac{a_0}{R} \quad (4.38)$$

[15]The chosen dot-product is just one possible difference measure.

[16]see Figure 3.2 on page 31

must hold in order to guarantee that the Hessian of the cost function (4.36) is positive-definite. Here $1/a_0$ is the maximal gain of the activation function, and R the input resistance. For more general kernels the appropriate conditions have to be found numerically.

Note that the cost function (4.36) is a soft representation of the imposed constraints. Thus the total number of winners in the network is not guaranteed to be N_{max}. In fact as seen in the simple WTA network, the maximal loop-gain can be too small to determine clear winners at all for weak and ambiguous input. This is dominantly caused by the inhibitory influence of the finite input resistance of the analog units which requires a certain activity level for units to be active. In any case, the total loop-gain can be increased almost arbitrarily without running into multi-stable behavior by using cascades of mWTA networks[17].

Closing the recurrent feedback-loop between the motion selective network and the optical flow network allows the motion selective units to control the strength of the local bias conductances. Hence, we modify the bias constraint in (4.13) to

$$B = \sum_{ij} (\sigma_1 + (1 - A_{ij})\ \sigma_2)\ \left[(u_{ij} - u_0)^2 + (v_{ij} - v_0)^2\right], \tag{4.39}$$

with $0 < \sigma_1 \ll \sigma_2$. Using a typical sigmoidal activation function (3.5) $A_{ij} \in [0, 1]$, the effective bias conductances are therefore in the interval $[\sigma_1,\ \sigma_1 + \sigma_2]$.

Simulation results

Simulation results are presented for the 'tape-rolls' sequence once for preferred rightward motion $\boldsymbol{v}_{model} = (1, 0)$ and once for leftward motion $\boldsymbol{v}_{model} = (-1, 0)$ as displayed in Figure 4.16 and Figure 4.17 respectively. Every second frame of the sequence is shown, superimposed with the resulting optical flow field. Below each frame, the corresponding activity in the attentional network is indicated as a gray scale representation. The high gain within the attentional network expresses an almost binary activity pattern. It is a consequence of using a two-stage mWTA arrangement which is necessary for sufficient suppression ($1 - A_{ij} \to 0$). The network parameters were kept constant in both cases and set as follows: the mWTA parameters $\alpha = 0.02$, $\beta = 0.5$ and $\gamma = 0.05$, the input weight $\delta = 1$ and self-inhibition $a_0/R = 0.1$. These parameters were identical in both mWTA-stages. The default number of winners was slightly reduced in the preferred leftward motion case to account for the size difference of the objects sizes. The lateral conductance and the lower limit of the bias conductance in the optical flow network were kept as in the smooth optical flow example ($\rho = 0.075$ and $\sigma_1 = 0.0025$) and the maximal leak conductance $\sigma_2 = 0.8$.

[17] see Figure 3.3 on page 33

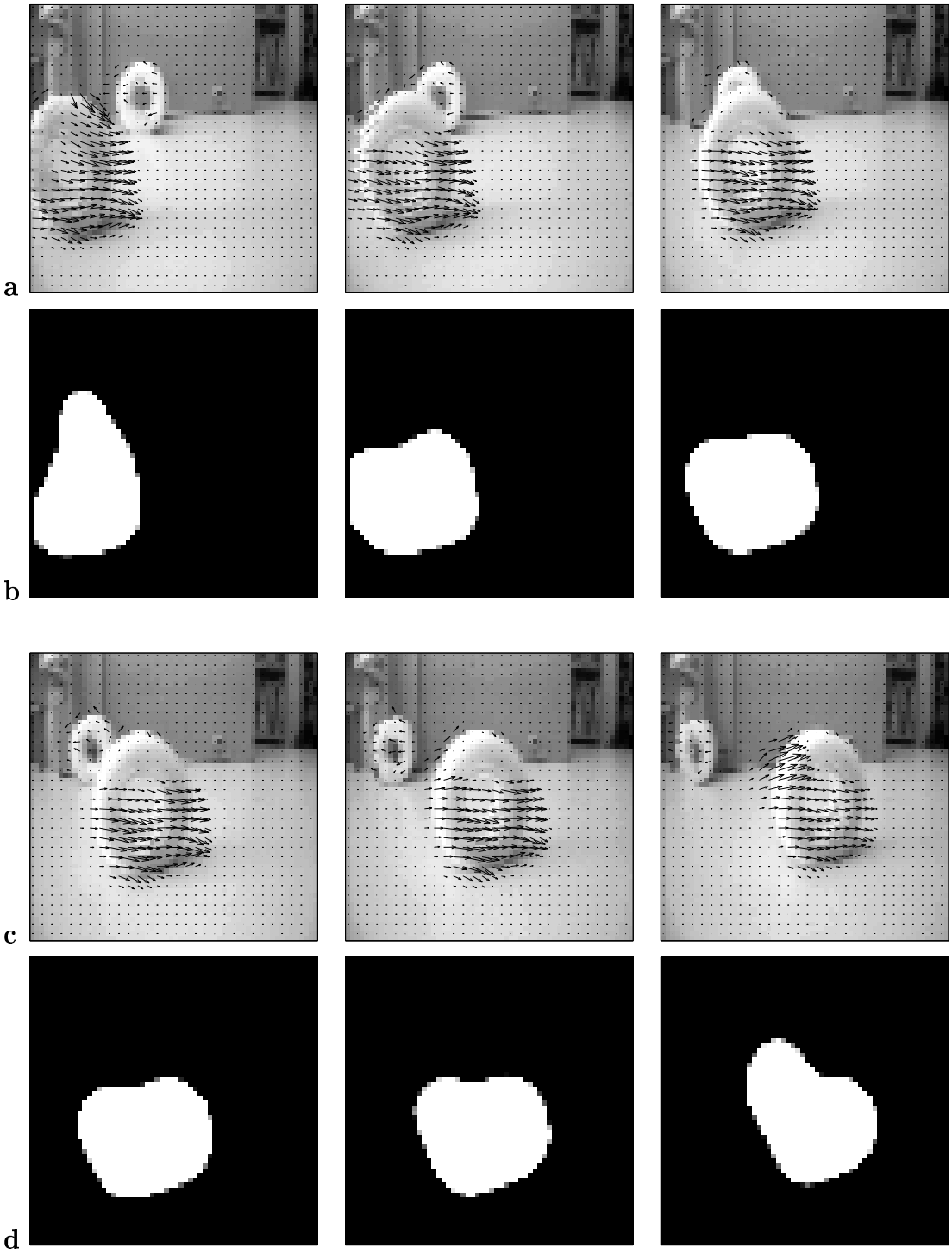


Figure 4.16: *The motion selective network tuned to rightward motion.* (a,c) The 'tape-rolls' sequence (frame # 2,4,6,8,10,12) is shown overlaid with the resulting optical flow field. The attentional feedback clearly suppresses the motion field associated with the smaller tape-roll. (b,d) The activity in the motion selective network ($N_{max} = 500$) is shown at the associated time instance. Note that no reset of the network states is performed in between the different frames.

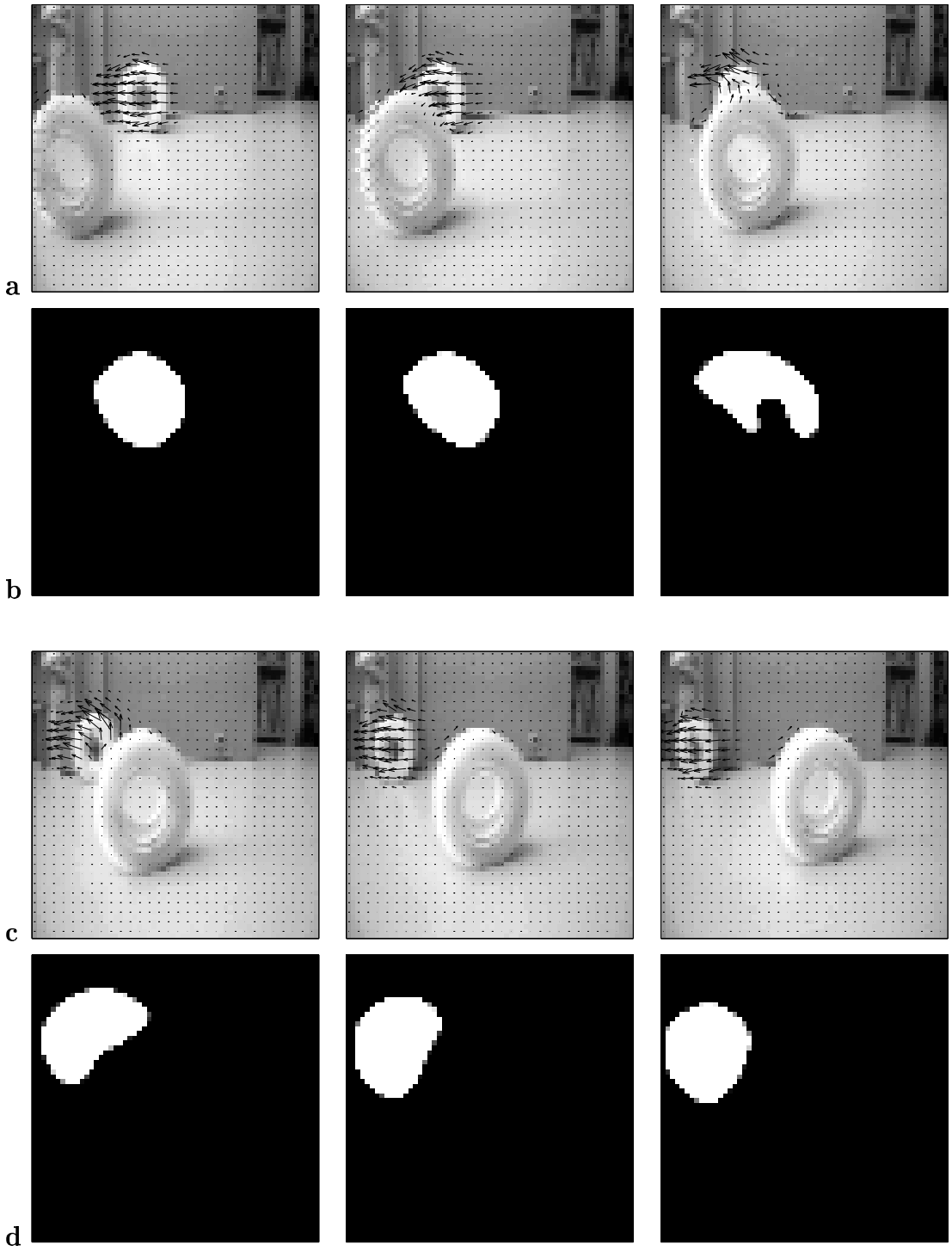


Figure 4.17: *The motion selective network tuned to leftward motion.* (a,c) Again the 'tape-rolls' sequence, but now leftward motion is favored. (b,d) The associated activity in the motion selective network is clearly assigned to the smaller roll. The size of attention is slightly reduced in comparison to the previous example ($N_{max} = 300$).

For preferred rightward motion, the leftward motion field of the smaller roll in the back is almost completely suppressed while the flow field on the front roll is correctly estimated. The activity in the motion selective network nicely matches the size of the tape-roll and does not show any shift towards the trailing edge which could be expected as the result of hysteresis. Preferring leftward motion (Figure 4.17), the system assigns all its attentional activity to the small roll in the back. Again, the induced local adjustment of the bias conductances strongly suppresses the optical flow field everywhere else. The parameters besides the preferred motion were kept identical in both cases. Again, there was no interframe reset of the network states.

Stability of the complete system is guaranteed because each sub-system, the optical flow network and the motion selective network, is asymptotically stable in any case[18]. Although a detailed analysis of conditions for global asymptotic stability of the complete system is not performed, it seems very likely that the positive feedback can elicit hysteretic behavior. In fact, hysteresis is observed in the simulation results such as *e.g.* in Figure 4.16, where the top part of the tape-roll in front is not or only partially detected although its motion matches the preferred motion. This is likely due to the occlusion of the two tape-rolls occurring in the first few frames which induces some non-preferred optical flow in that region. It prevents the network from becoming active again right away. It takes some time to overcome this 'experience' until the last frame, where a partial recovery can be observed. It has been verified that this effect is due to hysteresis: Reapplying the whole sequence but deleting the network history right after occlusion occurred (frame #8) leads to an immediate recovery of the activity pattern in the top part of the tape-roll (not shown).

Nevertheless, the effect of hysteresis seems to be less pronounced and distorting than in the previously discussed segmentation network. A possible reason can be the nature of the mWTA network: The network strongly encourages always N_{max} units to be active. In fact, the chosen value for σ_2 does not result in a complete suppression of the optical flow in non-attended areas. Thus, the suppressed inputs to the motion selective network still have the chance to win over the diminishing inputs at locations where motion has already passed.

The motion selective network also provides means to allow top-down influence. Firstly, it contains two global parameters, the preferred motion $\boldsymbol{v}_{model}$ and the attention kernel size N_{max} that can be thought of as controlled by the output of another, even higher level network. These parameters represent object related properties that are independent of a retinotopic representation. Any networks that provide such output will inevitably follow a different topological policy than the one introduced here. Secondly, the motion selective units can receive direct input, such that the attentional process is spatially

[18] We assume that the time constants of both networks are much shorter than that of the visual motion itself.

guided. This steering input can be generated by another retinotopically arranged place encoding network or some lower dimensional, variable-encoding map [Hahnloser et al. 1999].

4.3 Conclusions

I have shown how optical flow estimation can be formulated within the framework of constraint optimization. The introduced bias constraint thereby reveals many interesting properties. As rigorously proven, it is this constraint that turns the estimation of visual motion into a well-posed problem under any visual input conditions. The bias constraint also allows to apply *a priori* knowledge about the expected visual motion if visual input is ambiguous or absent.

More important, I have demonstrated how such optimization problems can be solved by simple analog electronic networks. A basic network architecture is proposed that reliably computes smooth optical flow. According to the strength of the lateral conductances in the network, the resulting output broadly varies between a normal flow and a global flow estimate. Unlike earlier proposals the network architecture is simple and does not contain any negative conductances [Hutchinson et al. 1988].

Extensions of the basic network have been discussed that allow the local adaptation of the bias and the smoothness conductance, in order to control the motion integration process and the suppression of network activity. The two proposed systems each use additional recurrently connected networks that allow them to perform motion segmentation and motion selective attention tasks where the activity in the layer directly represents motion discontinuities or the attentional strength respectively. Since segmentation and attention are decisive processes per definition, different variations of the classical WTA architecture form the core circuits of these additional network layers. In particular, a cross-excitation soft-WTA architecture is introduced that forces the formation of continuous motion boundaries while suppressing nearby parallel oriented line segments. A multi-WTA architecture is proposed for the motion selective attention network that assigns a preferred number of winning units to the input area that corresponds best to a preferred motion.

An important aspect of decisive processes within the temporal domain is their possible multi-stable behavior that can produce hysteretic effects. The proposed networks are designed such that they exhibit as little hysteresis as possible. This is achieved by respecting the weight conditions for mono-stability in the decisive network layers (self-excitation gain limits). Furthermore, the positive loop-gain is reduced by minimizing the feedback effect onto the inputs to the decisive stages (combined gradient differences and non-complete suppression). The presented simulation results using real-world image sequences demonstrate that these arrangements allow to keep hysteresis on an acceptable

level.

For all proposed networks, highest priority was given to simple (WTA-)architectures that permits a physical implementation as integrated electronic circuits. This requirement should be considered by those who draw comparisons of this approach, with alternative approaches reported elsewhere in the literature.

Chapter 5

Neuromorphic Implementation

A substantial part of this dissertation is dedicated to aVLSI implementations of some of the formulated constraint optimization networks. Idealized electronic circuits of these networks have been derived in the previous chapter. However, the implementation of such circuits imposes many problems arising from the fundamental physical properties of semiconductors and the inherent non-idealities of the applied process technologies. Definitions, symbols and the large signal transistor model used within this chapter are introduced in Appendix C.

It seems to be the appropriate place to acknowledge the pioneering work of John Tanner [Tanner 1986, Tanner and Mead 1986]. His attempt to implement a network for global motion estimation was both a source of inspiration and the challenge to prove that a functional implementation of such a network is possible. I am pleased to present here successful focal-plane implementations of the smooth optical flow network and the simplified motion discontinuity networks. Their characteristic behavior and performance under real-world visual conditions is measured and presented.

5.1 Computation of the Image Brightness Gradients

So far, the spatiotemporal brightness gradients that constitute the input to the proposed networks have been considered to be given. Nevertheless, they need to be extracted first. We show the circuits that are necessary to compute the spatiotemporal brightness gradients and discuss the influence of spatial sampling.

5.1.1 Adaptive photoreceptor circuit

At the basis of processing remains the transduction of the visual information into appropriate electronic signals. Preferably, such a transduction stage has to work over a wide range of light intensities, while remaining sensitive to small contrast changes. The **adaptive**

photoreceptor circuit [Delbruck 1993b, Delbruck and Mead 1994] meets these requirements and was an early but successful example of how biological phototransduction can be functionally translated into semiconductor technology. The circuit logarithmically encodes the irradiance values. A feedback loop connects the voltage-clamped photodiode through an adaptive element to the feedback transistor that provides the photocurrent. The adaptive element adjusts the operating point according to the slow dynamics of the global brightness while a high gain capacitive divider stage amplifies the transient response around the common operating point.

In the two presented implementation examples, adaptive photoreceptors are used where the adaptive element is modified as proposed by Liu [1998] (see schematics in Figure 5.4). The adaptive element consists of a well-transistor whose gate is controlled by a native-type source follower that replicates the output node voltage of the photoreceptor with a gain factor $\kappa < 1$ and an offset given by the voltage Bias_{PhA}. In first approximation we assume κ to be constant and therefore the source follower ensures that the adaptation current is exponential in the transient amplitude of the photoreceptor output voltage. Thus, the adaptation time constant is ideally uniform over a wide output-range and can be controlled by the gate voltage Bias_{PhA}. However, the time constant is not symmetric with respect to on- and off-transient responses. In addition to its temporal high-pass filter characteristics, the photoreceptor circuit shows also a low-pass filter behavior[1]. The cut-off frequency is determined mainly by the background light intensity, the capacitance of the photodiode and the parasitic (Miller-)capacitance between output node and photodiode. Lowering the bias current of the photoreceptor (by increasing Bias_{Ph}) reduces the speed of the high gain output amplifier, and so allows the cut-off frequency to decrease below that limit.

To summarize: The adaptive photoreceptor acts as temporal band-pass filter where the characteristic frequencies are adjustable to some degree. Such a filter stage at the very beginning of the visual motion process discards the non-relevant high- and low-frequency contents of the visual information (such as the absolute brightness level or flicker noise) from propagating into following processing stages.

5.1.2 Continuous temporal derivative circuit

A correct velocity estimation within gradient-based motion circuits is only possible if temporal differentiation is performed accurately over a broad temporal frequency spectrum. The circuit shown in Figure 5.1 does meet these specifications to a large extent. Originally called **hysteretic differentiator** [Mead 1989] and also considered in a slightly different configuration, it has been used in a number of motion detection circuits [Kramer et al. 1997, Higgins and Koch 1997, Harrison and Koch 1998, Liu 2000].

[1] see [Delbruck 1993b] for a detailed analysis

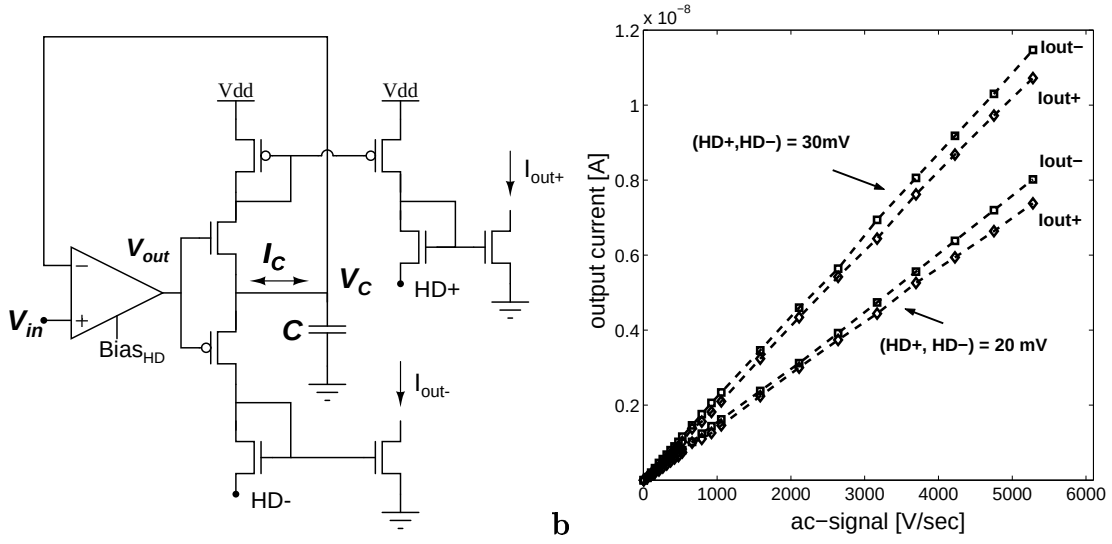


Figure 5.1: *The hysteretic differentiator circuit and its measured response curves.* (a) The hysteretic differentiator circuit consists of a transconductance amplifier, the capacitance C and a stage that provides the rectified output currents that are ideally proportional to the temporal derivative of the input voltage V_{in}. (b) The peak currents of I_{out+} and I_{out-} are shown as a function of the applied temporal voltage gradient for a saw-tooth input waveform with large peak-to-peak amplitude (2 Volts). The currents are plotted for two different levels of the adjustable source potentials $HD+$ and $HD-$. Note that the cut-off frequency is beyond the largest measured values of the temporal gradient (5000V/sec).

The circuit consists of a high gain transconductance amplifier that controls the gate voltages of a pair of transistors of opposite polarity, which are connected at their sources. The non-inverting input to the transconductance amplifier is driven by the photoreceptor output while the inverting input is the voltage V_C on the capacitance C. Depending on the output of the transconductance amplifier, current flows either through the upper transistor (nFET) to charge the capacitance C or through the lower transistor (pFET) to discharge C. The current through each branch is mirrored with current mirrors of variable source potentials $HD+$ and $HD-$ respectively to provide the rectified and potentially amplified currents I_{out+} and I_{out-} of the capacitor current I_C.

Ideal differentiation is performed by a single capacitor, where the current I_C flowing onto the capacitance C is proportional to the temporal derivative of the voltage V_C across it, thus

$$I_C = \frac{1}{C}\frac{\mathrm{d}V_C}{\mathrm{d}t} \quad . \tag{5.1}$$

If we assume the transconductance amplifier has zero output conductance and thus infinite gain, then the hysteretic differentiator acts as an ideal differentiator, because V_C follows infinitely closely the input signal V_{in} and the current I_C is given as in Equation (5.1). For any time-varying input signal, the amplifier's output V_{out} is always such that it allows I_C to flow either from the upper or out of the lower branch of the rectifying circuit. In steady state, V_{out} is adjusted such that the currents through the upper and lower branch are equal. Still, V_C is infinitely close to V_{in} because the gain is infinite.

The finite gain of the transconductance amplifier causes some deviations form this ideal behavior. There is some finite potential difference $\epsilon = |V_{in} - V_C|$ needed to open either side of the rectifying circuit enough to provide the necessary currents to the capacitance. In steady state, ϵ is close to zero and its exact value depends on V_{in}, the amplifier gain and the transistor characteristics in the rectifying circuit. For a linearly increasing input signal, the current to the capacitance and therefore I_{out+} is at first smaller than the true derivative current until ϵ is large enough to provide the whole current. If the input signal stops increasing and stays constant, the output current is still positive for some transition period until ϵ reaches its steady state value. The same is true for a linearly decreasing input signal.

The hysteretic differentiator shows a threshold behavior for transitions between input signals with temporal gradients of opposite sign. In steady state, the leak current through the reverse-biased junction of the nFET source requires the offset $V_{out} > V_C$ to be large enough to provide the necessary amount of leak current. If the input signal now decreases, V_{out} has to surmount at least this offset in order for the pFET to conduct. Every transition between positive and negative gradients requires the amplifier output to perform this voltage step. The offset voltage can be as large as 1 Volt and leads typically to a threshold voltage $\epsilon_{th} = [10 \ldots 30]$ mV, depending on the amplifier gain. For signal amplitudes smaller than ϵ_{th}, the hysteretic differentiator does not provide a proper measure of the

temporal derivative.

For large input signal amplitudes where the temporal gradient is constant over some sufficiently large voltage range, however, the peak currents are a very close approximation of the temporal derivative. Figure 5.1b demonstrates the wide linear frequency-range of the differentiator output for large amplitude signals. As shown, the adjustable source nodes within the output current mirrors allow to tune individually the amplification of the output currents I_{out+} and I_{out-}. Although the upper frequency limit of accurate differentiation was not measured, it exceeds the upper cut-off frequency of the photoreceptor even under strong illumination conditions [Delbruck 1993b].

5.1.3 Spatial sampling and aliasing conditions

Analog networks naturally provide continuous-time behavior, which is important for any signal processing in the temporal domain. Although temporal aliasing conditions are not expected, a focal-plane implementation of an array of single processing units with individual photoreceptors inevitably implies *spatial sampling* of the irradiance distribution on a regular grid due to the finite size of each unit in the array. Figure 5.2a illustrates the situation for a one-dimensional array with spacing parameter d. Only part of the incoming visual information is transduced into a current by a single photodiode, leading to a classical reconstruction problem as dealt within sampling theory. Spatial aliasing[2] can occur depending on the array parameter d and the spatial frequency of the irradiance distribution. The finite size D of the photodiode leads to patch-sampling rather than point-sampling of the irradiance distribution. Here the *fill factor* is given by $\delta = D/d$. For any 2D active pixel sensor $\delta < 1$, and usually decreases as increasing computational complexity is built into a single pixel. While strict point-sampling ($\delta = 0$) preserves spatial aliasing effects for arbitrarily high frequencies, a finite photodiode size introduces a low-pass filter characteristics due to the spatial averaging (Figure 5.2b).

We consider a simple visual input presented to a simplified model of the optical flow network in order to investigate how spatial aliasing and patch-sampling influence the motion estimation. Assume an arrangement as depicted in Figure 5.2a, where a sinusoidal brightness grating is moving with a fixed velocity $\boldsymbol{v}$ while it is seen by a one-dimensional optical flow network. We describe the irradiance distribution as

$$E(x,t) = \sin(kx - \omega t) \qquad \text{with } x, t \in \mathbb{R}, \tag{5.2}$$

where $\omega = kv$ and k is the spatial frequency of the projected stimulus.

The estimation of the spatial brightness gradient of a spatially sampled image is an ill-posed problem per definition [Bertero et al. 1987] and requires a model and optimization

[2]Definition: *aliasing* - folding of higher into lower frequency components in a discrete spectrum due to undersampling of the signal.

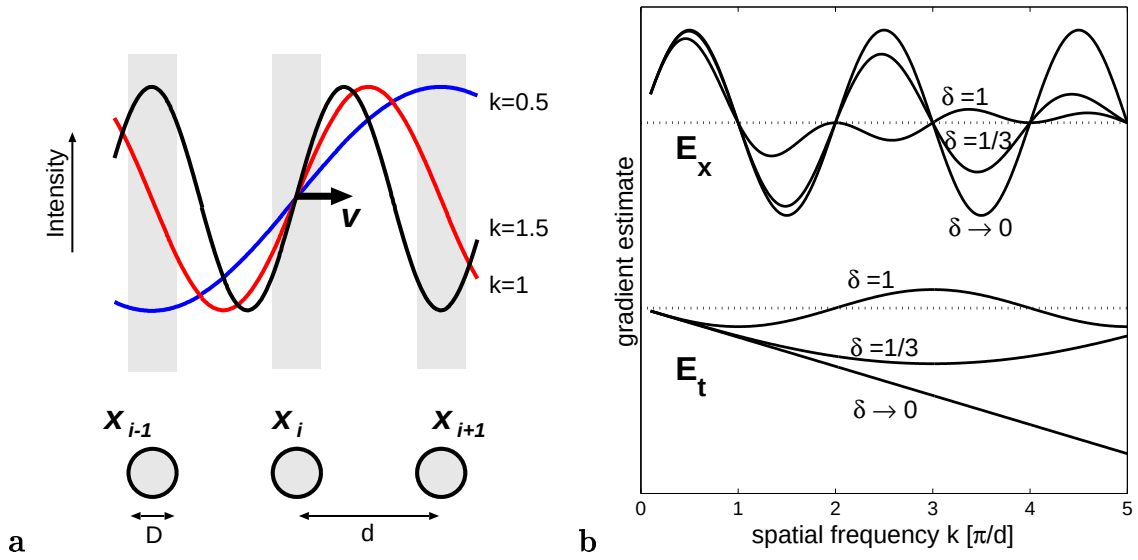


Figure 5.2: *Spatial sampling and continuous-time sensing of the irradiance distribution.* (a) Spatial sampling of sinusoidal patterns of different spatial frequencies k (given in units of the Nyquist frequency $[\pi/d]$) in one visual dimension. The discrete spatial patch-sampling of sinewave patterns discards the visual information outside the shaded areas and thus leads to a classical reconstruction problem. The approximation of the spatial gradient by a difference operator is influenced by aliasing and becomes significantly incorrect for higher spatial frequencies. (b) An increasing photodiode size affects both the temporal and spatial gradient estimation by introducing a low-pass filter behavior but might also change sign. However, the effect is only significant for high fill factors and/or very high spatial frequencies. For pure point-sampling ($\delta \to 0$) the peak amplitude of the spatial gradient estimate remains constant while the temporal gradient estimate increases linearly with spatial frequency.

procedure very much in the way as introduced in the previous chapters. A smoothness regularizer at the photoreceptor level could be implemented with a resistive mesh. Instead the brightness distribution was assumed to be linear in between the sampling points. The continuous one-dimensional spatial differentiation at location x_i thereby reduces to the difference operator

$$\Delta x_i = \left(\frac{x_{i+1} - x_{i-1}}{2d} \right) \tag{5.3}$$

Considering the average irradiance over D, the discrete spatial gradient becomes

$$\begin{aligned} E_x(x_i,t) =& \frac{1}{2d} \left(\frac{1}{D} \int_{x_i+d-D/2}^{x_i+d+D/2} \sin(k(\xi - vt))\, \mathrm{d}\xi \quad - \quad \frac{1}{D} \int_{x_i-d-D/2}^{x_i-d+D/2} \sin(k(\xi - vt))\, \mathrm{d}\xi \right) \\ =& \frac{1}{kDd} \left(\sin(k((x_i+d) - vt)) \cdot \sin(kD/2) - \sin(k((x_i-d) - vt)) \cdot \sin(kD/2) \right) \\ =& \frac{2}{kDd} \sin(kd) \cdot \sin(kD/2) \cdot \cos(k(x_i - vt)). \end{aligned} \tag{5.4}$$

Similarly, the temporal gradient at location x_i is

$$\begin{aligned} E_t(x_i,t) =& \frac{\partial}{\partial t} \frac{1}{D} \int_{x_i-D/2}^{x_i+D/2} \sin(k(\xi - vt))\, \mathrm{d}\xi \\ =& -\frac{v}{D} (\sin(k(x_i - vt + D/2)) - \sin(k(x_i - vt - D/2))) \\ =& -\frac{2v}{D} (\cos(k(x_i - vt)) \cdot \sin(kD/2)). \end{aligned} \tag{5.5}$$

According to (4.22), a single unit of the optical flow network in equilibrium, with no lateral interconnection, is expected to report normal flow that in the one-dimensional case reduces to

$$V_{out}(x_i,t) = -\frac{E_t(x_i,t) E_x(x_i,t) + u_0 \sigma}{\sigma + E_x(x_i,t)^2} \tag{5.6}$$

For simplicity we assume the reference motion $u_0 = 0$. Substituting (5.5) and (5.4) into (5.6), we can predict the motion response of a single unit depending on the spatial frequency of the stimulus and the fill factor δ as well as on the weight of the bias constraint σ. Figure 5.3 shows the predicted peak output response for different parameter values as a function of the spatial stimulus frequency k, given in units of the Nyquist frequency $[\pi/d]$. Note that the response strongly depends on k. Only for low spatial frequencies, the motion output approximates well the correct velocity. For very low frequencies, the locally perceived stimulus contrast diminishes and the non-zero σ biases the output response towards the reference motion. As the spatial frequency approaches the Nyquist frequency, the response increases to unrealistically high values and changes sign at $k \in \mathbb{Z}^+[\pi/d]$. A non-zero σ and an increasing fill factor reduce this effect. However, their influence is only marginal for reasonable values of δ and small σ.

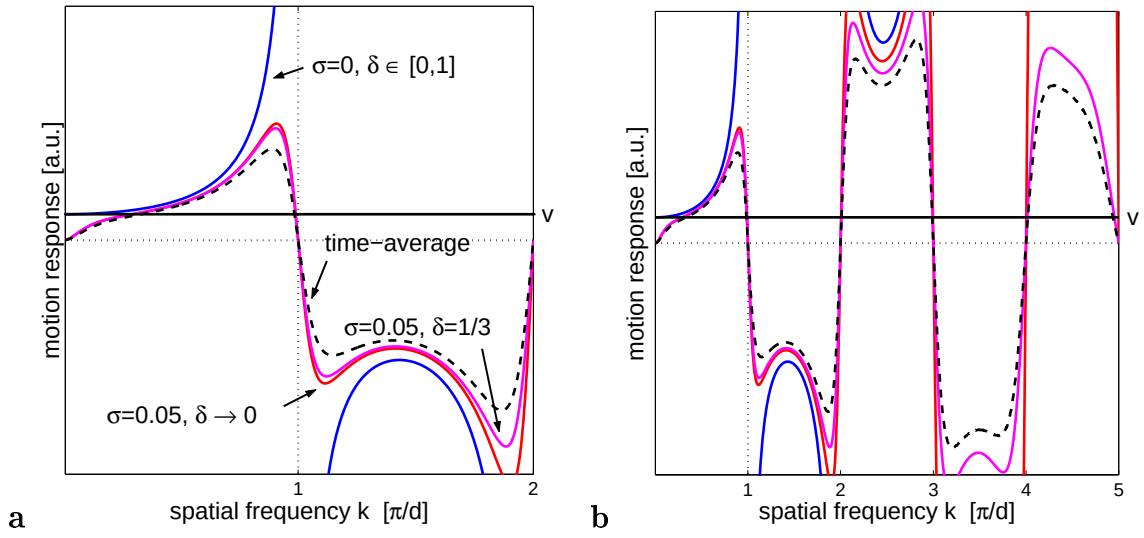


Figure 5.3: *Motion response dependence on spatial stimulus frequency* (a) The expected response of the optical flow chip according to (5.6) as a function of the spatial frequency k (given in units of the Nyquist frequency $[\pi/d]$) of a sinewave stimulus moving with velocity v: All solid curves are taken for a single motion unit at location $x = 0$ and time $t = 0$. Only for low spatial frequencies the motion response resembles the true velocity. For $k > 0.5$ the discrete approximation of the spatial gradient leads to a significantly increased response. At multiples of the Nyquist frequency $k = [1, 2, \ldots, n]$ spatial aliasing reverses the direction of the perceived motion. Without the bias constraint ($\sigma = 0$) the computation is ill-posed at multiples of the Nyquist frequency and the response goes to $\pm\infty$ in their vicinity. Increasing the fill factor does *not* affect the motion response unless $\sigma > 0$. A finite σ biases the response at very low frequencies to the reference motion $u_0 = 0$. The dashed curve is the time-averaged response over the duration of one stimulus cycle. (b) A magnified view covering the same frequency-range as shown in Figure 5.2: The label of the curves are equivalent to (a). Note that the influence of the fill factor only becomes significant for unrealistically high spatial frequencies $k > 1$.

In the extreme and hypothetical case where $\sigma = 0$, (5.6) reduces to

$$V_{out} = v \, \frac{kd}{\sin(kd)} \qquad \text{with } u_0 = 0. \tag{5.7}$$

The computation is ill-posed at spatial frequencies $k \in \mathbb{Z}^+$ and the motion response approaches infinity close to their vicinity. For $k \to 0$, however, the correct velocity estimate is found. Two things are remarkable: Firstly, D drops out in Equation (5.7) which means that the low-pass filter characteristics of a non-zero fill factor only influences the response for a non-zero σ. Thus, for high k the overall motion response increases because the estimated spatial gradient amplitude remains constant but the temporal gradient scales linearly (see Figure 5.2). Secondly, the motion output response does not depend on time nor space, meaning that a uniformly moving sinewave pattern induces a constant output response.

However, this is not the case for $\sigma > 0$. According to (5.6) we find the motion response to be

$$V_{out}(x_i, t) = v \, \frac{kd}{\sin(kd)} \cdot \frac{\cos(kx_i - \omega t)^2}{\alpha + \cos(kx_i - \omega t)^2} \tag{5.8}$$

with

$$\alpha = \sigma \, \frac{k^2 d^2 D^2}{4 \sin^2(kd) \sin^2(kD/2)} \, .$$

Since $\alpha > 0$, the motion response $V_{out}(x_i, t)$ is in the range between zero and the hypothetical constant response given by (5.7) at any time t. The exact value depends on the spatial frequency k of the stimulus and the array parameters d and D. Furthermore, the motion response is now **phase-dependent** due to the remaining frequency terms in (5.8). The response follows the stimulus frequency ω and thus oscillates sinusoidally. As a consequence, the time-averaged motion response over the duration of a complete stimulus cycle (dashed curve in Figure 5.3)

$$\overline{V}_{out}(x_i) = \int_t^{t+T} V_{out}(x_i, t) \, \mathrm{d}t \quad \text{with } T = \frac{1}{\omega} \tag{5.9}$$

is always less than the peak response at the particular place $x = 0$ and time $t = 0$, where the sinewave stimulus shows its highest temporal gradient.

Phase-dependence of the motion response of an isolated motion unit is an inherent property of the approach caused by the non-zero σ and the single spatial sampling scale in the gradient estimation. The equivalent problem is encountered in motion energy approaches where quadrature pairs of local spatiotemporal filters with a spatial phase-shift of 90° have to be combined in order to achieve phase-independent motion energy responses [Adelson and Bergen 1985, Grzywacz and Yuille 1990, Heeger 1987a]. However, the collective computation amongst the units of the optical flow network results in the spatial integration of the visual information and allows to partially overcome the phase-dependent response of a single unit.

5.2 The Optical Flow Chip

An aVLSI implementation of the smooth optical flow network derived in the previous chapter (see Figure 4.2) is presented. Articles on preliminary implementations of this network have been published already [Stocker 1998, Stocker and Douglas 1999]. The discussion in the following, however, will be restricted mainly to the latest version, briefly named *the optical flow chip*. This prototype implementation consists of a 10x10 quadratic array of single motion units, each having a size of $(124\,\mu\text{m})^2$, as fabricated in a $0.8\,\mu$m BiCMOS process. Detailed process specifications are listed in Appendix E.

5.2.1 Single motion unit

The complete schematics of a single motion unit of the optical flow chip are drawn in Figure 5.4. Besides the photoreceptor and the hysteretic differentiator circuit, it consists of two identical sub-parts for each of the two optical flow components, including the feedback loops necessary to embed the brightness constraint. The general encoding scheme throughout the entire cell is such that each variable is encoded as the *difference* of two electrical signals, either currents or voltages. Thus for example, the two estimated components of the optical flow vector u and v are given as the potentials U_+ and V_+ with respect to some reference voltage V_{ref}.

The circuit is easiest understood if we consider the current equilibrium at the two capacitive nodes C_u and C_v. For each node, the sum of four bi-directional currents has to be zero, namely

- the differential output current of the high output-gain stage $(Fx_- - Fx_+)$ and $(Fy_- - Fy_+)$ respectively,
- the output current I_B of the transconductance amplifier,
- the current I_S flowing to neighboring units in the resistive network,
- and finally the capacitive current I_C.

The *only* purpose of the circuit is to adjust the potentials U_+ and V_+ on the nodes according to the current equilibrium. The design of the circuit ensures that the dynamics of the potentials map as closely as possible the dynamics derived in (4.16). Since we have shown that these dynamics are a sufficient condition for solving the optimization problem, the solution will be provided for free: we just let the circuit behave according to its natural dynamics!

In steady state when $I_C = 0$, the circuit has solved the constraint optimization problem and the current equilibrium at the capacitive node reflects the necessary condition for the global solution (4.14).

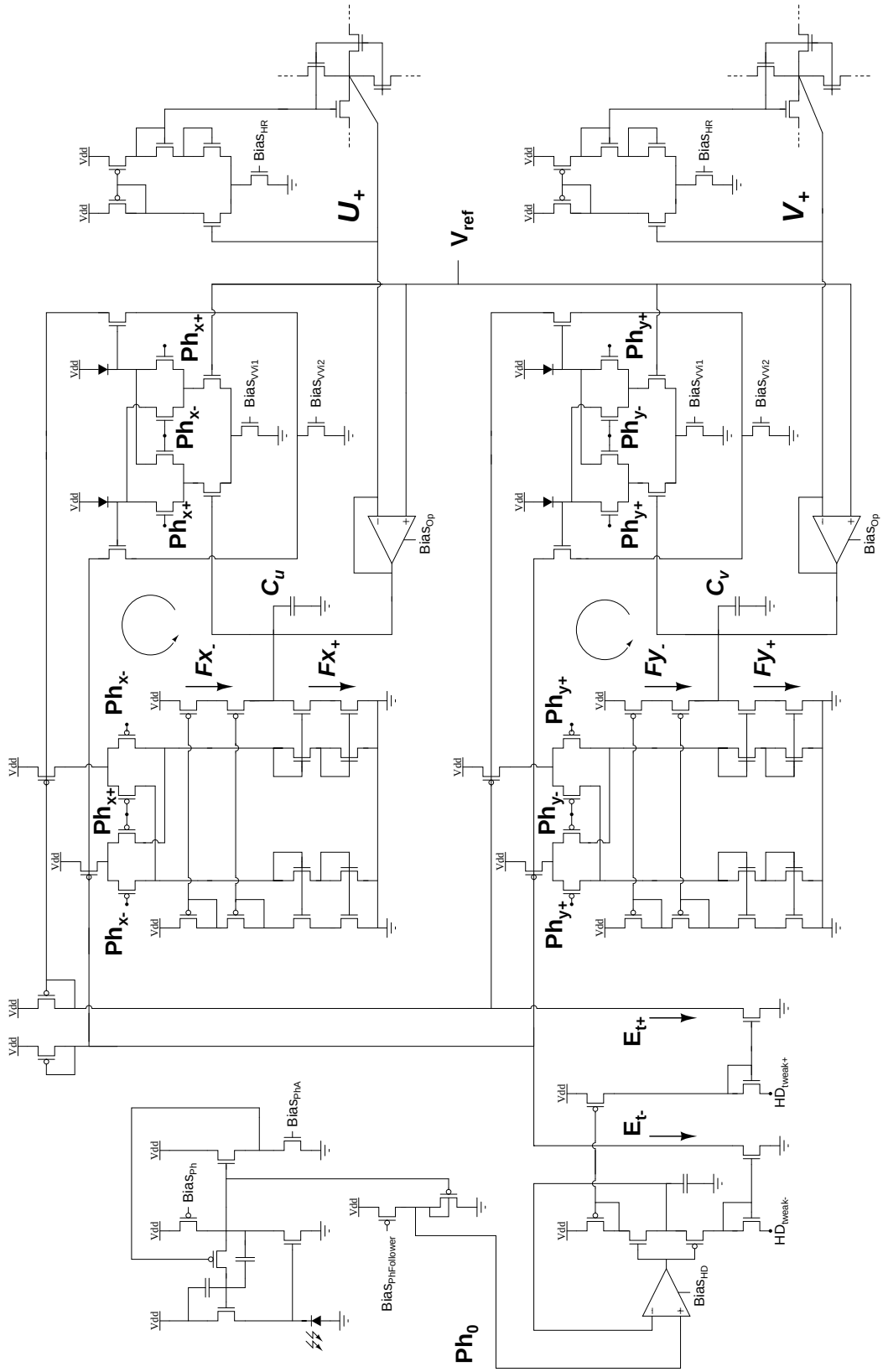


Figure 5.4: *Schematics of a single motion unit.* The complete circuit diagram of a single 2D motion unit is shown, including the photoreceptor and the temporal differentiator circuit.

We apply the following mapping between the physical quantities in the circuit (currents) and the mathematical expressions in the dynamics of the optimization problem, exemplarily shown for one optical flow component:

$$\begin{aligned} (Fx_+ - Fx_-) &= -E_{x_{ij}} \left(E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}\right) \\ I_S &= \rho \left(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}\right) \\ I_B &= \sigma \left(u_{ij} - u_0\right) \end{aligned} \tag{5.10}$$

Hence, each of the above currents represents the momentary influences of the three constraints: The lateral currents flowing in the resistive network ensure smoothness; the conductances of the transconductance amplifiers bias the optical flow estimate U_+ and V_+ to some reference potential V_{ref}; and the correction currents (Fx_-, Fx_+) and (Fy_-, Fy_+) force the brightness constraint to hold. However, the size of these currents depends on the actual estimate of the optical flow. While this dependence is intrinsically respected in the resistive network and the transconductance amplifier, an explicit feedback loop is needed to generate the correction currents. This feedback loop (indicated by circular arrows in Figure 5.4) is expensive from a circuit's perspective. It has to compute the differential correction currents which requires the emulation of several mathematical operations (5.10). Two wide linear-range Gilbert multipliers provide the four-quadrant multiplication $E_x u$ and $E_y v$ between the spatial gradients and the instantaneous estimate of the optical flow.

Addition of the output currents of the multipliers and the output currents of the temporal derivative circuit is performed at the pair of current mirrors. The summed currents are then mirrored to a second pair of Gilbert multipliers that perform the outer multiplication with the spatial intensity gradient. Finally, the loop is closed by cascoded mirroring and re-connecting the multipliers' outputs to the capacitive nodes C_u and C_v respectively.

5.2.2 Wide linear-range multiplier

The design of the multiplier circuit needed to compute $E_x u$ and $E_y v$ is critical for a successful implementation of the optical flow network. The demands for such a circuit are ambitious, including four-quadrant operation and a wide linear-range while keeping the circuit still as compact as possible. Linearity, in this context, means that the output current is proportional in each term of the multiplication. The standard Gilbert multiplier circuit implemented with MOSFETs and operated in weak inversion provides compact four-quadrant multiplication but offers only a fairly limited linear range [Mead 1989].

We present a modified version of the standard Gilbert multiplier circuit, shown in Figure 5.5. The circuit embeds the original multiplier within an additional differential pair that enables the multiplier core to operate above threshold while still providing subthreshold output currents. As shown in Appendix D, above-threshold operation of the

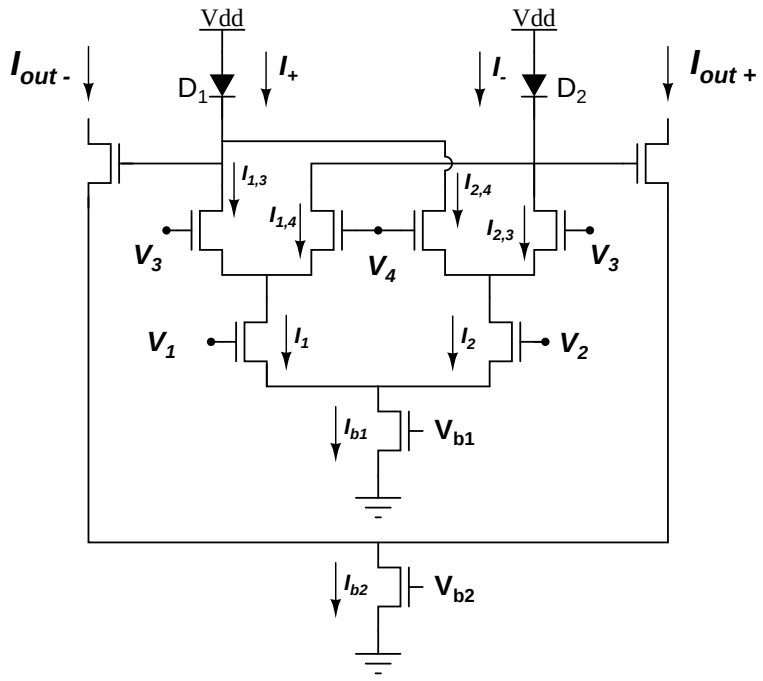


Figure 5.5: *Wide linear-range multiplier circuit.* The main advantage of the circuit is its compactness and the fact that its linear range and the output current level can be independently controlled. The bias current I_{b1} directly determines the linear range in above-threshold operation and can be adjusted with V_{b1}. Two ideal diodes logarithmically transform the currents I_+ and I_- into a voltage difference that serves as the input to the outer differential pair and provides the scaled down sub-threshold output currents I_{out+} and I_{out-}. The voltage V_{b2} allows to adjust the total output current level to match the temporal differentiator output.

simple differential pair circuit increases substantially its linear range. Since a Gilbert multiplier is a stack of three differential pairs, we expect its linear-range to be enlarged as well if biased above threshold. In the following, we provide a detailed analysis of this multiplier circuit and compare the results with measured data.

First, we consider the multiplier core of the circuit that is identical with the original Gilbert multiplier. We assume all three differential pairs to operate above threshold. The input to the multiplier is given by the differential voltages $\Delta V_A = V_1 - V_2$ and $\Delta V_B = V_3 - V_4$, respectively. From Appendix D we know that the saturation currents in the two legs of the lower differential pair can be described as

$$I_1 = \frac{\beta\kappa}{8}\left(\Delta V_A + \sqrt{\frac{4I_{b1}}{\beta\kappa} - {\Delta V_A}^2}\right)^2 \quad \text{and} \quad I_2 = \frac{\beta\kappa}{8}\left(-\Delta V_A + \sqrt{\frac{4I_{b1}}{\beta\kappa} - {\Delta V_A}^2}\right)^2 .$$

These currents are the bias currents for the upper two differential pairs. Thus we find the total differential output currents to be

$$I_+ = I_{1,3} + I_{2,4} = \frac{\beta\kappa}{8}\left[\left(\Delta V_B + \sqrt{\frac{4I_1}{\beta\kappa} - {\Delta V_B}^2}\right)^2 + \left(-\Delta V_B + \sqrt{\frac{4I_2}{\beta\kappa} - {\Delta V_B}^2}\right)^2\right]$$

and

$$I_- = I_{1,4} + I_{2,3} = \frac{\beta\kappa}{8}\left[\left(-\Delta V_B + \sqrt{\frac{4I_1}{\beta\kappa} - {\Delta V_B}^2}\right)^2 + \left(\Delta V_B + \sqrt{\frac{4I_2}{\beta\kappa} - {\Delta V_B}^2}\right)^2\right]$$

The total output current of the multiplier core then simplifies to

$$I_{out}^{core} = I_+ - I_- = \frac{\beta\kappa}{2}\Delta V_B\left(\sqrt{\frac{4I_1}{\beta\kappa} - {\Delta V_B}^2} - \sqrt{\frac{4I_2}{\beta\kappa} - {\Delta V_B}^2}\right) \tag{5.11}$$

Yet, Equation (5.11) does not seem very intuitive to understand. We can derive a more imaginative form if we expand both square-roots in (5.11) around ΔV_B^2, according to the Taylor series

$$\sqrt{a - x} \;\approx\; \sqrt{a} - \frac{x}{2\sqrt{a}} - \frac{x^2}{8a\sqrt{a}} \;\cdots$$

Neglecting second and higher order terms, we find the following approximation:

$$I_{out}^{core} \approx \frac{\beta\kappa}{\sqrt{2}}\,\Delta V_A \Delta V_B\left[1 + \frac{{\Delta V_B}^2}{\frac{4I_{b1}}{\beta\kappa} - 2{\Delta V_A}^2}\right] . \tag{5.12}$$

Equation (5.12) illustrates more clearly the behavior of the multiplier core. In a first approximation the output current is proportional to the product of the two differential

input voltages ΔV_A and ΔV_B which is what we require. However, the term in brackets is not constant because it depends on the input. According to the input limit for the differential pair (Appendix (D.13)), the denominator has to remain positive, thus in the interval $[0, 4I_{b1}/\beta\kappa]$. The complete term in brackets is always > 1 and is monotonically increasing with increasing inputs ΔV_A and ΔV_B. As a consequence, the multiplier circuit shows a *super-linear* behavior, at least as long as all transistors are biased above threshold. We expect the effect to be pronounced for changes in the input voltage at the upper differential pairs because the numerator changes directly with ΔV_B^2.

Before we examine the (super-)linear behavior more closely, we first consider the rest of the circuit and solve for the total output current I_{out}. According to Figure 5.5, the currents I_+ and I_- from the multiplier core induce a logarithmic voltage drop across the diodes D_1 and D_2. These voltages form the input to the outer differential pair, biased in the sub-threshold regime according to the bias voltage V_{b2}. Given the general diode current (Shockley-equation) as

$$I_{diode} = I_0(\exp\,(\frac{1}{n}V_d\frac{q}{kT}) - 1) \tag{5.13}$$

and considering the characteristics of the differential pair in sub-threshold operation[3], we find the the total output current of the multiplier circuit to be

$$I_{out} = I_{out+} - I_{out-} = I_{b2}\tanh\Big(\kappa n\frac{\log\left((I_+ + I_0)/(I_- + I_0)\right)}{2}\Big)\ . \tag{5.14}$$

For typical values of κ and n, I_{out} is some exponential function in the above-threshold output currents of the multiplier core as illustrated in Figure 5.6. However, if we were to assume $\kappa n \approx 1$, equation (5.14) would simplify to

$$I_{out} \approx I_{b2}\ \frac{I_{out}^{core}}{I_+ + I_- + 2I_0}. \tag{5.15}$$

Since I_0 is very small compared with the above-threshold bias current I_{b1}, we could safely neglect it and finally would find

$$I_{out} \approx \frac{I_{b2}}{I_{b1}}\ I_{out}^{core}\ . \tag{5.16}$$

Hence, the total output current of the multiplier circuit would be the normalized output of the multiplier core running above threshold, scaled by the bias current I_{b2}. In general, however, κ of the outer nFET's and the ideality factor $1/n$ of the diodes do not match. Nevertheless, we show in the following that the assumption of linear current scaling (5.16) is valid when choosing an ideal implementation of the diodes D_1 and D_2.

[3] see Appendix D

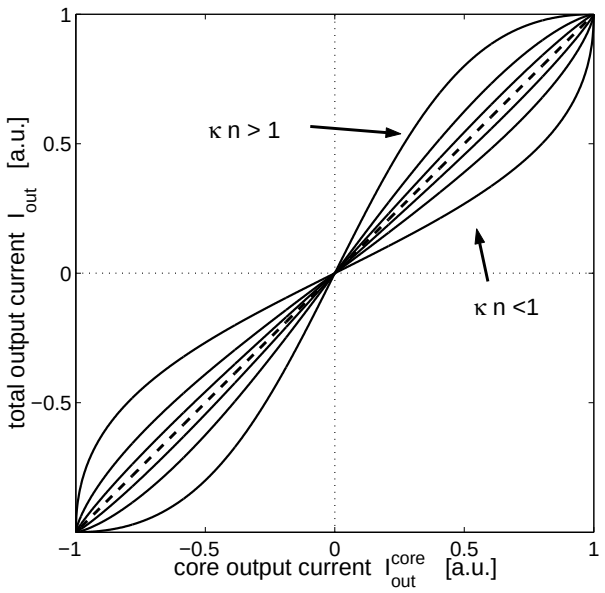


Figure 5.6: *Scaling characteristics of the output current of the multiplier circuit as a function of κn.* The normalized scaling characteristics of the core output current I_{out}^{core} to the total output current I_{out} is displayed for different values of κn = [0.5, 3/4, 9/10, 1 10/9, 4/3, 2]. The curves are computed according to (5.14) with $I_{b1}/I_{b2} = 1$ where the differential currents I_+ and I_- are swept completely $[0 \ldots 1]$ in each case. The dashed line shows the ideal, linear scaling ($\kappa n = 1$).

The voltage drop across the diodes D_1 and D_2 is such that the gate voltages of the outer nFET's are typically within 1 V below V_{dd}, meaning that the gate-bulk potentials are large. In this voltage range, κ asymptotically approaches 1 because the capacitance of the depletion layer becomes negligibly small compared to the gate-oxide capacitance[4]. Thus we can assume $\kappa \approx 1$ and equal for both nFET's. Furthermore, the emitter current of a bipolar transistor shows an ideal diode behavior ($n = 1$) as a function of the emitter potential when connecting base and collector to a common reference potential [Gray and Meyer 1993]. This ideal behavior has also been verified by measuring the emitter currents of diode-connected native BiCMOS and vertical CMOS bipolar transistors (data shown in Figure 5.9). Deviations occur only at high current densities when high-level carrier injection becomes relevant. Therefore, implementing D_1 and D_2 as diode-connected npn-bipolar transistors allows us to assume $\kappa n \approx 1$ and thus (5.16) to be a valid approximation for the total output current of the wide linear-range multiplier.

Figure 5.7 shows the measured output characteristics of the wide linear-range multiplier for varying input voltages ΔV_A and ΔV_B respectively. At the macroscopic level, the output current is typically a sigmoidal function of the input voltage. However, a closer

[4] see Appendix C

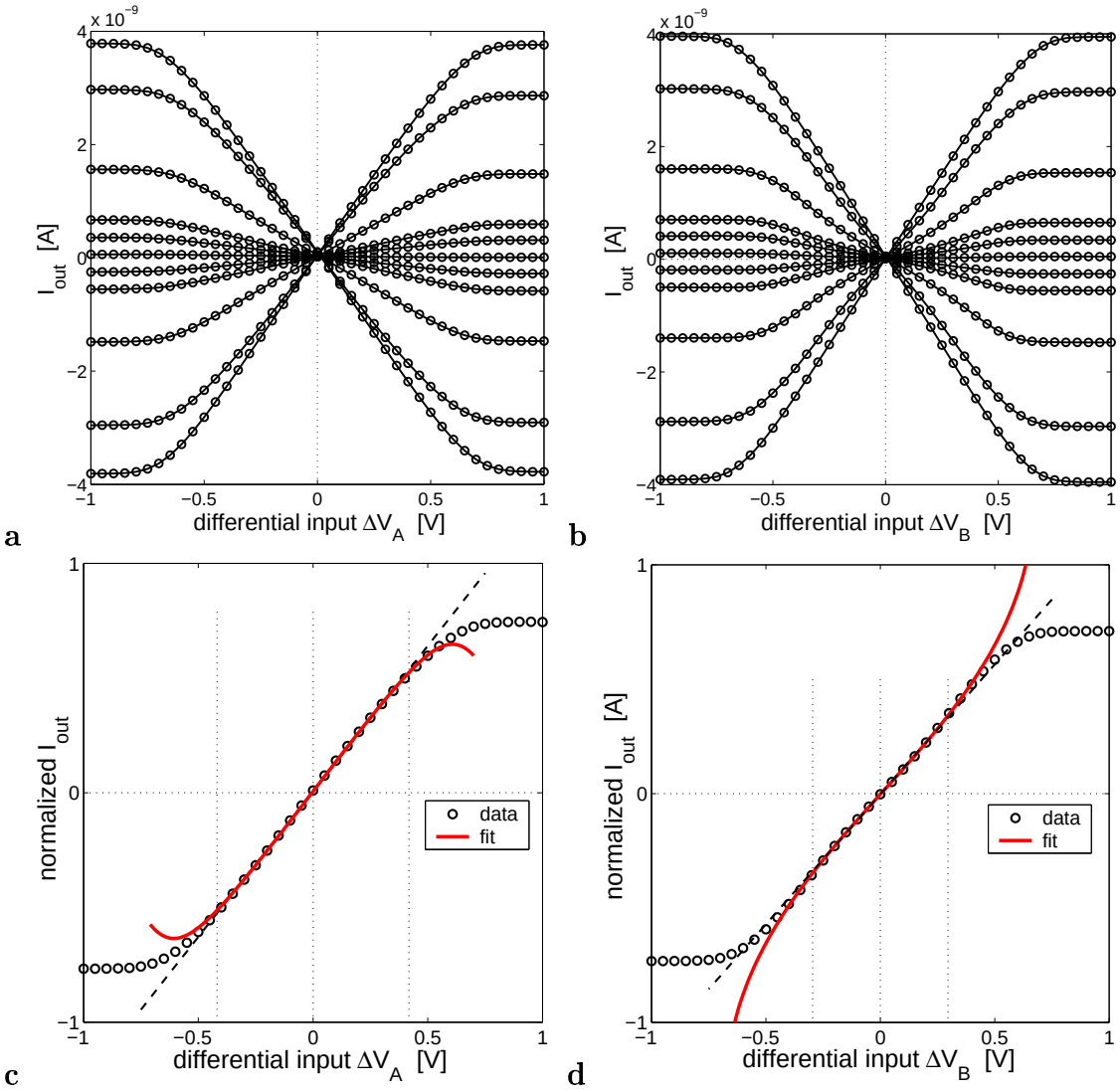


Figure 5.7: *Output characteristics of the wide linear-range multiplier.* (a) Measured output currents of the wide linear-range multiplier as a function of the applied input voltage at the lower differential pair ΔV_A. (b) The same measurement but this time sweeping the input voltage at the upper differential pairs ΔV_B. In both figures, each curve represents the output current for a fixed differential voltage ΔV (0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.75 V) applied at the non-swept input. The bias voltages were $V_{b1} = 1.2$ V and $V_{b2} = 0.725$ V respectively. (c) and (d) Single traces for sweeping ΔV_A and ΔV_B respectively are shown, together with a least-square fit according to Equation (5.11) (bold line) and a linear fit (dashed line). Equation (5.11) characterizes accurately the measured output currents within the voltage range that obeys the input limits (vertical dotted lines). Note the small super-linear behavior of the output current for sweeping ΔV_B.

examination reveals the predicted super-linearity in the response curves. The effect is almost not noticeable when varying ΔV_A, but is significant when changing input ΔV_B as shown in Figure 5.7d. The measured behavior nicely matches the fitting functions according to the derived expression (5.11) and it also supports the linear scaling of the output currents to sub-threshold level according to (5.16). Of course, the behavior is only described accurately as long as all transistors in the multiplier core are saturated and biased in the above-threshold regime.

Operational limits

To keep all transistors in saturation, we have to define some lower limits for the input voltages. Similar considerations as for the simple differential pair circuit (Appendix D) lead to the following *saturation limits* of the multiplier circuit:

$$\begin{aligned} \max(V_1, V_2) &> V_b + (V_b - V_{T0}) \qquad \text{and} \\ \max(V_3, V_4) &> \max(V_1, V_2) + (V_b - V_{T0}) \; . \end{aligned} \tag{5.17}$$

These are important operational limits that the design of the single motion unit has to account for (see Figure 5.4). As an immediate consequence, a unity-gain source follower is added which allows the photoreceptor output voltages to be increased sufficiently to obey (5.17); because they form the input to the upper differential pairs (V_3, V_4) of the multiplier circuits. The input to the lower differential pair is given by the differential voltage $U_+ - V_{ref}$ and $V_+ - V_{ref}$ respectively. Since V_{ref} is controllable, it can be set such that it meets condition (5.17).

In order to keep all transistors of the multiplier circuit core in above-threshold operation, the maximal differential input voltages are limited. We find the following *input limits*

$$\Delta V_A < (V_{b1} - V_{T0}) \qquad \text{and} \qquad \Delta V_B < \frac{\sqrt{2}}{2}(V_{b1} - V_{T0}). \tag{5.18}$$

We note that the derived limits (represented by the dotted vertical lines in Figures 5.7c,d) correspond well with the measured voltage range within which Equation (5.11) describes accurately the output currents.

Linear range

The linear range of the presented multiplier circuit is almost an order of magnitude larger than the one of the standard Gilbert multiplier running in sub-threshold. Figure 5.8a displays the linear range for both differential input stages as a function of the bias current I_{b1}, where the linear range is defined as the input voltage where the measured output

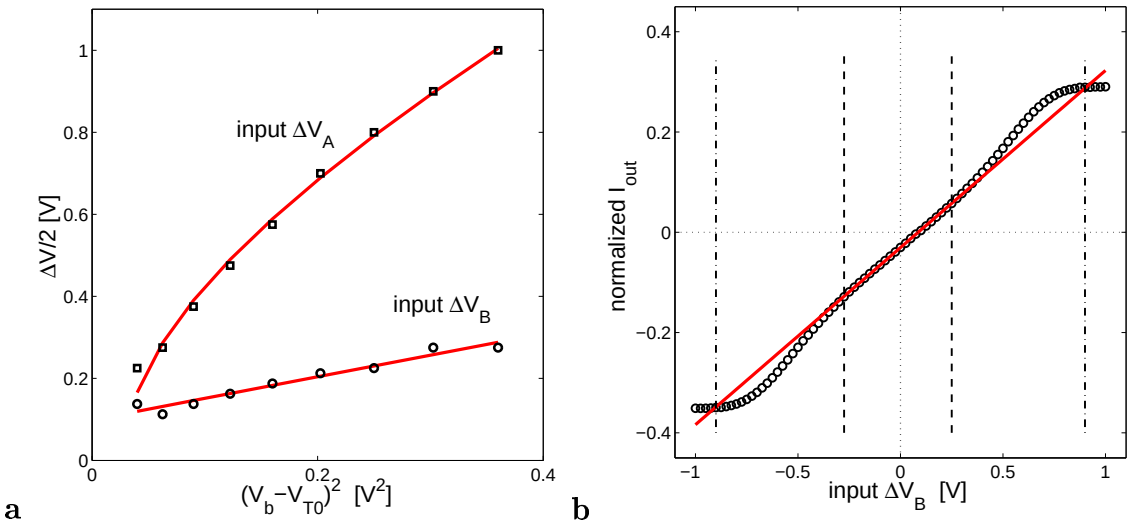


Figure 5.8: *Linear and saturation range of the wide linear-range multiplier.* (a) The linear range is shown as a function of the bias current I_{b1} for both input stages. Linearity is massively increased in comparison to the sub-threshold regime. The linear range grows inversely proportionally to the transconductance of the multiplier circuit, as illustrated by the least-square fit based on (5.19) and (5.20). Due to the super-linear behavior, the linear range is smaller for variable input ΔV_B. In practice however, saturation is of much greater importance because it signals the transition point where the multiplier output changes drastically its behavior. (b) The output current as a function of input ΔV_B clearly shows the super-linear behavior ($V_{b1} = 1.4$ V). The range of linearity (dashed lines) is significantly smaller than the range where the output is not saturating (dash-dotted lines).

current deviates by less than 5% from a linear fit through the origin. Due to the expansive non-linearity, the linear range for the input ΔV_B is much smaller as shown in Figure 5.8b.

We can estimate the size of the linear range as a function of the bias current of the multiplier core. The maximal slope of the normalized output currents for given inputs ΔV_A, ΔV_B is proportional to the transconductance. Under the assumption that the output currents represent similar sigmoidal functions, the linear range is inversely proportional to the transconductance of the multiplier. The two fitting functions, based on the inverse of the transconductances (5.19) and (5.20), describe fairly well the measured linear ranges as shown in Figure 5.8a and prove that above assumption is valid.

In the following, the transconductances of the multiplier circuit will be determined for each of the two differential input voltages. Differentiating (5.16) with respect to ΔV_A and ΔV_B respectively leads to

$$g_{mA} = \frac{\partial I_{out}}{\partial \Delta V_A}\Big|_{\Delta V_A \to 0, V_B = const.} = \frac{1}{\sqrt{2}} \frac{I_{b2}}{I_{b1}} \beta\kappa\Delta V_B \sqrt{\frac{I_{b1}}{I_{b1} - \beta\kappa\Delta V_B^2/2}} \tag{5.19}$$

and

$$g_{mB} = \left.\frac{\partial I_{out}}{\partial \Delta V_B}\right|_{\Delta V_B \to 0, V_A = const.} = \frac{1}{\sqrt{2}} \frac{I_{b2}}{I_{b1}} \beta \kappa \Delta V_A \ . \tag{5.20}$$

The effect of the expansive non-linearities on the motion output of a single motion unit is much less significant than the impact of saturation of the multiplier observed for large inputs. Saturation starts when the inputs ΔV_A, ΔV_B exceed the input limits (5.18)[5]. The input range for which the multiplier is not in saturation, is almost equally large for both inputs and slightly beyond the linear range found for ΔV_A.

Diode implementation

In principal, base-emitter junctions can be exploited either using native bipolar transistors in a genuine BiCMOS process or the vertical bipolar transistors in standard CMOS technology. Vertical bipolar transistors are intrinsically present in all well-type MOSFETs as schematically shown in Figure 5.9b. The use of vertical bipolars is insofar limited as they have a common collector which is the substrate. As a consequence, only one type of vertical bipolars is available depending on the substrate doping (e.g. only pnp-bipolars with a p-doped substrate) which would require to invert the complete multiplier circuit[6].

It is important that the bipolar implementation of the diodes D_1 and D_2 guarantees an ideal behavior over a sufficiently large above-threshold current range. To test this we measured the base-emitter voltage V_{BE} as a function of the applied emitter current I_E for both, a vertical pnp-bipolar transistor in a typical p-substrate CMOS process and a native npn-bipolar transistor in a genuine BiCMOS process. As shown in Figure 5.9c, the two emitter currents depend exponentially on the base-emitter potential for low current values. At current levels above 1 μA, the vertical bipolar starts to deviate significantly from its exponential characteristics due to high-level injection that is caused by the relative light doping of the well (base) [Gray and Meyer 1993]. Nonetheless this occurs already at a current level that is below the range where the multiplier core is preferably operated at. The exponential regime of the native bipolar, however, continues up to 0.1 mA. Although a CMOS implementation of the diodes would be preferable to avoid the more complex and expensive BiCMOS process, that approach is inconsistent with a correct operation of the multiplier circuit. Figure 5.9d shows that the standard deviation between different devices of both types of bipolar transistors reveals only a small current mismatch in the order of 1%.

[5] see also Figure 5.8b (dashed lines)

[6] i.e. replace all native with well-transistors and switch the rails

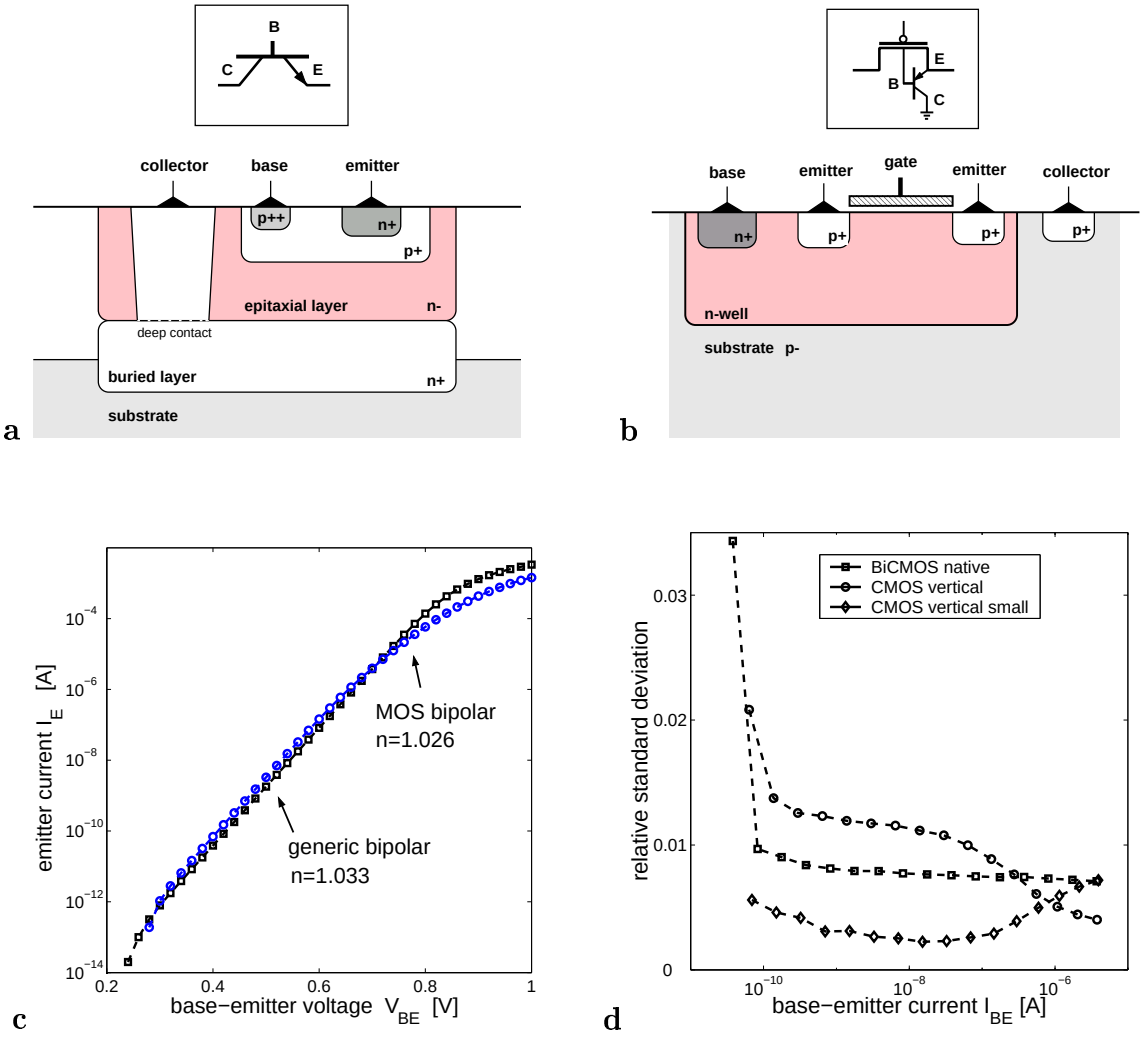


Figure 5.9: *BiCMOS bipolar versus vertical CMOS bipolar transistor.* (a) Cross-section of a native npn-bipolar transistor. BiCMOS processes usually use a highly (n+) doped buried layer to achieve low collector resistivity and thus fast switching. (b) In standard CMOS technology, the intrinsic vertical pnp-bipolar can be activated within a normal pFET. Considering either drain or source as the emitter, the well serves as base and the substrate as the common collector. Obviously, this heavily constrains the use of multiple vertical bipolars. (c) The measured emitter currents for the native npn-bipolar and the vertical pnp of a pFET are shown. In both cases base and collector are shorted and the emitter potential is moved. Note that the deviation from the exponential characteristics for the pFET bipolar occurs already at 1μA, thus below the range the wide linear-range multiplier is preferably biased. This deviation is mainly caused by high-level injection due to the relatively lightly doped well (base). (d) The measured relative standard deviation (n=6) for the native bipolar ($A_E = 4\,\mu\text{m}^2$) and for two pFETs with two different emitter sizes $(4\,\mu\text{m})^2$ and $(2.6\,\mu\text{m})^2$. Data is given for currents within the exponential regime of the vertical bipolar.

Non-linearities in the feedback loop

The output current of the multiplier circuit saturates for large input voltages. What does this mean in terms of the expected motion output of a single unit of the optical flow chip?

For the sake of illustration, we once more assume a one-dimensional array of the optical flow network. Furthermore, we disable all lateral connections and neglect the bias constraint ($\rho = 0, \sigma = 0$). Then, for given spatiotemporal brightness gradients, the circuit ideally satisfies the brightness constraint

$$E_x u + E_t = 0 \ . \tag{5.21}$$

Now, assume that the multiplication $E_x u$ is replaced by the saturating function $f_{E_x}(u)$ with E_x held constant, where f describes the output characteristics of the proposed multiplier circuit and E_x and u represent its input voltages[7]. Since f is one to one, we can solve for the motion response

$$u = \ f^{-1}(-E_t) \ . \tag{5.22}$$

Unfortunately, Equation (5.11) only describes the multiplier circuit accurately if all its transistors are running above threshold which is, in particular, not valid in the saturation region of the multiplier. For reasons of simplicity and because we are only interested in the qualitative effect of the saturation behavior, we assume the output I_{out}^{core} of the multiplier core to follow a simple sigmoidal function $\tanh(u)$ for any given E_x. Considering the total output current of the wide linear-range multiplier (5.16), we now rewrite the saturating function $f_{E_x}(u) = c_{E_x} \ I_{b2}/I_{b1} \tanh(u)$, where the constant c_{E_x} is according to E_x, and find

$$u = -\text{artanh}(\frac{I_{b1}}{I_{b2}} E_t c_{E_x}^{-1}) \ . \tag{5.23}$$

Figure 5.10 illustrates the expected motion response according to (5.23) for different values of the bias current I_{b2}. We see that the motion output overestimates the stimulus velocity the more the multiplier circuit starts to saturate and finally ends up at the rails! The more the output of the multiplier circuit is below the true multiplication result, the more the feedback loop raises u in order to make the brightness constraint (5.21) hold.

Considering the correct characteristics of the real multiplier circuit, the response will look different from the curves in Figure 5.10 insofar as the linear range is increased and the saturation transition is sharper and more pronounced. Nevertheless, Figure 5.10 illustrates the qualitative response behavior nicely. Increasing I_{b2} decreases the slope of the motion response. Thus, the bias current of the outer differential pair acts as **gain control** that allows, *e.g.* the linear range of the motion response to be adjusted to the expected motion range so obtaining the highest possible signal-to-noise ratio.

[7] According to the notation in the schematics of the optical flow unit (Figure 5.4), the inputs to the multiplier circuit are the differential voltages $E_x = Ph_{x+} - Ph_{x-}$ and $u = U_+ - V_{ref}$.

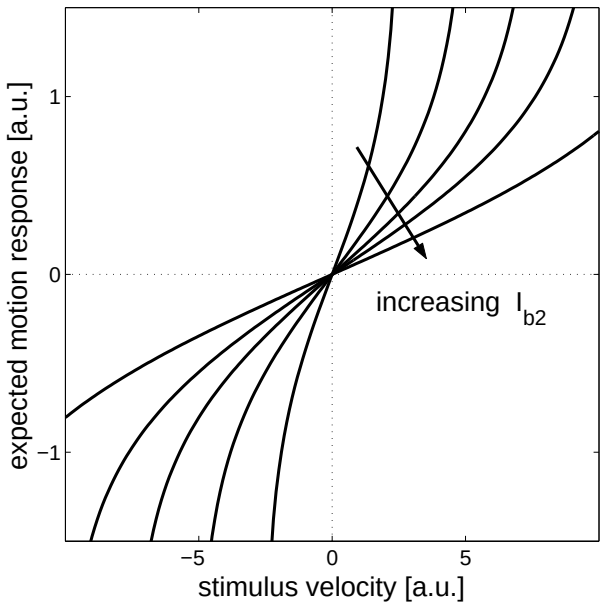


Figure 5.10: *Expected motion response due to saturation of the multiplier circuit.* The figure shows qualitatively the expected speed tuning curves according to Equation (5.23). Note that the more the function f saturates, the more expansive is the motion response curve. The different curves correspond to different values of the bias current I_{b2}. The bias voltage V_{b2} controls the response gain and thus allows to adjust it to the velocity range of the expected visual motion.

The presented multiplier circuit provides compact four-quadrant multiplication where the linear/saturation range and the multiplication gain can be separately controlled by two bias voltages V_{b1} and V_{b2}. The disadvantages are the increased power consumption caused by the above-threshold operation and the need for BiCMOS process technology to implement diodes showing an ideal behavior over a sufficiently large current range.

5.2.3 Cascoded current mirror and the effective bias constraint

As shown in the schematics of the single motion unit (Figure 5.4), the *bias constraint* is implemented by a transconductance amplifier configured as unity-gain follower. Instead of an ohmic current (5.10), it generates a saturating leak current I_B that is a function of the voltage difference between the voltage $U+$ and $V+$ respectively, and V_{ref}, the virtual ground or zero-motion level in the circuit. The bias constraint transforms from a quadratic to an absolute-value function for increasing voltage differences[8]. The saturation current and therefore the strength of the bias conductance is controlled by the bias voltage Bias_{OP} of the transconductance amplifier. Consider the current equilibrium at the capacitive

[8]see also previous discussion in Section 4.2.1

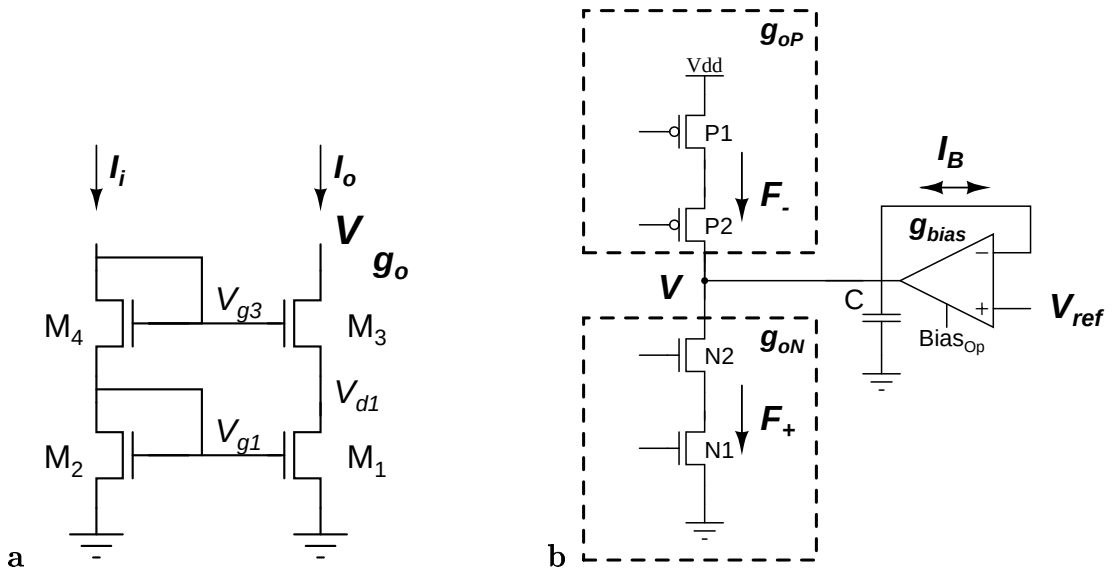


Figure 5.11: *Cascoded current mirror and the effective bias conductance* (a) A n-type cascoded current mirror. (b) The effective bias conductance at the capacitive feedback node is the combination of the output conductances of the current mirrors g_{oN} and g_{oPS} and the transconductance of the operational amplifier g_{bias}.

nodes C_u, C_v in the single motion unit. In steady state, we assumed all currents onto these nodes to represent the different constraints according to the dynamics (4.19). To be completely true, this would require the feedback current to be generated by ideal current sources. However, the small but present output conductances of the feedback nodes induce a deviation from such an ideal model.

Figure 5.11b shows one of the feedback nodes used in the single motion unit. According to the figure, an incremental increase of V requires the total correction current $F_+ - F_-$ to be larger than actually defined by the constraint dynamics by an amount that is proportional to the sum of the conductances g_{oN} and g_{oP}. The extra current has to be provided by the feedback loop and thus alters the computational result. Since this current depends proportionally on the voltage V, we can understand it also as the *second bias current* in addition to the current through the transconductance amplifier. It biases the capacitive node to a reference voltage, however, that cannot be controlled and that depends on various issues like the strength of the currents F_+ and F_- or the Early voltages of the transistors. In general, this reference voltage is not identical with V_{ref}. Thus, the superposition of the two bias currents has an asymmetric influence on the final motion estimate. Since the total correction currents are typically weak, the effect is significant and cannot be neglected. It is therefore necessary to reduce the output conductances as much as possible.

A cascoded current mirror circuit as shown in Figure 5.11a reveals a significant decrease

in output conductance compared with a simple current mirror. Using the notation as in Figure 5.11a and neglecting the junction leakage in drain and source, we find the output conductance to be

$$g_o = g_{o1}\frac{g_{ds3}}{g_{ms3}} \ , \tag{5.24}$$

where g_{ds3} and g_{ms3} are the drain- and transconductance of transistor M_3 and g_{o1} is the output conductance of transistor M_1. We immediately see, that the decrease in output conductance compared to g_{o1} (the output conductance of a simple current mirror circuit) is in the order of a factor

$$\frac{g_{ds3}}{g_{ms3}} = \frac{U_T}{V_{E,M3}} \approx \ 250 \ldots 500 \ , \tag{5.25}$$

where U_T is the thermal voltage kT/q and $V_{E,X}$ denotes the Early voltage of transistor X. Applying the same analysis to the p-type current mirror, the final output conductance of the feedback node (see Figure 5.11b) becomes

$$g_o = g_{oN} + g_{oP} = \frac{F_+ \cdot \ U_T}{V_{E,N1} \ V_{E,N2}} + \frac{F_- \cdot \ U_T}{V_{E,P1} \ V_{E,P2}} \ . \tag{5.26}$$

The output conductance depends equally strongly on both transistors in the cascoded current mirror. Since the Early voltage is proportional to the length of a transistor[9] both transistors should be long in order to reduce the output conductance. If restrictions in implementation area do not allow this, then the length of the mirror transistor preferably has to be increased at the expense of the length of the cascode transistor because of matching reasons.

The total output conductance depends also on the strength of the correction currents F_+ and F_-. The correction currents are mainly determined by the bias voltage Bias_{VVI2} of the multiplier circuit. We have previously shown that increasing this bias voltage allows to adjust the motion output range in order to detect higher visual speeds. However, a higher total output conductance also pronounces the unwanted second order effects (asymmetry). As a consequence, g_{bias} has to be sufficiently large to counter-balance these asymmetric influences and allow the implementation of the exact bias constraint. From an algorithmic point of view, however, we want to keep the strength of the bias constraint minimal to achieve a best possible optical flow estimate. Thus again, it is very important to minimize the output conductances of the feedback nodes.

5.2.4 Implementation of the smoothness constraint

The implementation of the smoothness constraint consists of two resistive networks of saturating resistors. The general characteristics of these elements has been discussed previously[10] where we have shown that their non-ohmic conductances reduce smoothing

[9] see Appendix C

[10] see Section 4.2.1

across large voltage differences. This anisotropic smoothing process allows motion integration to take place preferably in areas of common motion. The cost function associated with the smoothness constraint reduces to the absolute-value function for voltage differences that are in the saturation region of the saturating resistors. The resulting optical flow estimation is enhanced compared to the case of using ohmic conductances[11]. Interestingly, it is by far more convenient to implement saturating resistive elements in CMOS than ohmic resistors that are adjustable.

The saturating resistances are implemented using the Horizontal Resistor (HRes) circuit proposed by Mead [1990]. As shown in the schematics in Figure 5.4, a single transconductance amplifier with an additional diode-connected transistor in the output leg and four pass transistors are required to implement the resistive connections to the four neighboring units. For voltage differences within the small-signal linear range of the amplifier, the conductance is proportional to the bias current controlled by the gate voltage Bias_{HR}. Above the linear range, the current through the pass-transistor saturates and is equal to half of the bias current. Thus, the bias voltage Bias_{HR} sets the saturation current globally whereas the smoothing is anisotropic and **locally controlled** by the instantaneous voltage distribution in the resistive networks.

5.3 Performance of the Optical Flow Chip

In the following, the measured behavior of the optical flow chip to moving real-world stimuli is presented and discussed. For all measurements the stimuli were projected directly onto the chip via an optical lens system. The stimuli were either generated electronically and displayed on a computer screen, or were physical moving objects *e.g.* printed patterns on a rotating drum. The wide-range characteristics of the adaptive photoreceptor makes the optical flow chip relatively insensitive to the absolute irradiance level. The measured on-chip irradiance in all the experiments was within one order of magnitude with a minimal value of 9 mW/m^2 for very low contrast computer screen displays. At such low values, the rise time of the photoreceptor circuit is in the order of a few milliseconds [Delbruck 1993b]. This is sufficiently fast considering the applied stimulus speed in those particular measurements. The bias voltages are named according to the labels provided in the schematics shown in Figure 5.4. Unless indicated differently, the bias voltages were all sub-threshold and were kept constant at their mentioned standard values. The measured output signals are always shown with respect to the potential voltage V_{ref}.

[11] as shown in Figure 4.11

5.3.1 Characterization of the motion circuit

Moving sinewave and squarewave gratings were applied to characterize the motion response for varying speed, contrast and spatial frequency of the gratings. Furthermore, the orientation tuning of the optical flow chip and its ability to report the intersection-of-constraints estimate of visual motion are tested. In order to increase the robustness of these measurements, the presented data constitutes the *global* motion signal, thus the unique, collectively computed solution provided by the 10x10 units of the optical flow chip. Although we measured the response of single isolated units to be very similar to the shown global responses they were less robust. In particular, pronounced mismatch effects require an increased weight of the bias constraint that deteriorates the output response with respect to its sensitivity to contrast and spatial frequency. The impact of mismatch will be discussed in a later section.

Speed tuning

Figure 5.12 shows the time-averaged response of the optical flow chip to a moving sinewave grating of high contrast (80%) and low spatial frequency (0.08 cycles/pixel). Each data point represents the mean value of 20 single measurements, each being the output voltage time-averaged over one stimulus cycle. As a consequence, the measured standard deviation is rather small. Only the component of the optical flow vector is displayed that is orthogonally oriented to the moving direction of the grating.

The response is almost linear in stimulus speed within the voltage range that is equivalent to the non-saturated range of the multiplier circuit. For the given bias strength of the multiplier core ($\text{Bias}_{VVI1} = 1.1$ V), the linear range is approximately ± 0.5 V as indicated by the dash-dotted lines. Beyond the linear range, the response quickly increases/decreases and finally hits the voltage rails on either side (not shown), as predicted[12]. Within a small range around zero motion, the optical flow output slightly underestimates the stimulus velocity.

The two graphs in Figure 5.13 show each the different speed-tuning curves as a function of the bias voltage Bias_{VVI2} for a moving sinewave and a squarewave grating of equal contrast and spatial frequency respectively. As expected, increasing Bias_{VVI2} leads to a smaller response gain and thus expands the linear output range. The solid lines represent the linear fits to the individual measurements. The slope of these lines roughly scales by a factor of one half for a fixed increase in bias voltage by 30 mV. Therefore, the gain is inversely proportional to the bias current as we predicted according to Equation (5.23). The bias voltage Bias_{VVI2} allows to optimally adjust the limited linear operational range of the optical flow chip to the speed of the expected visual motion. It also allows recalibration if scale changes are applied in the optical pathway. The slope values for the

[12] compare also with Figure 5.10

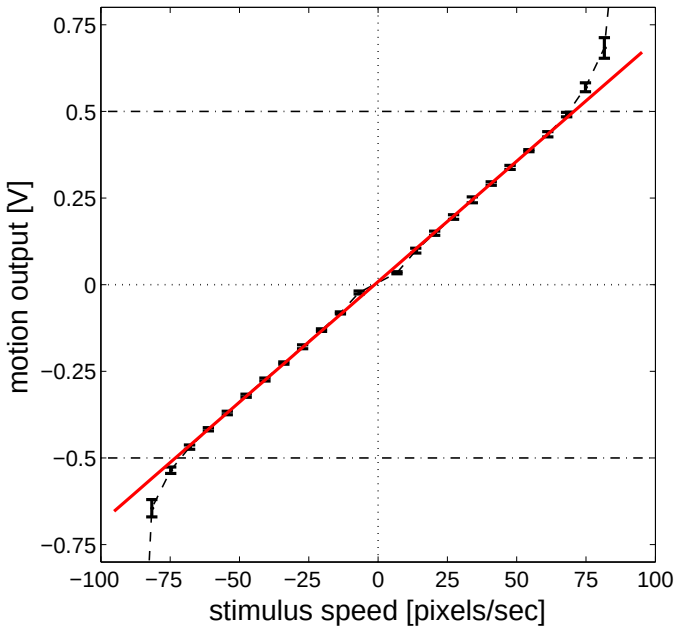


Figure 5.12: *Speed tuning of the optical flow chip.* The measured time-averaged speed tuning of the optical flow chip to a moving sinewave grating is shown (spatial frequency 0.08 cycles/pixel, contrast 80%). The tuning curve follows qualitatively the predicted response, showing the expanding non-linearity due to the saturation of the multiplier circuits.

two different stimuli are almost identical (see the insets in Figure 5.13). This suggests that the speed tuning is not dependent on the spatial frequency spectrum of the stimulus as long as the fundamental frequency is the same.

Figure 5.14 reveals a closer look to both, the low and high speed limits of operation. For appropriate values of Bias_{VVI2}, the range of reliable speed estimation spans **three orders** of magnitude, from 1 pixel/sec up to at least 3500 pixels/sec. Higher visual speed values are likely to be measured reliably but were not possible to be produced by the used test-setup. In principle, the upper bound of accurate speed estimation is given by the low-pass filter cut-off frequency of the photoreceptor and/or the hysteretic differentiator.

Very low speed values are difficult to be estimated correctly; even the direction of motion can be confounded (Figure 5.14a). Given the low output currents from the wide linear-range multiplier due to its low bias voltage $\text{Bias}_{VVI2} = 0.35$ V and considering the very low currents from the temporal differentiator for such small stimulus speeds, the total differential currents at the capacitive feedback nodes are also low and dynamic second order effects dominate, modulated by the slowly varying photoreceptor outputs. However, **zero motion** is reliably detected since the input is constant and the slow time constant induced by the weak bias constraint ($\text{Bias}_{OP} = 0.30$ V) is not expressed in the time-

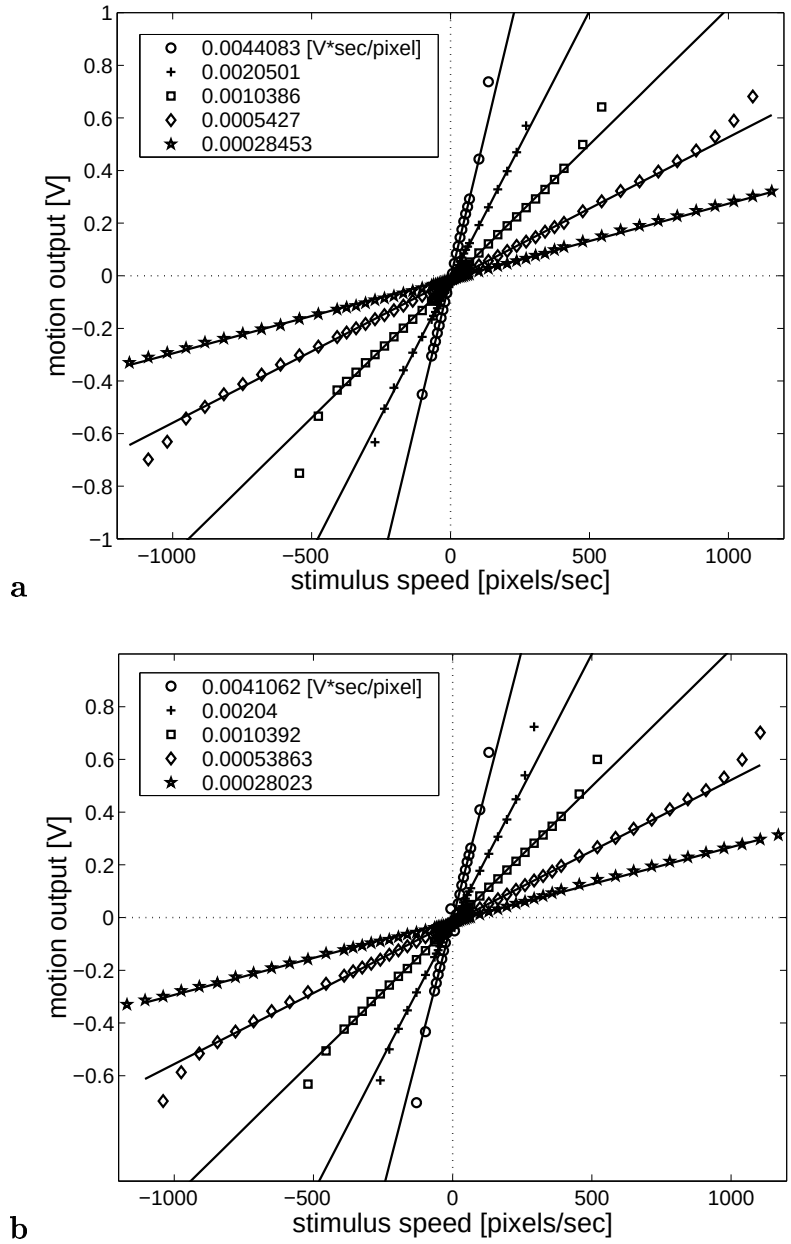


Figure 5.13: *Motion response gain.* (a) The measured time-averaged motion response of the optical flow chip for a moving sinewave grating is shown as a function of the multiplier bias voltage Bias_{VVI2} (0.5, 0.53, 0.56, 0.59, 0.62 V). The voltage gain decreases exponentially in the bias voltage (see insets). Thus, the linear range of the response can be adjusted to the speed range of the visual stimulus. (b) The same measurements for a squarewave grating of same spatial frequency and contrast. The comparison of the measured slope values for the two stimuli reveals an almost identical response behavior.

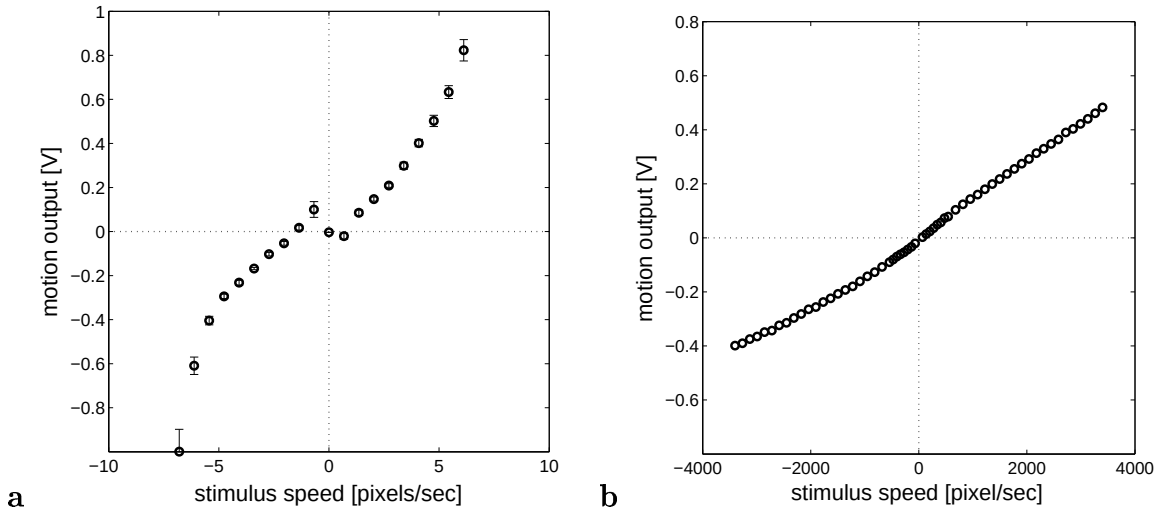


Figure 5.14: *Speed detection limits of the optical flow chip.* (a) The measured time-averaged motion response of the optical flow chip biased to detect very low speeds ($\text{Bias}_{VVI2} = 0.35$ V). The chip is able to reliably detect low speeds down to 1 pixel/sec. (b) Appropriately biased ($\text{Bias}_{VVI2} = 0.67$ V), high speeds up to 3500 pixels/sec were applied and could be measured reliably.

averaged measurements. Subsequentially, the deviation at low velocities is suppressed for higher current densities, that is higher values of Bias_{OP} and Bias_{VVI2}.

Contrast dependence

The output of the optical flow chip depends continuously on the contrast. The lower the contrast, the more dominant is the influence of the bias constraint, forcing the response towards the reference voltage V_{ref}. Figure 5.15a shows the output voltage as a function of stimulus contrast for a moving sinewave and squarewave grating stimulus of constant speed (30 pixels/sec) and identical spatial frequency (0.08 cycles/pixel). Each data point represents the mean value of 20 single measurements, each being the time-averaged output voltage over one stimulus cycle.

Below a critical contrast value of about 35%, the output signal decreases rapidly towards zero motion. The squarewave stimulus shows a slightly sustained resistance against the drop-off. A least-square fit is applied according to Equation (5.6) (solid lines). In particular, for the sinewave grating stimulus, the fit gets slightly worse around the critical contrast value because of the non-linear bias conductance implemented by the transconductance amplifier. The effective conductance decreases with increasing motion output signal while Equation (5.6) assumes a constant σ. As a consequence, the measured response curve rises faster and exhibits a flatter plateau than the fitting curve.

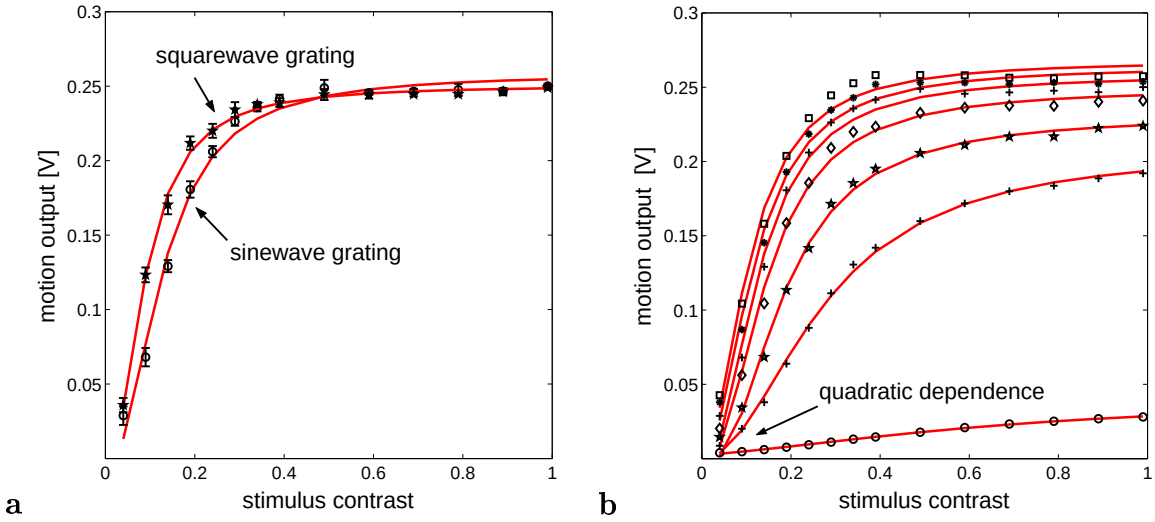


Figure 5.15: *Contrast dependence of the motion output.* (a) The motion output is measured as a function of stimulus contrast for sinewave and squarewave gratings of constant speed (30 pixels/sec) and spatial frequency (0.08 cycles/sec). The bias constraint forces the responses towards zero motion for low stimulus contrast. Especially for the sinewave grating, the measured response curves exhibits a sharper onset and a flatter plateau behavior than the fitting functions which is due to the non-ohmic bias conductance. (b) Increasing values of the bias voltage Bias_{OP} (0.15, 0.23, 0.26, 0.29, 0.32, 0.35, 0.45 V) lead to a stronger contrast dependence even at high contrast values, here shown for the sinewave grating stimulus. In the extreme case (bottom two curves), the output signal deteriorates and becomes proportional to the product of the spatial and temporal intensity gradient. Due to the early saturation of the second, simple Gilbert multiplier, the expected quadratic dependence is only noticeable at small signal levels (< 150 mV) and transfers above towards a linear function of contrast.

The influence of the bias conductance on the contrast tuning is shown in Figure 5.15b using the same sinewave grating stimulus as before. Increasing values of the bias voltage Bias_{OP} strongly decrease the output voltage at low contrasts and induce a contrast dependence even for high contrast values. For high values of the bias voltage, the bias conductance dominates the denominator in Equation (5.6). Thus the motion signal becomes quadratically dependent on contrast ($E_x E_t$). The motion estimation reduces to the multiplication of the spatial and temporal brightness gradients which simplifies the computational complexity drastically (see also Figure 4.3). In fact, such simple spatiotemporal gradient multiplication makes the costly feedback architecture of the circuit obsolete and can be implemented in a much more compact way [Horiuchi et al. 1994, Deutschmann and Koch 1998a]. The expected quadratic dependence can be observed in Figure 5.15b only at very low contrast levels, because the simple Gilbert multipliers responsible for

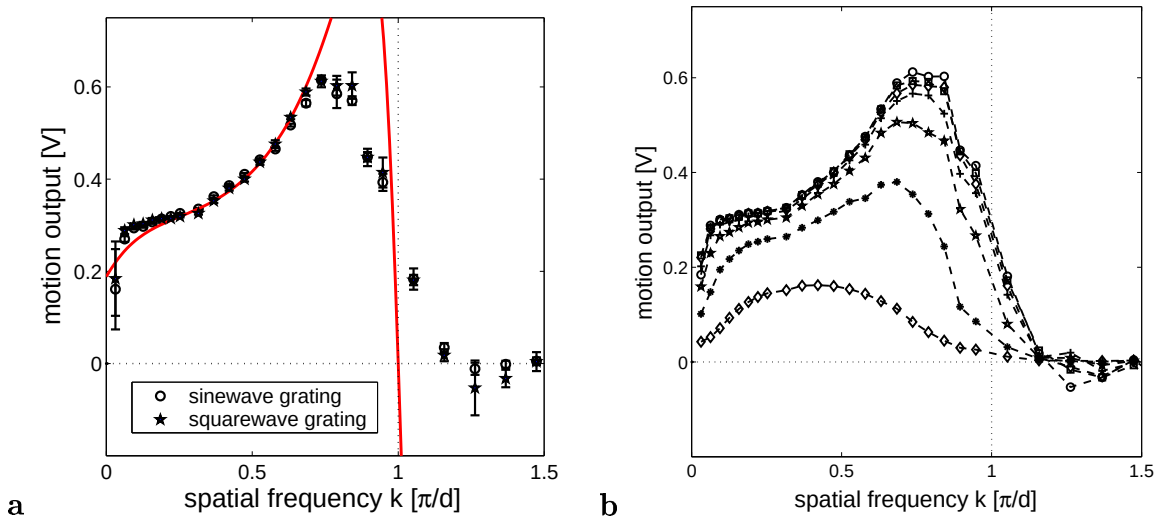


Figure 5.16: *Spatial frequency tuning of the optical flow chip.* (a) The spatial frequency response was measured for a sinewave and squarewave grating of constant speed (30 pixels/sec). For frequencies $k < 1$ $[\pi/d]$, the response follows closely the expected response (fit - solid curve). (b) Increasing the bias voltage Bias_{OP} (0.1, 0.15, 0.2, 0.25, 0.30, 0.35, 0.40 V) reduces the overall response and shifts the peak response towards the spatial frequency $k = 0.5$ $[\pi/d]$, where the spatial gradient and thus the local contrast is highest. Only the response to sinewave gratings is shown.

the product of the spatial and temporal gradient saturate early and lead then to a linear dependence on contrast.

Spatial frequency tuning

The third stimulus parameter tested was spatial frequency. Moving sinewave and squarewave gratings (80% contrast, 30 pixels/sec) of varying spatial frequency were presented to the chip. Figure 5.16a shows the motion output of the optical flow chip to sinewave and squarewave gratings as a function of increasing spatial frequency. Again, each data point represents the mean value of 20 single measurements, each being the time-averaged output voltage over one stimulus cycle. The spatial frequency k is given in units of the Nyquist frequency $[\pi/d]$.

The difference in waveforms does not have a significant effect onto the response. A least-square fit according to Equation (5.6) was performed on the response to the sinewave grating for frequencies $k < 0.75$. For spatial frequencies $k < 0.06$, the locally measured spatial gradients on the chip surface are very low and thus the bias constraint dominates and forces zero motion. As in the contrast measurements, we recognize a steeper slope of the response curve compared with the fit for frequencies below the critical value. Again, this is caused by the non-ohmic bias conductance. As a positive consequence, the response

shows an extended plateau behavior in the frequency range $0.06 < k < 0.3$. As spatial frequencies become high $k > 1$, the response curves rapidly drop towards zero and remain small. The deviation from the predicted response behavior, as shown in Figure 5.3 where large negative values occur as the spatial period drops below the double inter-pixel distance d, were induced by the non-optimal test setup that was not able to project a high spatial frequency stimulus with sufficiently high local contrast onto the chip surface.

The strength of the bias voltage Bias_{OP} affects the amplitude of the motion response as well as the spatial frequency for which the motion response is maximal as shown in Figure 5.16b. Recall that, increasing Bias_{OP} decreases the motion output the more, the smaller the local spatial gradient is. Because E_x is largest at $k = 0.5$ (see Figure 5.2) the spatial frequency for which the motion output is maximal moves more and more towards this value with increasing values of Bias_{OP}.

Orientation tuning

The ability of the optical flow chip to accurately estimate two-dimensional visual motion is superior to other reported hardware implementations. An orthogonal sinusoidal plaid pattern (spatial frequency 0.08 cycles/pixel, contrast 80 %) moving with constant speed (30 pixels/sec) was presented to the chip and the global motion response was measured for different directions-of-motion of the pattern. Each data point represents once more the mean value of 20 single measurements, each being the time-averaged output voltage over one stimulus cycle.

Figure 5.17 shows the applied plaid stimulus and the measured global motion components U and V of the optical flow chip as a function of the direction-of-motion. The output fits well the theoretically expected sine and cosine functions. Note that the least-square-error fit (solid lines) reveals a slight orientation offset of about 6°caused by the non-perfect calibration of the stimulus orientation with the orientation of the optical flow array.

Intersection-of-constraints solution

The optical flow chip is able to solve the aperture problem. To demonstrate this property, we presented a high contrast visual stimulus to the chip that consisted of a dark triangle on a light non-textured background, moving in different directions. The particular object shape requires the collective computation of the intersection-of-constraints solution in order to achieve the correct global motion estimate. The triangle was moving with a speed of 30 pixels/sec and the figure-ground contrast was 80%.

Figure 5.18 shows the applied stimulus and the global motion output of the chip for a constant positive or negative object motion in the two orthogonal directions. Each data point thereby represents the time-averaged global motion vector for a given strength of

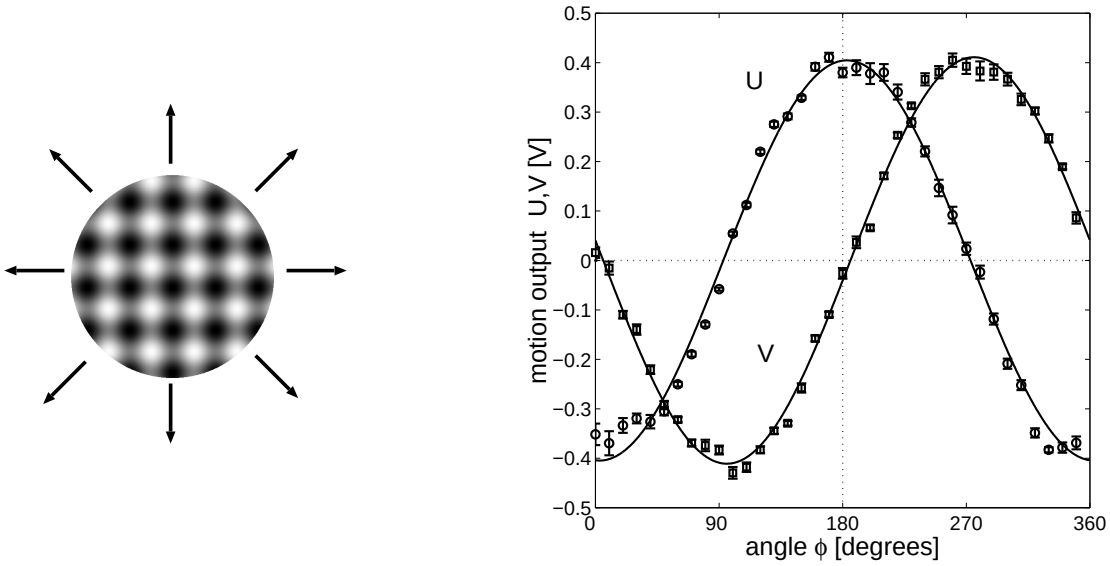


Figure 5.17: *Orientation tuning of the optical flow chip.* The optical flow chip exhibits the expected cosine- and sinewave tuning curves for the orthogonal motion components U and V. The sinusoidal plaid stimulus was moving at a velocity of 30 pixels/sec and the bias voltages were set as Bias_{VVI2}=0.46 V and Bias_{OP}=0.28 V respectively.

the bias constraint.

For strengths of the bias constraint smaller than a particular threshold value ($\text{Bias}_{OP} <$ 0.25 V) the optical flow chip almost perfectly reports the true object motion in either of the four tested directions. A small tilt of the motion vector in the order of less than 10°remains as the result of the remaining intrinsic output conductance of the feedback node (see Section 5.2.3). By contrast, the vector average integration scheme would lead to a deviation in direction of $22.5^{\circ}$for this particular object. As the strength of the bias constraint exceeds the threshold value, the reported speed decreases rapidly and the resulting direction estimation resembles the characteristics of a more vector average estimation scheme. A further increase in the bias voltage Bias_{OP} strengthens the bias constraint further such that the chip is not able to respond at all anymore. These findings match very well our theoretical predictions. Because an increase in bias strength is equivalent to a decrease in stimulus contrast, the recorded trajectories also qualitatively fit the ones from simulation[13].

Note, that the threshold value of the bias voltage is slightly smaller than the minimal value required in any practical applications. A minimal level of Bias_{OP} is needed in order to counterbalance any second order effects and is defined as the minimal voltage for which the chip still robustly reports a zero motion output if the visual contrast vanishes and the

[13]see Figure 4.6

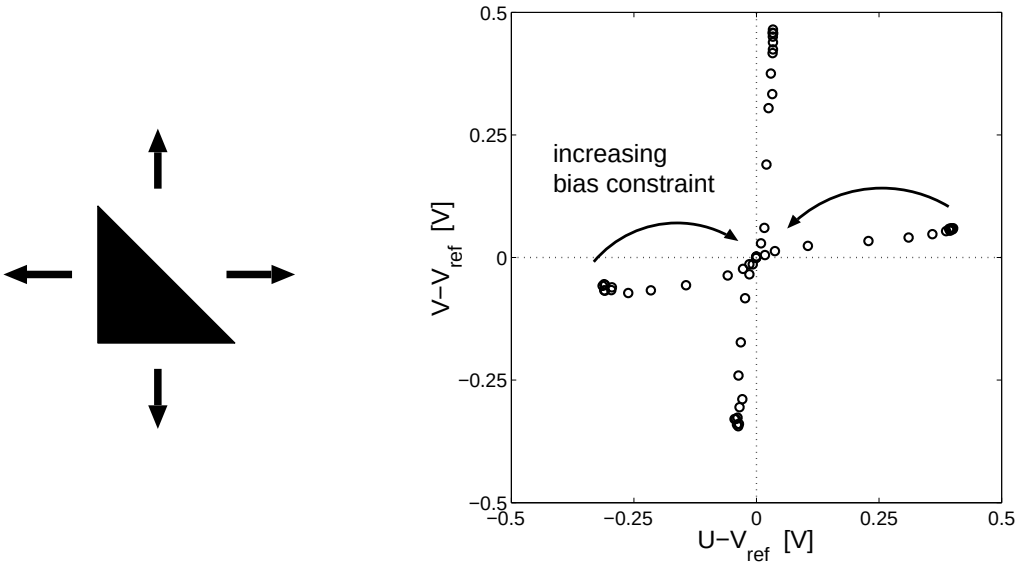


Figure 5.18: *Solving the aperture problem.* The time-averaged output of the optical flow chip is shown for a triangular object moving in orthogonal directions. The collective computation of the intersection-of-constraints solution is required in order to achieve the correct global motion estimate. Data points along each response trajectory correspond to particular bias voltages Bias_{OP} (0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.33, 0.36, 0.39, 0.42, 0.45, 0.48, 0.51, 0.75 V) where increasing voltages decrease the length of the motion vectors. The optical flow chip is capable of solving the aperture problem. Increasing the influence of the bias constraint, however, leads to a global motion estimate that represents more a vector average solution (compare also with Figure 4.6).

photoreceptors are completely adapted.

5.3.2 Flow field computation

The optical flow chip contains scanning circuitry that allows the optical flow field and the local photoreceptor signals to be simultaneously scanned at rates up to 5000 frames/sec. Figure 5.19 shows images of the photoreceptor signals and the instantaneous flow fields to various stimuli such as a dark dot on light background moving to the left (a,b), a black and white grid moving to the right (c,d) and part of a hand on a light background, moving to the lower right (e,f). For all these stimuli, the optical flow chip provides qualitative reasonable local estimates of visual motion.

The lateral coupling of the units was set such that interaction was fairly limited. As a consequence, the resulting flow field approximates normal flow in areas where the aperture problem holds. This can particularly well be observed for the grid pattern where the optical flow vectors tend to be perpendicularly oriented to the edges of the grid.

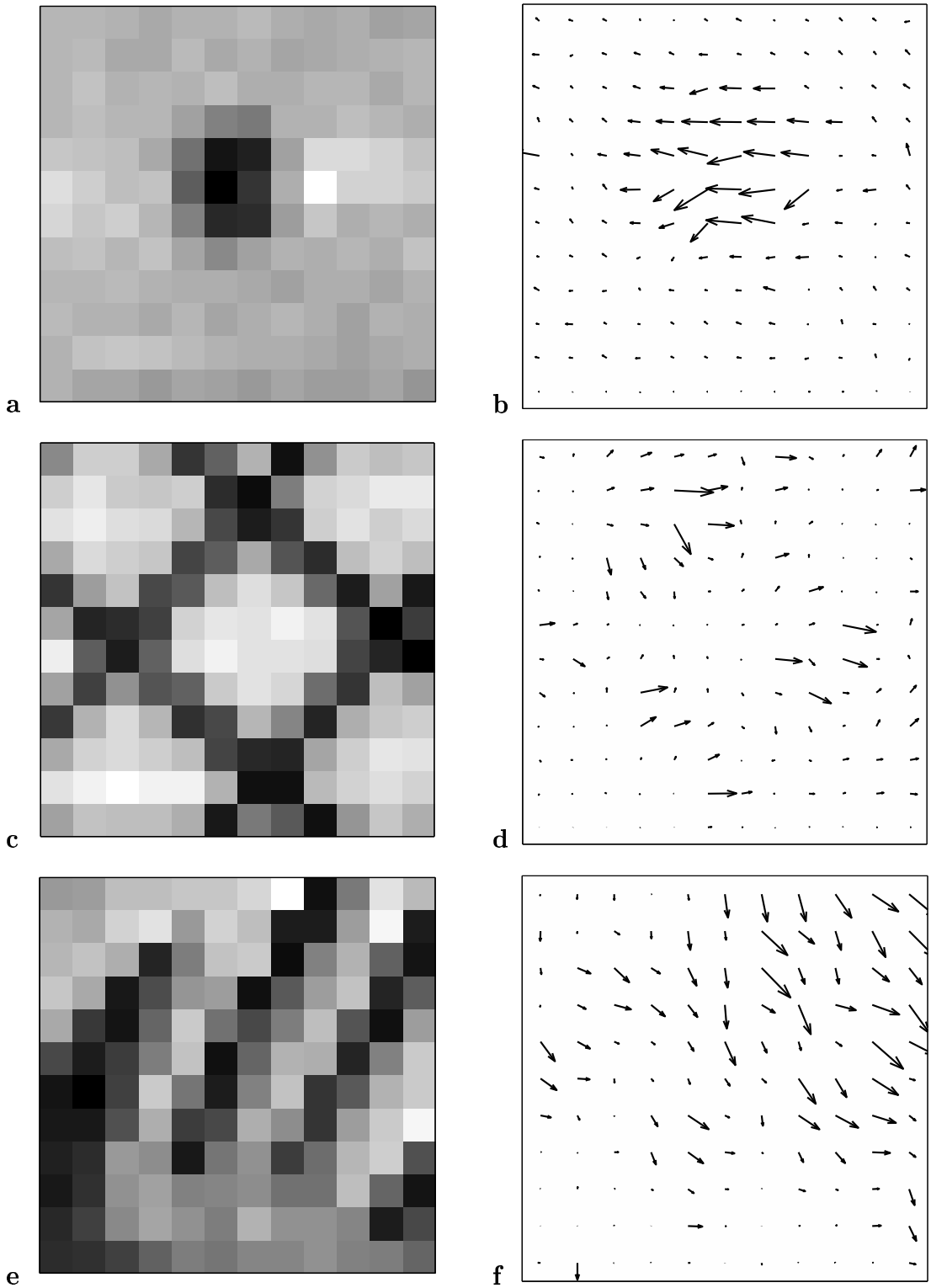


Figure 5.19: *Samples of reported optical flow fields.* (a,b) The scanned photoreceptor signal and the optical flow field are displayed for: a dark point horizontally moving to the left on a light background (a,b); a grid pattern that moves to the right (c,d) and a hand moving in the direction of the lower right corner of the image.

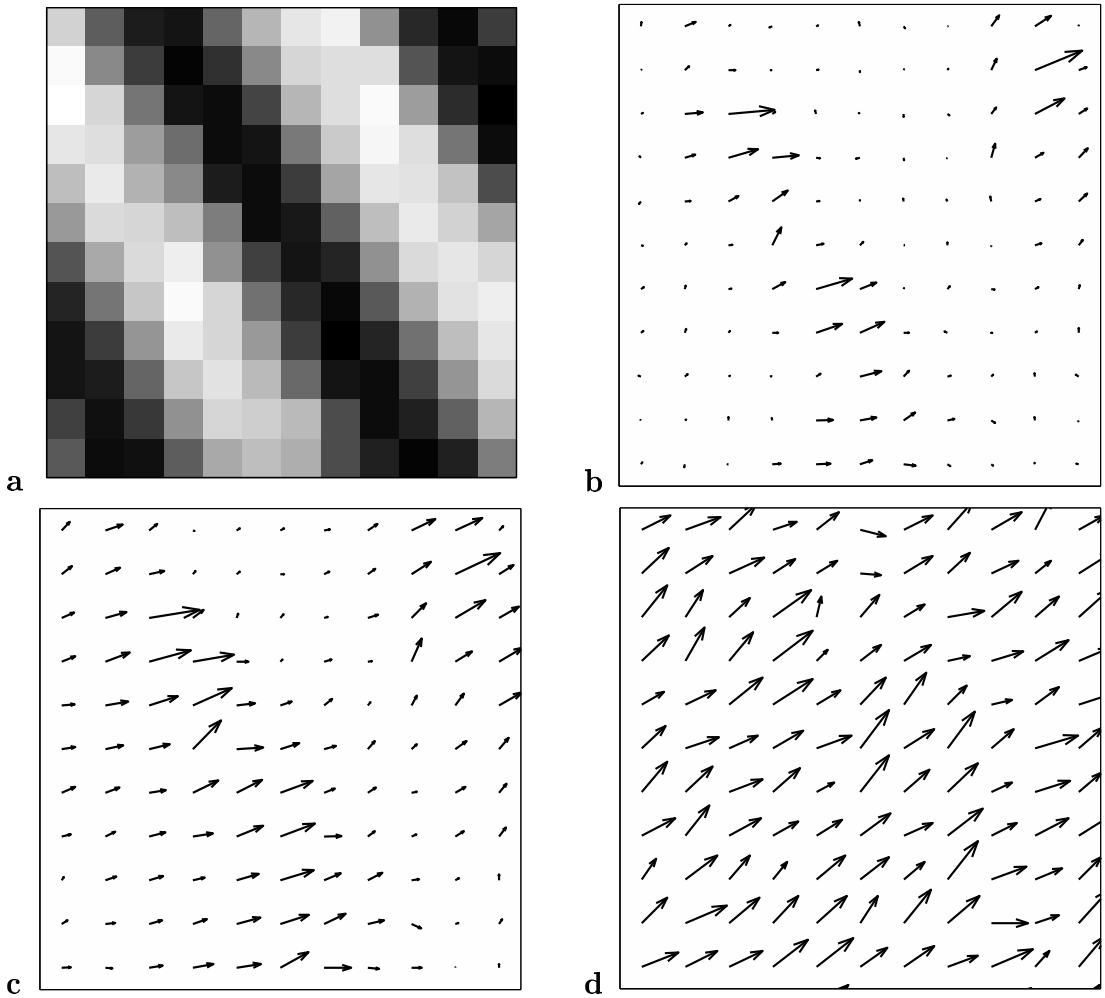


Figure 5.20: *Optical flow fields of increasing smoothness.* The optical flow fields of a sinewave grating moving to the upper right are shown for increasing values of the bias voltage Bias_{HR} (b-d). The image of the associated photoreceptor output is displayed in the upper left (a).

Only at the grid nodes, spatial gradients of different orientations are provided within the local kernel of interaction which makes the collective computation of the correct pattern motion possible. Anisotropic smoothing due to the characteristics of the HRes circuit can be observed in the optical flow field of the moving dot stimulus. The flow field within the dot area is smooth whereas the flow gradient between figure and background is steep. This leads to an improved optical flow estimate that is closer to the correct (global) intersection-of-constraints solution.

Figure 5.20 demonstrates how the applied HRes circuits allow to control the strength of the smoothness constraint and thus the smoothness of the resulting optical flow field: A

sinewave grating stimulus is moving to the upper right direction and the estimated optical flow is shown for three different values of the bias voltage Bias_{HR} (0.25, 0.38 and 0.8 V). As the voltage increases, the optical flow estimate becomes smoother and finally represents a global estimate that does not exhibit any local motion information anymore. The bias constraint controlled by Bias_{OP} allows the chip to compute normal flow. Note however, that most of the estimated flow vectors in (b) and (c) do have a slight bias towards a direction straight to the right-hand side that disappears in the global estimate (d). This shows that robustness and accuracy increase with larger and larger spatial interaction.

5.3.3 Non-idealities and limitations

As discussed in Section 5.1, the instantaneous motion response is expected to be phase-dependent on the stimulus frequency. Measuring a single, isolated motion unit confirms this expectation. We also see deviations from the ideal expected behavior that are the result of potential offsets in the intensity gradient estimation. These offsets are predominantly induced by mismatch in the responsible circuits due to local variations in the fabrication process.

Figure 5.21a shows the measured trajectories of the two optical flow components of a single, isolated motion unit. A sinewave grating moving with constant velocity is presented that is oriented such that the U-component of the flow vector is positive and the V-component is ideally zero. The top trace represents the photoreceptor output of the unit. It is clear that the mean response of the motion output over a complete stimulus cycle is in qualitative agreement with the presented visual motion, having a positive U-component and an approximately zero V-component. The individual trajectories, however, typically vary widely and the U-component even takes on values indicating motion in opposite direction.

Offsets in the estimation of the spatial and temporal gradients can explain the observed behavior to a large extent. Figure 5.21b shows the *simulated* instantaneous response of a single motion unit to a sinewave grating stimulus (top trace) in the presence of offsets in the spatial gradient estimation. Each trace represents the response according to Equation (5.6) with a constant offset Δ added to the computed spatial gradients. If there is no offset assumed, the output reports the correct visual motion most of the time except when the spatial gradient becomes zero and the bias constraint dominates and enforces zero motion. Adding offsets in the order of 25% or 50% of the maximal spatial gradient changes the response behavior drastically. If offset increases the spatial gradient estimate, the motion output decreases. It even changes sign if the effective gradient is smaller than the mismatch induced. When the offset decreases the effective gradient, the motion response is increased. The amount of increase and decrease in response is not equally large, due to the divisive influence of the spatial gradient in the computation of

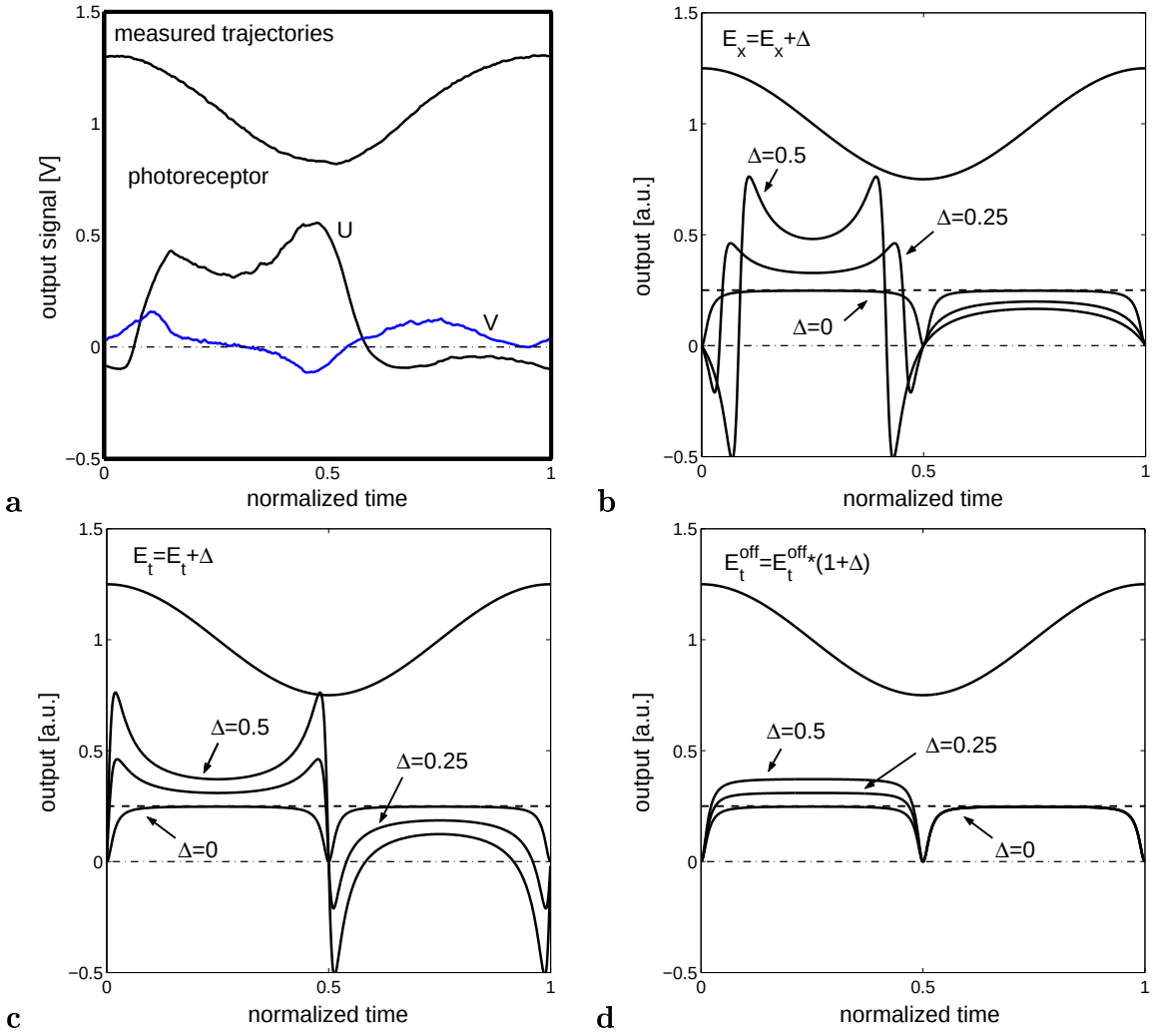


Figure 5.21: *Mismatch in the computation of the spatiotemporal brightness gradients.* (a) The typical instantaneous output of a single isolated motion unit to a sinewave grating exhibits non-ideal behavior due to mismatch. The photoreceptor output (top trace) and the two voltages U and V, representing the components of the local optical flow vector, are shown. Offsets in the computation of the spatial and temporal brightness gradients due to mismatch in the photoreceptor and in the hysteretic differentiator circuit can explain to a large extent the observed behavior, as shown in the following simulation results. (b) Additive offsets in the spatial gradient estimation result in an asymmetric and phase-shifted motion output. (c) Additive offsets in the temporal gradient estimation also induce asymmetric changes in the response but no phase-shift. (d) A gain offset in the off-transient (or on-transient) current of the hysteretic differentiator (which is likely to be induced through mismatches in the current mirrors of the circuit) results in a asymmetric response that significantly affects the time-averaged response. Note in all above cases, that even for no offsets, the response is phase-dependent on the stimulus due to the bias constraint.

the motion output. As a consequence, the output becomes increasingly asymmetric with increasing offsets. The locations of the extrema in the simulated trajectories undergo a phase-shift which can be observed in the measured trajectories as well.

The effective additive mismatch in the spatial gradient computation of the optical flow chip was estimated by measuring the variations of the individual photoreceptor outputs in their adapted state to a uniform visual input. These variations lead to an absolute mean of the mismatch in spatial gradient estimation of 21 mV with standard deviation of 18 mV. The relative impact of the mismatch depends of course on the effective spatial gradient imposed by the stimulus. For example, the sinewave grating stimulus used previously with 80% peak contrast induces a peak-to-peak amplitude of the photoreceptor output of approximately 500 mV. Given a spatial stimulus frequency of 0.08 cycles/pixel, the maximal[14] local spatial gradient signal is approximately 125 mV. Thus the measured mismatch is in the order of 17% of the maximal effective spatial gradient. Obviously, the photoreceptors are not adapted when exposed to a time-varying stimulus. Although the gain mismatch of the photoreceptor output in the transient regime is usually smaller than in the adapted state, the overall offset certainly increases.

Additive offsets in the temporal gradient computation have a similar effect on the resulting output which is illustrated in Figure 5.21c. However, the responses to both half-cycles of the stimulus are more symmetric than for offsets in the spatial gradient. The previously observed phase-shift of the peak responses does not occur. Additive offsets are likely to occur in the hysteretic differentiator circuit due to the previously discussed difference in leakage currents of the source implants between the native and the well-transistor in the rectifying circuit[15].

A third possible source of mismatch consists of the gain differences between the rectified output current of the hysteretic differentiator for positive (on-transient) and negative (off-transient) brightness changes. Figure 5.21d depicts the expected motion output for an increased gain in the off-transient. Although the effect is much smaller compared with the additive effects discussed before, gain offsets amplify the effects of the additive offsets in gradient computation. Gain deviations in the on- and off-currents of the hysteretic differentiator are likely to occur due to the necessity to mirror these currents several times between the hysteretic differentiator circuit and the feedback node.

Flow field offsets

It is interesting now to measure the output variations between the individual units of the optical flow network. We measured the response of each unit to a sinewave grating moving with constant speed diagonally to the array orientation such that the u-component and the negative v-component of the motion should be equally large. The smoothness

[14] with respect to the on-chip spacing in the network array and the given stimulus frequency

[15] see description in Section 5.1.2

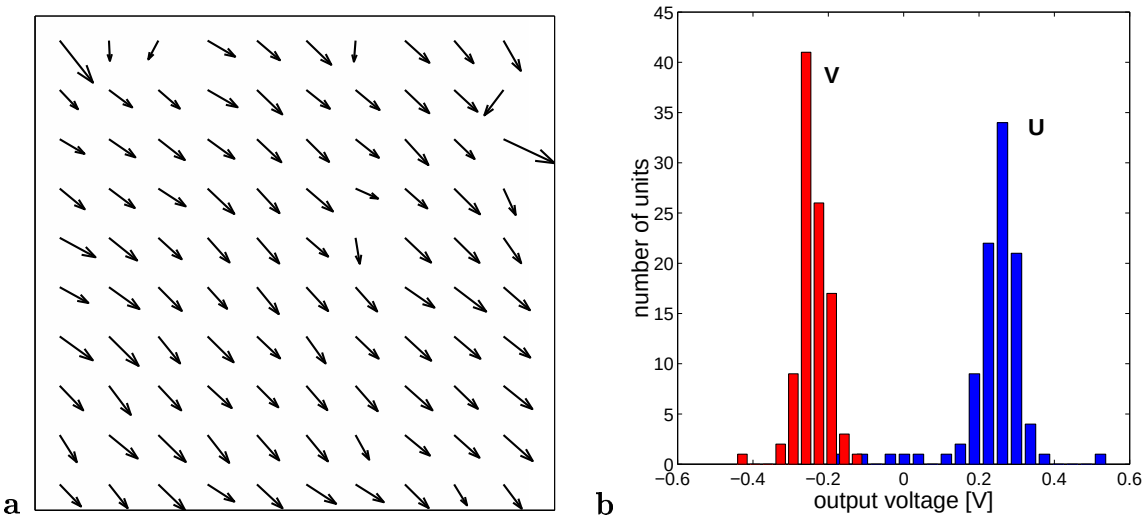


Figure 5.22: *Flow field offsets.* (a) The measured time-averaged motion output of individual and isolated single optical flow units is plotted for a sinewave grating stimulus, diagonally oriented and constantly moving to the lower right corner. (b) The histograms for both components of the optical flow vector quantitatively show the distribution of the individual motion estimates. We find mean values of $\overline{U} = 0.25$ V and $\overline{V} = -0.24$ V with standard deviations of $U_{std} = 0.09$ V and $V_{std} = 0.04$ V respectively. Bias settings were $\text{Bias}_{VVI2} = 0.42$ V and $\text{Bias}_{OP} = 0.38$ V.

constraint was completely disabled ($\text{Bias}_{HR} = 0$ V). The stimulus speed and the bias settings were chosen such that the output values are clearly within the linear range of the optical flow units to prevent any distortion of the output distribution due to the expansive non-linearity of the circuit.

Figure 5.22a presents the measured *time-averaged* optical flow field. It reveals some moderate offsets between the optical flow vectors of the different units in the array. The output of some units is clearly incorrect, but these units seem to be located preferably at the array boundaries. Because motion is computed component-wise, deviations due to mismatch are causing errors in speed as well as orientation of the optical flow estimate. Figure 5.22b shows the histogram of the output voltages U and V of all motion units. Mean values are found to be $\overline{U} = 0.25$ V and $\overline{V} = -0.24$ V with standard deviations of $U_{std} = 0.09$ V and $V_{std} = 0.04$ V respectively. Very similar results were found for other chips of the same fabrication batch. The outputs consistently approximate normal distributions as long as the motion signals are within the linear range of the circuit. This is in agreement with randomly induced mismatches due to the fabrication process and rules out any systematic errors.

The above offset values are for completely isolated motion units. Already weak coupling among the units increases noticeably the homogeneity of the individual responses

and decreases the offsets in the individual optical flow estimates.

5.3.4 Processing speed

An argument often made in favor of analog processing systems is their intrinsic fast processing speed. As discussed earlier[16], the instantaneous output of the optical flow chip can be assumed to be the optimal solution of the defined optimization problem given that the time constant of the implemented network is negligible compared to the dynamics of the visual input. The obvious implication of this qualitative statement is the existence of a **processing speed limit** beyond which the optical flow estimate eventually deviates largely from the optimal solution of the problem. Recalling the only function of the network, namely to reach as quickly as possible the global minimum of the cost function, we immediately recognize that the **time constant** of the network represents a measure of its maximal processing speed: If the dynamics of the visual scene exceed the time constant of the network, the global minimum can change its location in energy space faster than the network is able to reduce the distance to it and thus it will potentially never reach it.

The time constant of the optical flow chip is mainly determined by the capacitances C_u, C_v and the various attached conductances of a single motion unit (schematics in Figure 5.4). C_u and C_v are not explicitly designed but represent the sum of the various parasitic capacitances present. The total conductance at the capacitive nodes mainly consists of the transconductance of the multiplicative feedback circuit, but also of the transconductance of the amplifier that implements the bias constraint and the remaining output conductance of the cascoded current mirror. As a consequence, the time constant and thus the processing speed **varies** and depends on the instantaneous optical flow estimate, the visual input and the bias settings in the circuit. In particular, the transconductance of the feedback loop decreases as the optical flow estimate U and V fulfills more and more the brightness constraint. The time constant therefore increases and slows down the processing. Thus, we are forced to redefine processing speed as *the minimal time the optical flow chip requires to reach an estimation level of the optimal solution of a certain accuracy.* With respect to this definition, we now list and qualitatively discuss some of the above mentioned influences on the processing speed:

- Low spatial contrast
  If the local stimulus contrast is very low, the spatiotemporal brightness gradients vanish and the correction currents at the capacitive nodes become small. The time constant then is large because the conductance is given mainly by the transconductance of the amplifier and the remaining output conductance of the cascoded current mirror, which both are ideally small. The lower the visual contrast, the slower is the processing speed. This is a sensible property of the optical flow network because

[16] see Section 4.1.3

in this way, it can distinguish between noisy input of low contrast but high spatiotemporal frequencies and reliable high contrast but low frequency input moving with the same constant velocity; in the first case, the integration time is too short to elicit a noticeable output signal while in the latter case, the low spatial frequency of the stimulus provides the necessary integration time to build up a correct motion estimate.

- High spatial contrast
  Conversely, high spatial stimulus contrast increases the feedback currents and leads to faster processing in the optical flow chip, which is also sensible because locations of high spatial contrast are more reliable stimulus features that need a shorter integration time and thus allow a higher temporal resolution. Similarly, fast visual motion increases the transconductance of the feedback loop and therefore increases the processing speed.

- High current densities
  The overall current densities in the feedback loops determine their transconductances (5.26). These densities depend on the bias voltage Bias_{VVI2} of the wide linear-range multipliers and the amplification of the output currents of the hysteretic differentiator given by the voltages HD_{tweak+} and HD_{tweak-}. In practice, the requirement for sub-threshold operation limits the largest possible current densities.

In the following, some quantitative measurements of the processing speed of the optical flow chip are presented for one particular visual stimulus and a given typical current density. As illustrated in Figure 5.24a, the stimulus consisted of a plaid pattern composed of two orthogonal sinewave gratings of which the contrast of one grating was varied $[0 \ldots 50\%]$ while the contrast of the second one was kept constant at 50%. The spatial and temporal frequency of each grating was identical and given as 0.08 cycles/pixel and 30 pixels/sec respectively. The plaid pattern was presented such that the orientation of each grating was parallel to one of the motion component axes. Each data point represents the mean and standard deviation for a total of 10 samples. Processing time was defined as the time needed by the circuit to raise the voltages U and V up to 75% of their asymptotic signal level after the onset of the moving stimulus. Before the onset of the stimulus, the visual input to the chip was a uniformly illuminated surface of constant brightness. Figure 5.23 shows the minimal processing time for each component of the global optical flow estimate as a function of the varying contrast of one of the gratings. If both gratings are of high contrast, the processing time is approximately 30ms and equal for both components of the optical flow estimate. As the contrast of one grating decreases, the processing time for the motion component pointing perpendicularly to the grating's orientation strongly increases. At 0% contrast, no visual information is present in that direction and the

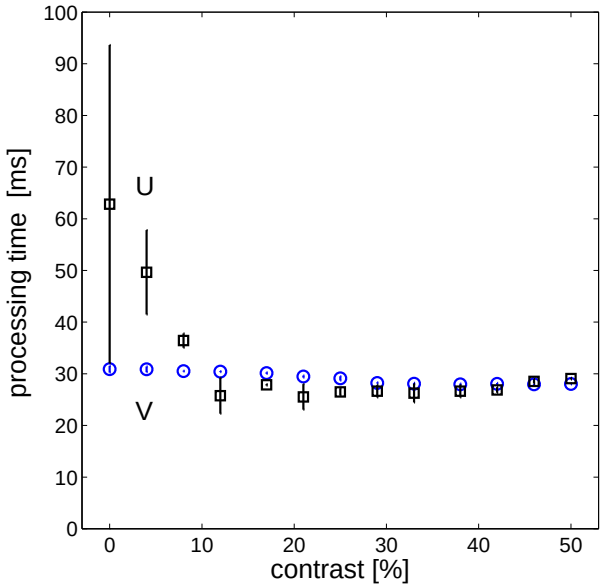


Figure 5.23: *Processing speed of the optical flow chip.* The graph shows the time needed for the output voltages U and V to reach 75% of their asymptotic output level after the onset of a moving plaid stimulus. The plaid consisted of two sinewave gratings where the contrast of the grating oriented perpendicular to the u-component of the motion varies $[0 \ldots 50\%]$ while the contrast of the other grating is constant (50%). For low contrasts, the processing speed significantly decreases. Beyond some contrast level ($> 12\%$) the processing speed stays constant because the Gilbert multipliers saturate and therefore limit the correction currents.

measured processing time is only determined by the bias conductance[17]. The processing speed saturates for contrast values $> 12\%$ because the pFET Gilbert multipliers feeding into the cascoded current mirrors saturate and therefore limit the transconductance of the feedback loop (see schematics in Figure 5.4).

5.4 A Model for Motion Processing in Visual Cortex ?

The proposed optical flow chip clearly does *not* represent the attempt of a close structural embodiment of the primate's visual motion system in silicon; and it was never intended to be so. Considering the large differences in the computational substrates of silicon and the brain, such an attempt also does not seem very reasonable at all.

[17] The error bar is evidently huge because the component's value is ideally constant and zero and the 75% limit only determined by noise.

There is strong belief that the processing of motion information in visual cortex is mainly a two-stage process. The primary visual cortex V1 thereby serves as the first stage where neuronal activity represents the spectral energy of the visual input within narrow spatiotemporal frequency bands [Adelson and Bergen 1985]. Neurons in V1 are therefore sensitive to the orientation and direction of the visual stimulus, as has been known for a long time [Hubel and Wiesel 1962]. Also because of their small receptive field sizes, such neurons are inherently confronted with the aperture problem. A second processing stage is then required that has to compute local velocity **explicitly** and unambiguously. It is widely believed that area MT is the earliest cortical region where sub-populations of neurons show such properties. Although often suggested [Movshon et al. 1985] it has been clearly shown only recently that some MT neurons exhibit in fact an explicit velocity encoding scheme [Perrone and Thiele 2001]. Several models have been proposed to encode velocity where neurons receive the combined input from those spatiotemporally tuned neurons of V1 that are consistent with a particular image velocity and thus populate a plane in the spatiotemporal frequency domain [Heeger 1987a, Grzywacz and Yuille 1990]. Thus, it seems that V1 neurons rather span the complete range of spatial orientations than only subsets that are consistent with specific stimulus-dependent patterns [Schrater et al. 2000]. It is obvious, that visual motion energy induces activity in multiple velocity planes in the frequency space due to noise and visual ambiguities. As a consequence, several velocity-tuned neurons in MT will be active and there has to be a *competitive process* taking place in order to generate an unambiguous vote for a particular local velocity. There is physiological and psychophysical evidence that this competitive process is affected by many different stimulus relevant parameters such as spatial frequency, shape, contrast and disparity [Adelson and Movshon 1982, Movshon et al. 1985, Yo and Wilson 1992, Burke and Wenderoth 1993, Lorenceau et al. 1993, Bradley et al. 1995].

Knowing this, we may now ask to what extent the optical flow network is at least a plausible *functional model* of parts of the cortical motion processing circuitry in area MT or even MST. Clearly, the optimization problem of finding the intersection-of-constraints solution is functionally equivalent to the competitive process necessary to determine unambiguously local motion[18]. Furthermore, the bias constraint of the optical flow network, which was introduced from an engineering point-of-view, can explain several perceptual effects such as the tendency to report normal flow when local ambiguities occur [Nakayama and Silverman 1988a, Nakayama and Silverman 1988b].

Experiments

We tested the global motion response of the optical flow chip to moving sinusoidal plaid patterns, a class of two-dimensional visual stimuli that are generated by the superposition

[18] We assume that the velocity space is densely encoded by velocity-tuned neurons.

of two one-dimensional sinewave gratings of different orientation [Adelson and Movshon 1982]. These stimuli have been used for many physiological and psychophysical experiments on biological visual motion processing systems.

The first test stimulus consisted of a type-I plaid pattern [Ferrera and Wilson 1990], that is the superposition of two orthogonal gratings. As Figure 5.24a shows, the plaid was chosen such that the orientation of each grating is mapped to one of the component axes U and V of the optical flow chip. Besides their orientation, the gratings were identical in spatial and temporal frequency and their normal motion was positive with respect to the voltages $U - V_{ref}$ and $V - V_{ref}$. The circular aperture of the stimulus was large enough to ensure that the chip exclusively received visual input originating only from the plaid pattern.

Figure 5.24b displays the time-averaged output of the optical flow chip, recorded for different contrast ratios of the two gratings where the contrast of one grating was kept constant at 50% while the other one varied [0...50%]. If both gratings have identically high contrast, the response is in good agreement with the intersection-of-constraints solution. As the contrast of one of the gratings decreases, the reported motion continuously changes direction until it signals pure component motion of the other grating. Ideally, the output would not show such an immediate deviation of the pattern motion for decreasing contrasts but rather would stay close to the pattern motion estimate and then switch rapidly only if the contrast ratios become small (compare also with Figure 4.6). However, this ideal behavior could not be reproduced because the strength of the bias constraint had to be slightly increased compared with previous measurements ($\text{Bias}_{OP} = 0.32$ V) to assure normal flow estimation when the contrast of one grating becomes zero.

Also type-II plaid stimuli were tested (Figure 5.25), where the gratings are such that the component motions are within a different quadrant of the velocity space than the combined pattern motion of the plaid. Figure 5.25b shows the output response for different contrast ratios of the gratings. If the contrast of both gratings is identically high, the optical flow chip reports the correct pattern motion that has a negative U-component whereas the component motion is always positive. As the contrast of either of the two gratings decreases, the output is increasingly biased towards the individual component motions.

The behavior of the optical flow chip in the two experiments is at least qualitatively in good agreement with psychophysical data reported. It has been shown early that under high-contrast and unambiguous conditions, human visual motion perception performs functionally an intersection-of-constraints computation [Adelson and Movshon 1982]. However, Stone et al. [1990] found, that humans subjects report increasing deviations of the perceived plaid motion towards the normal motion of the higher contrast grating if contrast of the other grating is successively reduced. Other authors reported a continuous transition between the perception of the vector average of the component mo-

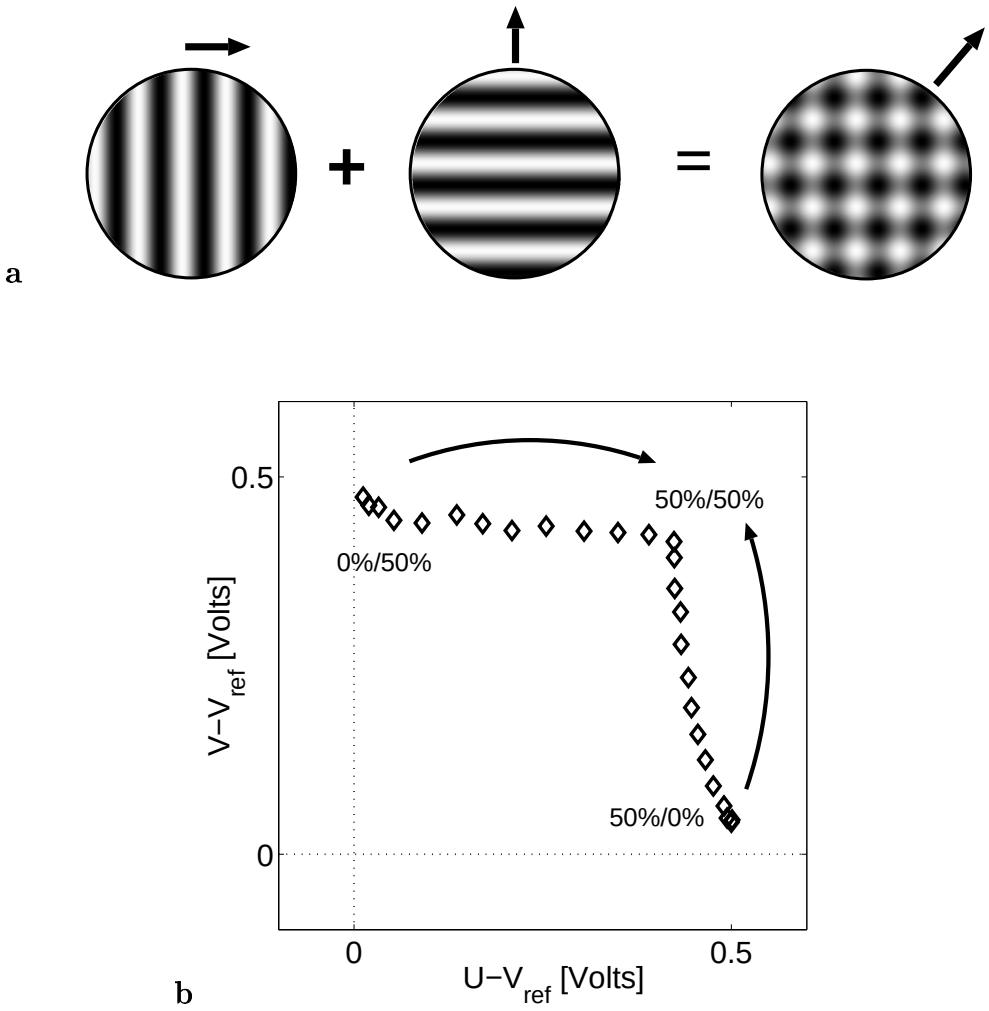
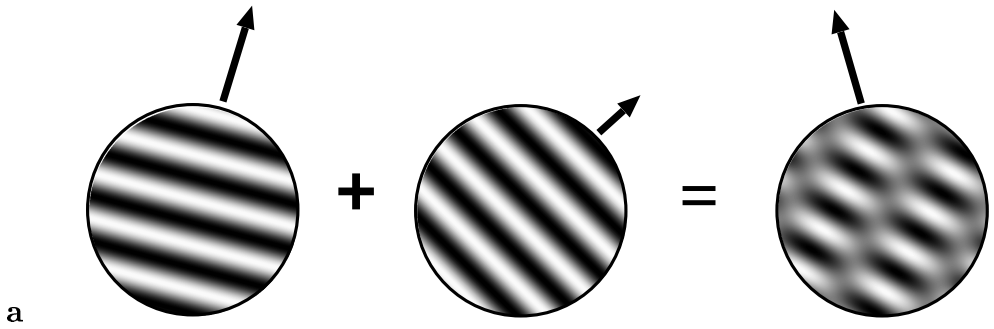
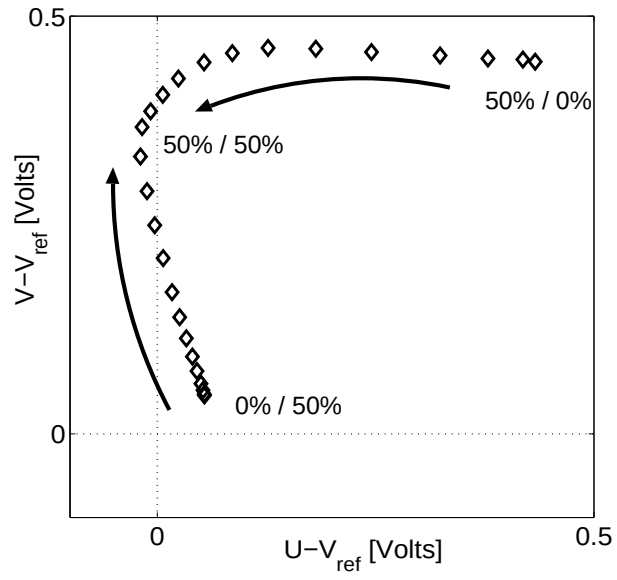


Figure 5.24: *Type-I plaid stimulus of varying contrasts.* (a) A type-I plaid stimulus was generated by superimposing two orthogonal sinewave gratings of same spatial and temporal frequency. (b) The global motion response of the optical flow chip is shown as a function of contrast $[0 \ldots 50\%]$ of one of the underlying gratings while keeping the other grating at a constant contrast value of 50%. The response of the optical flow chip exhibits a continuous transition between component and pattern motion.



a



b

Figure 5.25: *Type-II plaid stimulus of varying contrasts.* (a) A type-II plaid pattern consists of two sinusoidal gratings that each have a normal motion in the same quadrant of the velocity space but superimposed create a pattern motion that lies in a different quadrant. (b) The optical flow chip reports the correct pattern or component motions, depending on the individual contrast of the gratings.

tion vectors and the intersection-of-constraints solution [Nakayama and Silverman 1988a, Burke and Wenderoth 1993, Weiss and Adelson 1998], depending on the contrast differences of the gratings.

Another functional similarity between the optical flow chip and the primate visual motion system can be found in the dynamics of the motion responses. In order to solve local motion ambiguities, integration has to take place over some spatial scale. When presented with a type-II plaid stimulus, humans perceived motion to be biased towards the vector average motion at short durations, approaching the intersection-of-constraints solution only after some time lag [Yo and Wilson 1992]. Furthermore, Yo and Wilson found that the time lag increases strongly with decreasing stimulus contrast. Just recently, Pack and Born [2001] supported these psychophysical results by intra-cellular recordings in area MT of alert macaque monkeys. They showed that in the early response phase after stimulus onset, the direction tuning-curves of MT neurons are strongly biased towards the orientation perpendicularly to the local stimulus orientation. Only after some time (60ms) was the direction-tuning consistent with the neuron's preferred direction of the overall pattern motion. Considering the results of the previous section on the processing speed of the chip and its dependence on contrast, we expect a very similar behavior of the optical flow chip. And in fact[19], tested with the type-II plaid stimulus shown in Figure 5.25b, the resulting motion estimate is initially biased towards the average component motion, thus pointing into the first quadrant in velocity space (positive U and V component) and only later on, after some time lag of ≈ 50 ms converges to the true pattern motion in the second quadrant.

5.5 The Motion Segmentation Chip

We present an extended implementation of the optical flow chip that includes the active control of the local smoothness conductances. The implementation follows the motion segmentation network system proposed in Section 4.2.2 that introduces two additional discontinuity networks with units P_{ij} and Q_{ij}. Active units in these networks represents active line segments oriented in x- and y-direction. The hardware implementation is modified insofar as there is no soft-WTA competition taking place amongst the discontinuity units to favor a continuation of the line process. This simplification was necessary because otherwise, the implementation would have required pixel sizes and wiring densities that exceed the feasible complexity level of focal-plane chip architectures. As a consequence, we will not consider the P and Q units to form networks in a strict sense because there is no connectivity between these units.

We modify the input dynamics of the discontinuity networks (4.32) with respect to

[19] not illustrated

the above mentioned simplifications and find, exemplarily for the P units:

$$\dot{p}_{ij} \;=\; -\frac{1}{C}\left[\,\frac{p_{ij}}{R} + \alpha - f(\Delta \boldsymbol{v}^{y}_{ij}) - h(\Delta \boldsymbol{F}^{y}_{ij})\right] \tag{5.27}$$

The quadratic norm in (4.32) is replaced by the more general functions f and h that are only required to be non-negative, symmetric and monotonic. The vector $\boldsymbol{F} = (Fx, Fy)$ is composed of two components that represent the correction currents generated by the optical flow unit for each motion component as given by (5.10). Using the difference in correction current as the input to the discontinuity units rather than the brightness constraint gradient as proposed in Section 4.2.2 incorporates additional information about possible *brightness boundaries* into the segmentation process.

Each discontinuity unit performs a threshold operation: The input p_{ij} increases as long as the sum of the local measure of the velocity differences and the brightness constraints is bigger than the threshold α plus the leakage term determined by the conductance $1/R$. Analogously, p_{ij} decreases as long as it can if the input does not reach threshold. Assuming $1/R$ to be small, the activation function $g : p_{ij} \rightarrow P_{ij}$ has an upper and lower bound and the output P_{ij} of the discontinuity unit is in either one of the two corresponding states. The recurrent loop is closed when allowing the discontinuity units P_{ij} to control the smoothness conductances between neighboring units in y-direction. In an equivalent manner, the dynamics of the units Q_{ij} are defined where their output control the presence or absence of line segments in the x-direction.

5.5.1 Single motion segmentation unit

Figure 5.26 shows the schematics of the complete motion segmentation unit. The core circuit consists of the optical flow unit (represented as black box) that is identical with the schematics in Figure 5.4, except that the HRes circuits are shown here explicitly to clarify the control of the smoothness conductances. The segmentation unit consists of twice identical circuitry that computes the control signal for the line process in the x- and y-direction. Thus it has to receive the appropriate optical flow and correction current signals from two neighboring cells and has to provide them to the two other neighbors. The correction current signals $Fx_{\pm}$, $Fy_{\pm}$ are represented by the gate voltages of the current mirrors at the feedback nodes that logarithmically encode the effective differential correction currents (see Figure 5.4).

The function $f(\Delta \boldsymbol{v})$ is implemented as the sum of the 'antibump' outputs of two bump circuits [Delbruck 1993a] which is an approximation of the quadratic vector norm of $\boldsymbol{v}$. Since the optical flow signals are referenced to a common potential V_{ref}, the components U_+ and V_+ of the unit and its neighbors can be directly the input to the bump circuits.

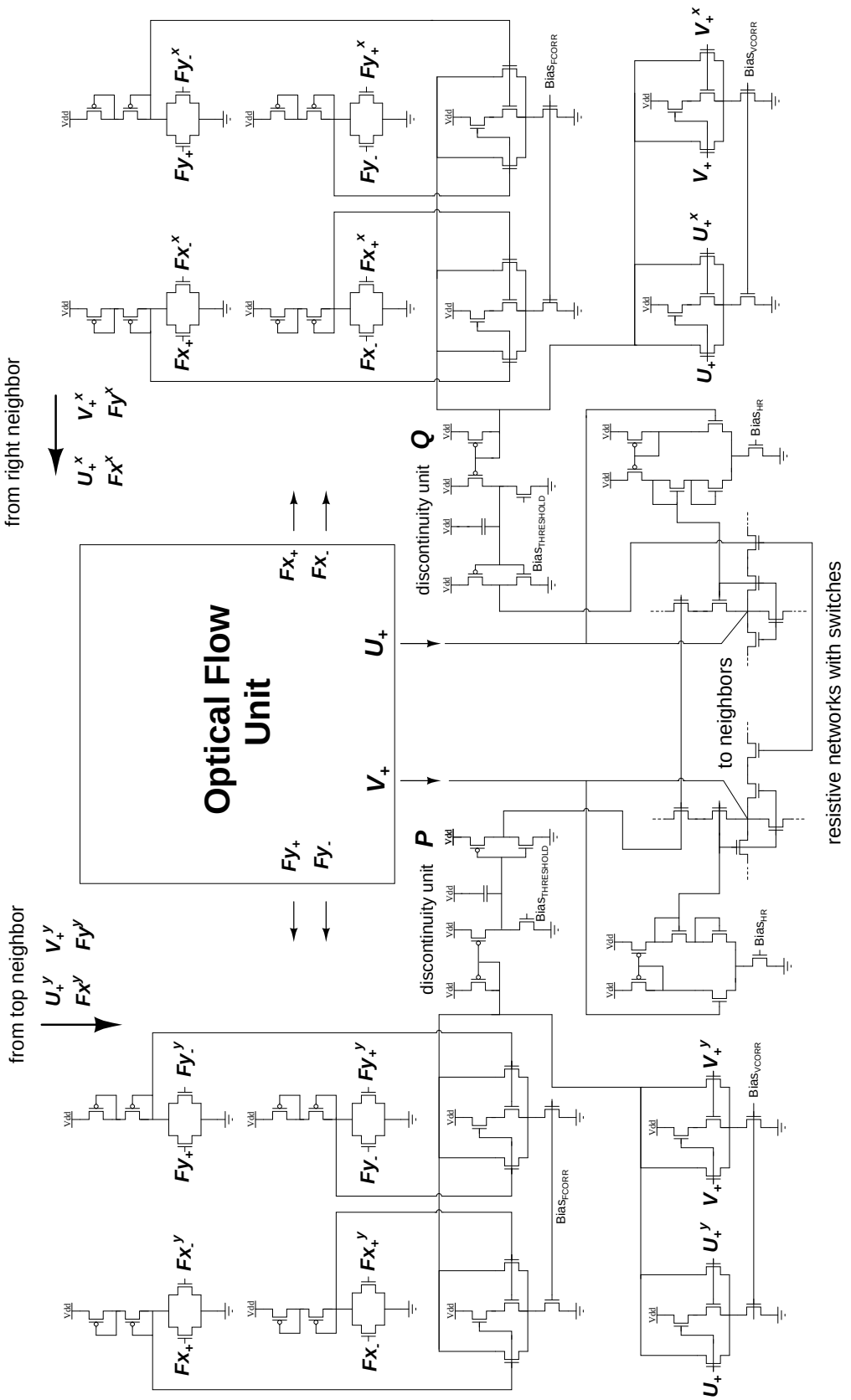


Figure 5.26: *Schematics of a single motion segmentation unit.*

The differential encoding of the correction current signals, however, requires a preprocessing step in order to extract the correction current gradient, which will be described in the next section. The function $h(\Delta \boldsymbol{F})$, however, is also based on the bump circuit dissimilarity measure. The total segmentation current for each of the two line orientations is then the sum of the output currents of the two bump circuits where the bias voltages Bias_{FCORR} and Bias_{VCORR} determine their relative weights. The segmentation current finally serves as input to the discontinuity units that qualitatively exhibit the dynamics (5.27): The input current is compared to a threshold current set by $\text{Bias}_{THRESHOLD}$, and the resulting current then either charges up or down the capacitance C_d. The leak conductance $1/R$ is given by the output conductance of the high-impedance node and therefore is dependent on the threshold and the segmentation current[20]. In any case, it is rather small. The activation function g is given by the inverter stage and is limited naturally by the voltage rails. Finally, the feedback loop is closed by the output of the inverter which either enables or disables the smoothness conductance in between two neighboring optical flow units as set by the horizontal resistor via a pass-transistor.

Computing the gradient of the brightness constraint error signal

The antibump output of the bump circuit biased in the sub-threshold regime provides a compact and appropriate dissimilarity measure between voltage signals. Its bias voltage further allows to adjust the weight of this measure which is crucial for the motion segmentation unit in order to scale the influence of the optical flow gradient versus the correction current gradients appropriately. Using bump circuits for the computation of the correction current gradients requires a stage that ideally transforms the difference between two differential currents linearly into a differential voltage.

Figure 5.27 illustrates the simple transformation circuit used in the motion segmentation unit. Its task is to transform a differential currents difference $\Delta I = I_A - I_B = (I_{A+} - I_{A-}) - (I_{B+} - I_{B-})$ into the voltage difference $\Delta V = V_1 - V_2$. The differential currents are cross-wise summed into currents I_1 and I_2. Stacked pairs of diode-connected transistors generate the voltages V_1 and V_2. The resulting voltage difference[21] then follows as

$$\Delta V = \frac{1}{\kappa_p} \left[\gamma \, \log \left(\frac{I_{A+} + I_{B-}}{I_{A-} + I_{B+}} \right) \right] \, , \tag{5.28}$$

where $\gamma = (\kappa_p + 1)/\kappa_p$ with κ_p being the gate coefficient of the diode-connected pFETs[22]. Assuming that the two differential currents I_A and I_B have a common *DC-current level* I_0, we can expand (5.28) around I_0. We rewrite the logarithm of the quotient as the difference

[20] see Equation (5.26)

[21] in units of kT/q.

[22] In first approximation, κ_p is assumed to be constant and equal for both pFETs.

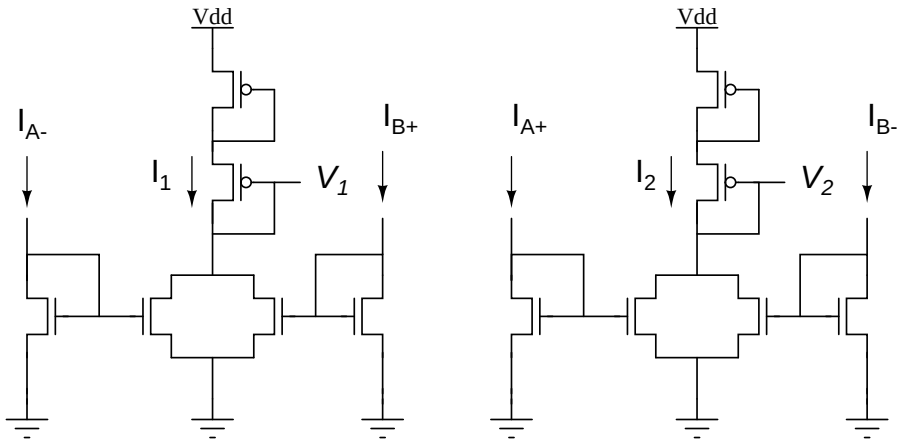


Figure 5.27: *Transforming a differential currents difference into a differential voltage.*

of two logarithms and expand with respect to each differential current component

$$\begin{aligned}
\log(I_{A+} + I_{B-})\Big|_{I_0} &\approx \log(2I_0) + \frac{1}{2I_0}[(I_{A+} - I_0) + (I_{B-} - I_0)] - \frac{1}{4I_0^2}[(I_{A+} - I_0)^2 + \\
&\quad (I_{A+} - I_0)(I_{B-} - I_0) + (I_{B-} - I_0)^2] + R_3 \\
\log(I_{A-} + I_{B+})\Big|_{I_0} &\approx \log(2I_0) + \frac{1}{2I_0}[(I_{A-} - I_0) + (I_{B+} - I_0)] - \frac{1}{4I_0^2}[(I_{A-} - I_0)^2 + \\
&\quad (I_{A-} - I_0)(I_{B+} - I_0) + (I_{B+} - I_0)^2] + R_3 \qquad (5.29)
\end{aligned}$$

where R_3 stands for the remainder of third and higher order terms. Taking the difference of the two logarithms in (5.29), the second order terms cancel because of the assumed common DC-level I_0. The remainder R_3 can be neglected for currents I_A and I_B moderately small compared to I_0 since the components of order n decrease inversely proportionally to I_0 to the power of n. Finally, using (5.29) we can rewrite (5.28) as

$$\Delta V \approx \frac{\gamma}{\kappa_p} \frac{(I_{A+} - I_{A-}) - (I_{B+} - I_{B-})}{2I_0} \propto \Delta I \ . \qquad (5.30)$$

The gain factor $\gamma > 2$ is induced by the additional diode-connected transistor and enlarges the transformation range. Simulations show that for typical sub-threshold levels of I_0, the useful output voltage range is up to 150 mV which matches the bump width of the bump circuit.

We assumed that the differential currents I_A and I_B have a common DC-level, in order for the linear transformation to hold. Recalling that these currents stand for the correction currents of two neighboring optical flow units, we recognize that this assumption is not completely true. The DC-level of each correction current is given as:

- half the bias current of the two wide linear-range multipliers, which is determined by the bias voltage Bias_{VII2} and thus remains constant throughout the array,

- plus a component provided by the hysteretic differentiator that is *not* constant because the sum of the differential output currents of the differentiator is not determined by some bias current[23].

Nevertheless, it is fair to assume that the difference in DC-level is small between neighboring units because the currents of the hysteretic differentiator are typically small compared to the bias currents of the wide linear-range multipliers. This is because the output currents of the multipliers and the hysteretic differentiator need to be of the same order of magnitude to allow a proper function of the feedback loop. And since the multipliers are preferably operated in their linear range, their differential output currents are rather small compared to their bias currents and thus also are the output current of the hysteretic differentiator. So, we can therefore assume that the shift in DC-level of the correction currents is small, and that the linear approximation (5.30) is sound.

5.6 Performance of the Motion Segmentation Chip

In the following, we present results from a fully functional implementation of a motion segmentation network. The data presented was recorded from chips of identical layouts, fabricated on a single production run in $0.8\,\mu$m BiCMOS technology[24]. The prototype chip, in the following called *the motion segmentation chip*, consists of a quadratic array of 12x12 functional optical flow units with a size of $(170\,\mu\text{m})^2$ each, thus having twice 11x11 functional discontinuity units. The motion segmentation chip provides five different signals from each unit: the photoreceptor signal, the two motion signals U_+ and V_+ and the output of the two discontinuity units P and Q.

5.6.1 Detecting motion discontinuities

We tested the ability of the motion segmentation chip to detect motion discontinuities. A relatively simple high-contrast visual stimulus was presented to the chip that consisted of a moving dark dot on a light, uniform background.

Figure 5.28 displays a sequence of the scanned output signals while the dot moves through the visual field of the motion segmentation chip. The rows (a,c) show the photoreceptor signals, overlaid with the local estimate of the optical flow. Below each frame, the associated activity pattern of the discontinuity units is shown (b,d) as a gray scale

[23] see schematics in Figure 5.4

[24] Exact specifications can be found in Appendix E.

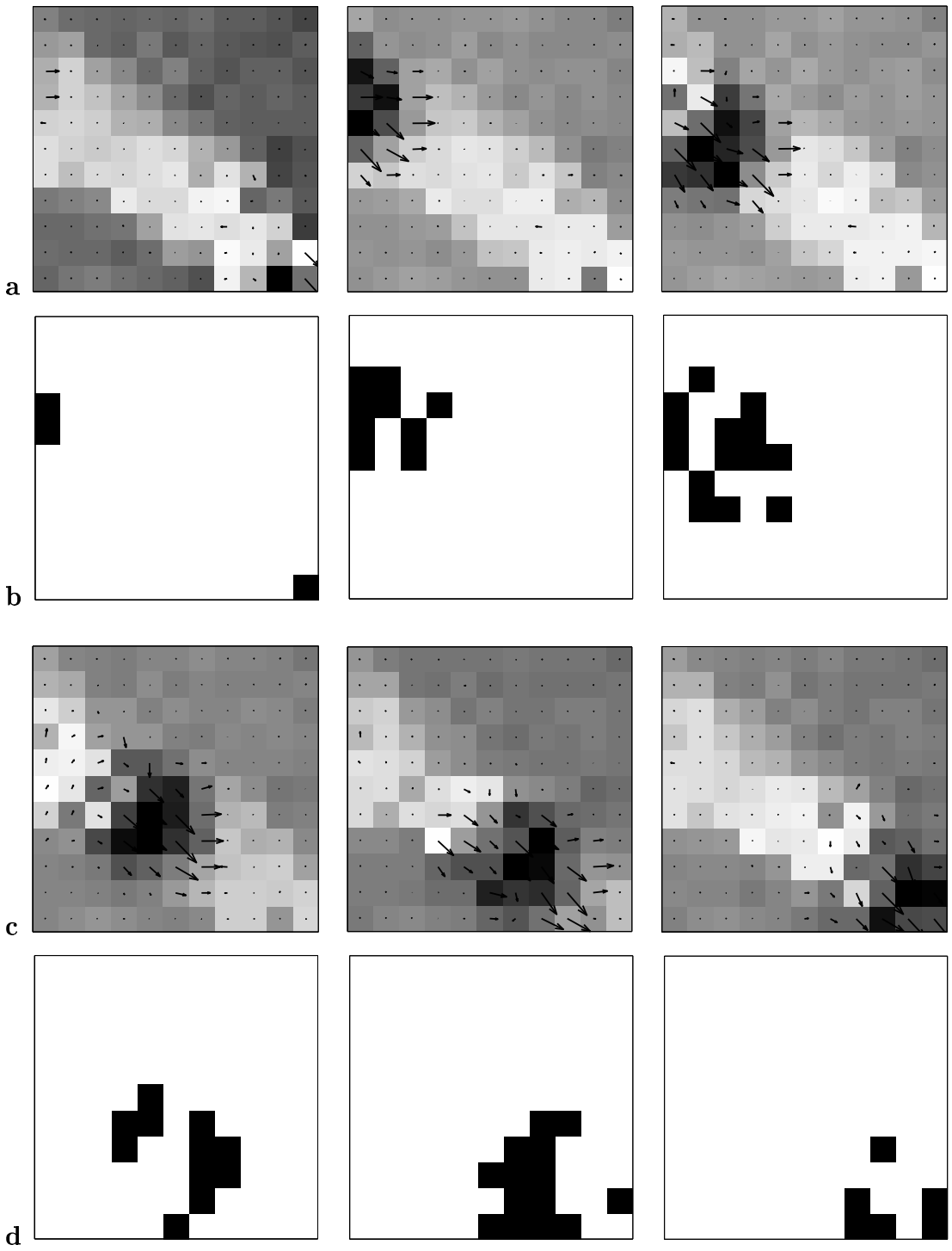


Figure 5.28: *Extraction of motion discontinuities.* A scanned output sequence of the motion segmentation chip is shown, which was presented with a visual stimulus that consisted of a moving dark dot on a light background. (a,c) The photoreceptor outputs are displayed together with the instantaneous estimate of the optical flow field. (b,d) The present state of the discontinuity units is reflected by a gray scale image where dark pixels represent locally active units (P and/or Q).

image where dark pixels represent locally active P and/or Q units. We see that the activity pattern approximately reflects the outline of the moving dot and thus the location of motion discontinuities. Nevertheless, it seems difficult to achieve a closed contour around the single object such that the activity of the discontinuity units leads to two fully disconnected areas in the optical flow network. The smoothing conductance was set rather low ($\text{Bias}_{HR} = 250$ mV) in order to prevent smoothing of the optical flow estimate due to the disrupted contour. As a consequence, the total segmentation current was mainly determined by the strength of the flow gradient which is a reasonable measure considering the relatively small object size.

Note that because the stimulus is repeatedly appearing on the same image trajectory, the photoreceptors in the background adapt over time to some uniform gray value that represents the logarithmically encoded background light intensity. Within the trajectory of the dot, however, the photoreceptors are forced to stay in their transient regime where the increased gain enhances the contrast between object and background and leads to the white stripe in the photoreceptor image.

5.6.2 Piece-wise smooth optical flow estimation

To demonstrate in principle the ability of the segmentation chip to separate completely two image areas around two different flow sources, a stimulus with a less complex motion boundary was applied.

As shown in Figure 5.29a, the applied stimulus consisted of two tightly joined, identical sinewave plaid patterns. One of the plaid patterns remained stationary while the other one moved horizontally with constant velocity, thus forming a linear motion discontinuity. The boundary between the two patterns was chosen such that in moments when they were in phase, the impression of an single uniform plaid pattern occurred. The stationary pattern roughly covered 2/3 of the visual field of the segmentation chip which is indicated by the black square. Although the motion discontinuity had a simple shape, the applied stimulus was very challenging because it did not provide any other cues for segmentation than its underlying motion. Unlike in the previous example of the moving dot, the absolute temporal gradient for example would not be a sufficient measure to find the true motion discontinuity because the average temporal gradient within the moving plaid pattern is constant and therefore the probability to detect a motion discontinuity is constant throughout the image region that belongs to the moving plaid pattern. The lateral smoothing was set high ($\text{Bias}_{HR} = 900$ mV) in order to obtain a piece-wise smooth flow field that does not reflect the spatial pattern of the plaid stimuli.

Two results are presented. In case the discontinuity units are disabled by setting $\text{Bias}_{THRESHOLD}$ sufficiently high, the resulting flow field is very smooth, spreading out far into the image region of the stationary pattern as shown in Figure 5.29b. Clearly, the

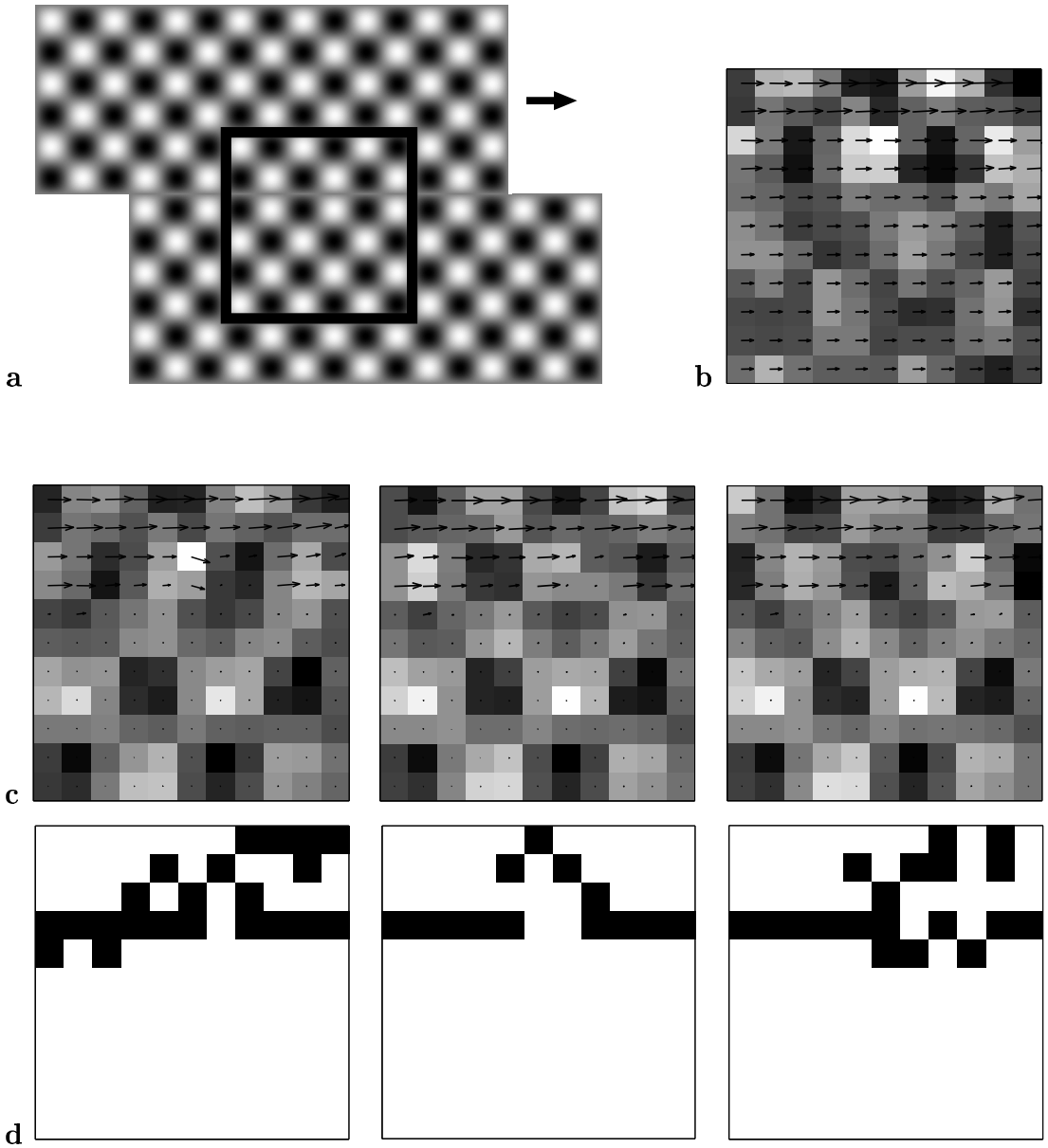


Figure 5.29: *Piece-wise smooth optical flow estimation.* (a) The motion segmentation chip was presented with two completely identical sinewave plaid patterns where one was static and the other one underwent translational motion to the right. (b) The photoreceptor output and the optical flow signal of the motion segmentation chip is shown when the discontinuity units are disabled: the optical flow field is smooth according to the large value of Bias_{HR}. (c,d) A short sequence under the same conditions, now with enabled discontinuity units. The units are preferably active at the motion discontinuity and lead to a clear separation between the two plaid patterns. Thus the optical flow field is smooth only within each separated region.

motion discontinuity is completely smoothed out and not visible from the estimated flow field. Any method that tries to find the motion discontinuity based on this smooth flow field only would fail because the gradient is very shallow and approximately constant over the entire image.

Using in addition the difference measure of the correction forces, however, allows the segmentation chip to find the true motion discontinuity despite the high smoothing conductance[25]. It is able to separate the two flow sources completely and to provide a nearly global flow estimate within each area. Figure 5.29c,d represents a scanned sequence of the chip's outputs, showing the photoreceptor output with overlaid optical flow estimate and the associated discontinuity pattern. The chip can clearly find the linear motion discontinuity although some discontinuity units remain active within the moving plaid pattern. The estimated optical flow field respects the motion discontinuity and is almost uniform within both separated areas.

5.7 Power Consumption

Neuromorphic aVLSI implementations are often claimed to be low power and energy efficient, especially if compared to digital, sequential processing hardware devices. This claim is not necessarily true. It depends very much on the design of the circuits. In any case, a comparison is almost impossible because exact numbers of commercially available ASICs (Application Specific Integrated Circuits) are mostly not published. Furthermore, such ASICs are highly optimized using latest technology whereas most known neuromorphic implementations are prototypes, fabricated in older, low-cost technologies which makes a comparison difficult. And finally, the exact power consumption of neuromorphic motion sensors is only rarely indicated in the literature (as *e.g.* in [Etienne-Cummings et al. 1993, Arreguit et al. 1996]).

The power consumption of a single pixel of the optical flow chip and the motion segmentation chip was measured as 72 μW and 80 μW respectively, not including the power dissipated in all of the necessary read out circuitry. This values are taken for standard bias settings under standard illumination conditions and with no visual input present. A large fraction of the power is dissipated in the wide linear-range multipliers and also the unity-gain follower because these circuits are constantly running in the above-threshold regime. These values are comparable with the power consumption of the 2D motion detector by Etienne-Cummings et al. [1993] which was measured to be 100 μW per pixel. How much an optimized implementation using a special low-power process can reduce power consumption was shown by Arreguit et al. [1996] where they claim a dissipated power of only 1.7 μW per pixel.

[25] compare also with Figure 4.13

5.8 Conclusions

Two focal-plane aVLSI implementations of constraint optimization analog networks for visual motion processing have been presented. They are both fully functional under real-world conditions. They prove the feasibility of embedding such networks in a physical substrate. The *optical flow chip* has been characterized carefully in detail. Its relatively large linear range of ± 0.5 V allows to reliably detect visual motion over three orders of magnitude in speed. The leak conductance ensures that the behavior of the chip is robust even for ambiguous visual input. The globally adjustable lateral conductances in the network offer the possibility to continuously control the degree of integration of visual motion information. If motion is estimated globally, the optical flow chip provides the intersection-of-constraints solution and thus solves the aperture problem for a single motion source.

Considering the 10x10 array size of the presented prototype chip, robustness can be greatly improved with increasing resolution. A higher resolution would allow to increase the number of pixels in the local integration process while having effectively the same smoothness of the flow field. A 32x32 array of the same layout in the same technology[26] would only require a die size of approximately $(4.5 \text{ mm})^2$, including the necessary read out and pad circuitry.

The *motion segmentation chip* represents to our knowledge the first 2D implementation attempt of its kind. It detects motion discontinuities and provides motion segmentation with piece-wise constant flow fields for real-world visual stimuli. Although based on the optical flow chip, its computational complexity is much higher because it includes the recurrent feedback loop between the discontinuity units and the optical flow network that is needed to find solutions to the computationally hard problem of segmentation. One of the key features of the motion segmentation chip is the combined use of the local velocity gradient and the gradient in the constraint currents in order to determine the locations of motion discontinuities. Only this combination allows a large smoothing conductance to be applied while still being able to perform motion segmentation. However, the performance of the segmentation chip depends rather strongly on the segmentation threshold and the weighting of the segmentation signal. This dependence can at least be partially overcome if discontinuity networks are applied as proposed in Section 4.2.2 where soft-WTA mechanisms lead to a more adaptive thresholding behavior. The implementation of such a network was waived because the pixel size and the wiring density would have become unfeasible large. Even in its current state, the pixel size of the motion segmentation chip, having approximately 200 active elements, is close to the limits of reasonable focal-plane implementations.

[26] AMS $0.8\,\mu$m BiCMOS

Chapter 6

Conclusions

This thesis outlined a particular understanding of the perceptual process of visual motion. Due to the inherently ambiguous nature of the visual information, the perception of apparent motion requires the **interpretation** of this information according to some *a priori* assumptions represented by the internal perceptual model of the observer. In a dynamical visual scene, many different motion sources can be present simultaneously. The perception of motion of each source requires the individual selection of an appropriate motion model and the region-of-support where it accounts for. Such selection necessarily reflects a competitive process that can be understood as an optimization problem of choosing the motion model that optimally explains the visual information with respect to the expectations of the observer. In addition, a second optimization process is necessary on the level of the individual motion models in order to extract the optimal model parameters for the given visual input. The framework of **constraint optimization** has been introduced that serves as the tool to formulate the models mathematically and to derive the architecture and the dynamics of equivalent analog networks that solve the competitive process of visual motion perception.

Within this context, I have presented an analog network solution for the estimation of **smooth optical flow**. It imposes a single translational motion model with a given and isotropic region-of-support. Optimization is only needed to find the optimal estimation of optical flow according to the model and its given parameters. I have shown rigorously that the particular optimization problem is well-posed for the complete visual input and parameter space, unlike previously proposed regularization approaches [Horn and Schunck 1981, Poggio et al. 1985, Grzywacz and Yuille 1991]. The introduced **bias constraint** plays an important role in this performance and also leads to solutions of the network that coincide with known psychophysical properties of the primate's visual motion perception system.

Furthermore, I have suggested two network architectures that allow dynamical changes

of the local model parameters[1] of the optical flow network in order to solve for computationally hard problems such as **motion segmentation** and **motion selective optical flow estimation**. The recurrent architecture of these networks reflects the increased computational complexity. The motion selective network is an illustrative example of how the interplay between top-down and bottom-up input leads to a selective perception of the visual input.

An important aspect of this thesis is to provide the complete analysis-synthesis loop, that is to emulate (and not only simulate) the suggested optical flow and motion segmentation networks in a physical substrate. The presented aVLSI implementations serve as the existence proof for the plausibility of the proposed networks architectures, because they allow the measurement of real-time behavior under real-world conditions. The experiments and their results confirm the robustness and functionality of the networks. The **optical flow chip** and the **motion segmentation chip** are fairly outstanding examples of neuromorphic visual motion processing systems with respect to their computational power and complexity. Both implementations perform continuous-time, distributed parallel processing in arrays of identical units.

The organization of this thesis may have evoked the impression of a sequential 'feedforwad' process that had taken place from the analysis of the problem of visual motion to the analytical derivation of the network architectures and up to the physical implementation of these networks. This was and is definitely not true. Literally spoken, it was a recurrent process between analytical description and physical implementation, leading to a continuous improvement in understanding the problem and finding new solutions. An illustrative example for this fruitful interaction was the introduction of the **bias constraint** for the optical flow network. It was not obvious until the attempt to implement the classical optical flow algorithm of Horn and Schunck [1981], that the unavoidable output conductance of the feedback-loop prevents the exact translation of the algorithm into aVLSI [Stocker and Douglas 1999]. In fact, the output conductance seemed to be the 'natural' mechanism of the circuit to keep itself well-posed[2]. After discovering that such a property is indeed "not a bug but a feature", the influence of the output conductance on the motion estimation was investigated, and finally the bias constraint was included in the optical flow model.

6.1 Outlook

An important contribution of this thesis was the demonstration that constraint optimization networks are indeed possible to be implemented in aVLSI successfully. This demonstration hopefully encourages researchers in the neuromorphic community to reconsider

[1] the conductances ρ and σ

[2] see Hadamard's quote on page 37

such an approach after it had been regarded for a long time - after some implementation failures [Moore and Koch 1991] - as a computationally promising but technically impractical way. Some immediate future directions are worth to consider:

- **Application to other perceptual tasks** Constraint optimization networks could certainly be derived to solve other perceptual tasks such as *e.g.* disparity computation with stereoscopic visual input or the estimation of depth-from-motion.

- **Improved aVLSI implementations - multi-chip systems** The motion segmentation chip already reaches the limits of feasible focal-plane implementation. Larger, more complex, systems will require careful considerations of how to split up such networks into several subsystems on different chips. It will be important to access what the necessary information is that must be passed between the subsystems, and how it should be encoded such that known communication schemes such as the Address-Event-Representation [Mahowald 1994] can guarantee reliable transmission. Superficially considered, these questions address only typical engineering problems. However, these are fundamental questions to solve for the understanding of computation in larger recurrent networks and thus are of vital importance in the field of computational neuroscience.

- **Analytical tools** The analytical methods proposed in Chapter 3 and 4 are sufficient to describe the stability behavior of simple uniform networks. However, for non-homogeneous systems containing different recurrently connected sub-networks, improved methods are needed. Such methods must be able to characterize system behavior based on the properties of its sub-networks rather than on the dynamics of the individual computational units.

- **Product applications** A successful translation of the visual motion chips presented here into a product would tremendously strengthen the claim that neuromorphic engineering can provide very efficient and superior solutions to particular class of problems. The optical flow chip operated as a global motion estimator might find an immediate application for simple, visually controlled human-machine interfaces. Its small size, low power consumption and the low dimensional signal output makes it a cheap and simple to use device, especially for mobile applications.

As my final remark, I propose that competitive mechanisms similar to those describe here for visual motion perception, could be a common processing strategy in higher-level cognitive areas. In order to describe such processes in a similar framework as demonstrated here, however, we have to find means to adapt and transform the network architectures from a topographic input-space representation towards a more abstract and sparse representation in object-space.

Appendix A

Variational Calculus

Let us consider the following minimization problem:
Given the *Lagrange function* $L : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ which is sufficiently regular (e.g. $\in C^2$) and the function $f : \partial\Omega \to \mathbb{R}^m$, find $\boldsymbol{q} : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ such that

$$\begin{aligned} H &= \int_\Omega L(\boldsymbol{q}(\boldsymbol{x}), \boldsymbol{q}'(\boldsymbol{x}), \boldsymbol{x}) \ \mathrm{d}\boldsymbol{x} = \min ! \qquad (A.1) \\ \text{and} \quad & \boldsymbol{q}(\boldsymbol{x}) = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \partial\Omega \qquad \text{(boundary condition)} \end{aligned}$$

Definition 1. A function $\boldsymbol{q}(\boldsymbol{x})$ is called a candidate for a solution of (A.1) if it is defined on Ω, obeys the boundary condition and is piece-wise differentiable.

A necessary condition for a candidate $\boldsymbol{q}_0(\boldsymbol{x}) = (q_{0_1}, \ldots, q_{0_m}), \boldsymbol{x} = (x_1, \ldots, x_n)$ to be solution of (A.1) is that the following system of differential equations hold:

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} L_{\frac{\partial q_j}{\partial x_i}}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{x}) - L_{q_j}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{x}) = 0 \qquad j = 1, \ldots, m \qquad (A.2)$$

This fundamental theorem was proven by Leonhard Euler in 1744 which at the same time the initiated variational calculus as a mathematical field.

Well-posed

Definition: An analytical problem is said to be *well-posed* according to Hadamard [Tikhonov and Arsenin 1977] when its solution i) exists, ii) is unique and iii) depends continuously on the data.
Otherwise, it is said to be *ill-posed*.

Convexity

Convex sets

Definition 2. The set $M \subset \mathbb{R}^n$ is said to be *convex* if all points on the connecting line between $u_1, u_2 \in M$ are in M. Hence, M is convex if for all $u_1, u_2 \in M$ and $t \in [0,1]$

$$u_1 + t(u_2 - u_1) = tu_2 + (1-t)u_1 \quad \in M \tag{A.3}$$

Examples for convex sets are the space $\mathbb{R}^n$ and intervals in $\mathbb{R}^n$.

Convex functions

Definition 3. A function $f(u)$ defined on a convex set M is called a convex function if

$$f(tu_2 + (1-t)u_1) \leq tf(u_2) + (1-t)f(u_1) \quad \forall (u_1, u_2) \in M \text{ and } \forall t \in [0,1]. \tag{A.4}$$

In particular, it is said to be *strictly convex* for $u_1 \neq u_2$ if

$$f(tu_2 + (1-t)u_1) < tf(u_2) + (1-t)f(u_1) \quad \text{for } 0 < t < 1. \tag{A.5}$$

Lemma 1. For a continuous differentiable strictly convex function $f(u)$ defined on a convex set M and $(u_1, u_2) \in M$ the following inequality holds:

$$f(u_1) + \text{grad}f(u_1)(u_2 - u_1) < f(u_2) \quad \text{for } u_2 \neq u_1. \tag{A.6}$$

Proof. According to the Taylor expansion we can rewrite $f(u)$ for every $(u_1, u_2) \in M$ as

$$f(u_1 + t(u_2 - u_1)) = f(u_1) + \text{grad}f(\hat{u})\, t(u_2 - u_2) \quad \text{with } t \in [0,1] \tag{A.7}$$

where $\hat{u}$ is between u_1 and u_2. According to the definition of a strictly convex function and recognizing that the left-hand sides of (A.5) and (A.7) are equivalent we obtain the inequality

$$f(u_1) + \text{grad}f(\hat{u})\, t(u_2 - u_2) < tf(u_2) + (1-t)f(u_1)$$

that can be simplified (division by t, $0 < t < 1$ (A.5)) such that

$$f(u_1) + \text{grad}f(\hat{u})(u_2 - u_1) < f(u_2).$$

Since $f'(u)$ is continuous, $\lim_{\hat{u} \to u_1}$ finally leads to

$$f(u_1) + \text{grad}f(u_1)(u_2 - u_1) < f(u_2). \tag{A.8}$$

This holds for all pairs (u_1, u_2) which are allowed in (A.5) and thus the proof is complete. It follows directly form Lemma 1 that u_1 is a global minimum of $f(u)$, if $\text{grad}f(u_1) = 0$. □

The Hessian matrix

Definition 4. The Hessian J of a twice continuously differentiable function $f(x)$ with $x = (x_1, \ldots, x_n)$ is a matrix where the entry at each position i, j is defined as

$$\mathsf{J}_{i,j} = \begin{cases} \partial^2 f(x)/\partial x_i^2 & \text{for } \; i = j \\ \partial^2 f(x)/\partial x_i \partial x_j & \text{otherwise} \end{cases} \tag{A.9}$$

Lemma 2. (with no proof) A function $f(x)$ is convex if its *Hessian* J is positive semi-definite, i.e. $\det(\mathsf{J}) \geq 0$. It is strictly convex if J is positive definite, thus $\det(\mathsf{J}) > 0$.

Global Solution of a Variational Problem with Strictly Convex Integrand

Proposition 1. Given the variational problem

$$H(\boldsymbol{q}) = \int_\Omega L(\boldsymbol{q}(\boldsymbol{x}), \boldsymbol{q}'(\boldsymbol{x}), \boldsymbol{x}) \;\; \mathrm{d}\boldsymbol{x} = \min ! \tag{A.10}$$

with $L : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ of type C^2 and $\boldsymbol{q} : \Omega \to \mathbb{R}^m, \Omega \subset \mathbb{R}^n$. The boundary conditions are free such that $L_{\boldsymbol{q}'}(\boldsymbol{q}(\boldsymbol{\xi}), \boldsymbol{q}'(\boldsymbol{\xi}), \boldsymbol{\xi}) = 0$ with $\boldsymbol{\xi} \in \partial\Omega$.
If the integrand $L(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{x})$ is continuous and strictly convex with respect to the variables $(\boldsymbol{q}, \boldsymbol{q}')$, then a candidate function $\boldsymbol{q}_0$ that fulfills the boundary conditions and the Euler-Lagrange equations is a global solution of the problem.

Proof. In the following we use the short notation $L(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{x}) = L(\boldsymbol{q})$ and $L(\boldsymbol{q}_0, \boldsymbol{q}_0', \boldsymbol{x}) = L(\boldsymbol{q}_0)$ respectively. We now formulate an estimate for the difference:

$$H(\boldsymbol{q}) - H(\boldsymbol{q}_0) = \int_\Omega L(\boldsymbol{q}) - L(\boldsymbol{q}_0) \;\; \mathrm{d}\boldsymbol{x}. \tag{A.11}$$

Since L is strictly convex we introduce the following inequality according to Lemma 1:

$$\begin{aligned} L(\boldsymbol{q}) &> L(\boldsymbol{q}_0) + \mathrm{grad}(L(\boldsymbol{q}_0))(\boldsymbol{q} - \boldsymbol{q}_0) \\ &> L(\boldsymbol{q}_0) + L_{\boldsymbol{q}}(\boldsymbol{q}_0)(\boldsymbol{q} - \boldsymbol{q}_0) + L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q}' - \boldsymbol{q}_0'). \end{aligned} \tag{A.12}$$

Thus we rewrite the difference (A.11) as

$$H(\boldsymbol{q}) - H(\boldsymbol{q}_0) > \int_\Omega L_{\boldsymbol{q}}(\boldsymbol{q}_0)(\boldsymbol{q} - \boldsymbol{q}_0) + L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q}' - \boldsymbol{q}_0') \; \mathrm{d}\boldsymbol{x}. \tag{A.13}$$

Partial integration of the second term under the integral

$$\int_\Omega L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q}' - \boldsymbol{q}_0') \;\; \mathrm{d}\boldsymbol{x} = L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q} - \boldsymbol{q}_0)\Big|_{\partial\Omega} - \int_\Omega \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}} L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q} - \boldsymbol{q}_0) \;\; \mathrm{d}\boldsymbol{x} \tag{A.14}$$

finally leads to

$$H(\boldsymbol{q})-H(\boldsymbol{q}_0) > \int_\Omega \left[L_{\boldsymbol{q}}(\boldsymbol{q}_0)(\boldsymbol{q}-\boldsymbol{q}_0)-\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q}-\boldsymbol{q}_0)\ \mathrm{d}\boldsymbol{x}\right]+L_{\boldsymbol{q}'}(\boldsymbol{q}_0)(\boldsymbol{q}-\boldsymbol{q}_0)\Big|_{\partial\Omega}. \quad \text{(A.15)}$$

The integrand in (A.15) describes the Euler-Lagrange equation which is a necessary condition for a weak solution of the optimization problem. Since we assume $\boldsymbol{q}_0$ to obey the Euler-Lagrange equations, the integral is zero. The second term is also zero according to the boundary condition. Since

$$H(\boldsymbol{q}) - H(\boldsymbol{q}_0) > 0 \qquad \forall\ \boldsymbol{q} \in \mathbb{R}^m, \quad \text{(A.16)}$$

the energy has a single minimum at $\boldsymbol{q}_0$. □

Appendix B

Simulation Methods

The simulation results presented in Chapter 4 are gained by numerically integrating the partial differential equations that describe the dynamics of the analog networks. Integration was performed using the explicit *Euler method*. Step size and stopping criterion were chosen sufficiently small because processing time was not crucial. The network size was determined as equally large as the spatial resolution of the applied image sequence. Although mostly displayed on a subsampled grid (see *e.g.* Figure 4.7), the flow field was always estimated for each pixel. Networks were randomly initialized before computing the first frame. If there were subsequent frames to calculate, no re-initialization took place in between frames.

Gradient Estimation

Discretization effects were reduced by *presmoothing* the image sequences in the spatial and temporal domain [Bertero et al. 1987]. The image sequences were convolved with a Gaussian kernel of fixed width σ. Depending on the image resolution, the kernel size was chosen differently; $\sigma = 0.25$ pixels for the triangle and the tape-rolls sequence and $\sigma = 0.5$ pixels for the other image sequences. After presmoothing, the gradients in each direction were computed as the symmetric difference between the two nearest neighbors.

Image sequences

The triangle sequence

This sequence consists of 20 frames with 64x64 pixel resolution and 8 bit gray scale encoding. It shows a white triangle moving with a speed of one pixel/frame horizontally to the right. The brightness values are normalized to the maximal brightness difference between the darkest points in the background and the surface of the triangle. The contrast

of the stationary sinusoidal plaid pattern in the background is half the maximal figure-ground contrast.

The tape-rolls sequence

This sequence is a real image sequence of two tape-rolls rolling in opposite direction on an office desk. It consists of 20 frames with a spatial resolution of 64x64 pixels and has 6-bit gray level depth. The frame-rate is 15 frames/sec.

The Yosemite sequence

The Yosemite sequence simulates an artificial flight over the Yosemite valley in California, U.S.A and is based on real topological data overlaid with artificially rendered textures. The clouds in the background are completely fractal. The sequence was generated by Lynn Quann at SRI and has been extensively used as a test sequence for optical flow methods.
The sequence consists of 16 frames with a spatial resolution of 316x252 pixels with 8-bit gray level encoding. Simulations were performed on an image partition containing 256x256 pixels.

The Rubic's cube sequence

The Rubic's cube sequence is another, widely used test sequence for visual motion algorithms. It shows a Rubic's cube on a rotating turn-table. The sequence consists of 21 frames each with a resolution of 256x240 pixels and 8-bit gray level depth. In simulations, an image partition with a size of 240x240 pixels was used.

The Hamburg taxi sequence

The sequence shows a typical traffic scene on a crossroad, seen out of the window of a high building. There are several motion sources present, including multiple cars and a pedestrian. The original sequence contains 21 frames with 190x256 pixel resolution. In simulations, an image partition of 190x190 pixels was applied.

The 'triangle' and the 'tape-rolls' sequence are available from the author. The other sequences can be found in various image sequence archives on the web, such as *e.g* `www.cs.cmu.edu/~cil/v-images.html`.

Appendix C

Large Signal MOSFET Model

The **Metal-Oxide-Semiconductor Field-Effect Transistor MOSFET** is the basic computational unit used in the implementation of the presented analog networks. Standard **Complementary-Metal-Oxide-Semiconductor CMOS** processes provide two types of MOSFETs, the native and the well transistor. Labeling of the two types is often based on their channel type, thus calling the n-channel transistor **nFET** (or NMOS) and the p-channel transistor **pFET** (or PMOS). Therefore, in a typical p-type substrate process, the nFET is equivalent to the native and the pFET to the well transistor. Throughout this thesis, schematic symbols for both transistor types are used as depicted in Figure C.1.

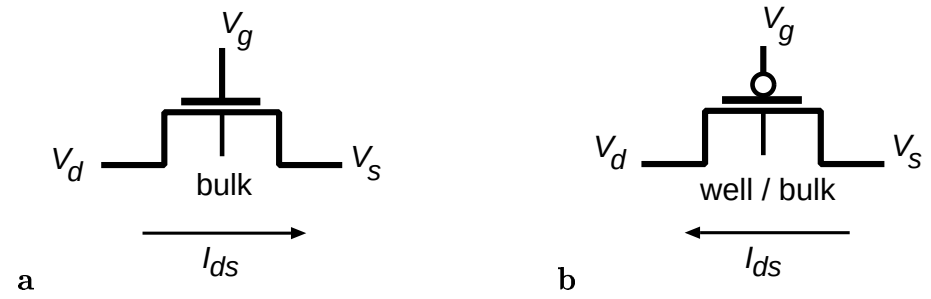


Figure C.1: *Symbolic representation of (a) the native and (b) the well MOSFET.*

The MOSFET has four terminals called *gate (g)*, *bulk (b)*, *drain (d)* and *source (s)*. It is a **symmetric** device insofar as drain and source are not intrinsically defined by the device but depend only on the voltage distribution on the terminals. The source thereby is the terminal, where the net current of majority carriers originates.

Detailed discussions and extensive models of MOSFET devices can be found in the literature. We recommend the excellent textbooks by Gray and Meyer [1993] and Sze [1981].

Sub-threshold regime (weak inversion)

The channel current in a MOSFET is in first order approximated by the difference of the current densities in the *forward* and the *reverse* direction, thus

$$I_{ds} = I_0 \frac{W}{L} [\underbrace{\exp(-\frac{q}{kT}(\kappa V_g - V_s))}_{\text{forward current}} - \underbrace{\exp(-\frac{q}{kT}(\kappa V_g - V_d))}_{\text{reverse current}}] \quad \text{(C.1)}$$

where q is the elementary charge of the majority carriers (i.e. for the nFET $q = -1.6022 \cdot 10^{-19}$ C), k the Boltzmann's constant, T the temperature and $\kappa < 1$ a parameter that accounts for the efficacy of the gate voltage V_g on the channel potential. The source V_s, gate V_g and drain V_d voltage are all referenced to the bulk potential.

κ can be approximated by the capacitor ratio $\kappa \approx C_{ox}/(C_{ox} + C_{depletion})$, where C_{ox} is the capacitance of the gate-oxide and $C_{depletion}$ is the capacitance of the depletion region under the channel. Since $C_{depletion}$ inversely depends on the width of the depletion layer, κ is not constant but increases with increasing gate-bulk potential.

For a drain-source potential difference $V_{ds} > 4kT/q \approx 100$ mV the reverse current can be neglected and the transistor is said to be in **saturation** where

$$I_{sat} = I_0 \frac{W}{L} \exp(-\frac{q}{kT}(\kappa V_g - V_s)). \quad \text{(C.2)}$$

Channel-length modulation – the Early effect

The saturated channel current of the MOSFET is linearly dependent on the width-to-length ratio W/L of the transistor gate as given by Equation (C.2). Although primarily given by the physical size of the gate the effective transistor length L depends also on the depletion width at the drain and source junctions. The depletion width increases with increasing reverse bias of the junctions. Thus, under the assumption of a constant source potential, we have to consider an effective drain conductances g_{ds} caused by the voltage modulation of the channel length. We rewrite the sub-threshold channel current as

$$I_{ds} = I_{sat} + g_{ds} V_{ds} \ , \quad \text{(C.3)}$$

where the drain conductance is defined as

$$g_{ds} = \frac{\partial I_{sat}}{\partial V_{ds}} = \frac{\partial I_{sat}}{\partial L} \frac{\partial L}{\partial V_{ds}} \ . \quad \text{(C.4)}$$

In first approximation we assume the channel-length modulation as negatively proportional to the the drain potential V_{ds}, thus

$$\frac{\partial L}{\partial V_{ds}} = -c_0 \quad \text{(C.5)}$$

where c_0 [m/V] is a constant for a given process and a given transistor length. With

$$\frac{\partial I_{sat}}{\partial L} = -\frac{I_{sat}}{L} \tag{C.6}$$

we find

$$g_{ds} = I_{sat}\ \frac{1}{V_{Early}}\ , \qquad \text{where} \quad \frac{1}{V_{Early}} = \frac{c_0}{L}\ . \tag{C.7}$$

The Early voltage $-V_{Early}$ is the hypothetical drain voltage where the extrapolated sub-threshold drain currents all intersect the voltage axis. It is named after Jim Early who first described this effect in bipolar transistors. Taking into account the effect of channel-length modulation, the total channel current then becomes

$$I_{ds}\ =\ I_{sat}\ (1 + \frac{V_{ds}}{V_{Early}})\ , \tag{C.8}$$

with I_{sat} as given in (C.2).

Above-threshold regime (strong inversion)

In strong inversion, the channel current can be described as a quadratic function of the gate voltage of a reverse and forward component

$$I_{ds}\ =\ \frac{\beta}{2\kappa}\big[\underbrace{(\kappa(V_g - V_{T0}) - V_s)^2}_{\text{forward current}} - \underbrace{(\kappa(V_g - V_{T0}) - V_d)^2}_{\text{reverse current}}\big]\quad . \tag{C.9}$$

V_{T0} is the threshold voltage and $\beta = \mu\ C_{ox}W/L$. The parameter κ does not have any more a clear physical interpretation as in the sub-threshold regime. It remains as a slope factor and is typically $0.5 < \kappa < 1$.

We consider the transistor to be in saturation if the reverse current vanishes, thus if the *saturation condition* $V_d > \kappa(V_g - V_{T0})$ holds. In that case Equation (C.9) simplifies to

$$I_{ds}\ =\ \frac{\beta}{2\kappa}\left((\kappa(V_g - V_{T0}) - V_s)^2\right)\ . \tag{C.10}$$

Appendix D

Differential Pair Circuit

Sub-threshold regime

Consider the differential pair circuit in Figure D.1. In the first instance, we assume all transistors to be in saturation and neglect any second order effects. For simplicity we assume $W/L = 1$ and write all voltages in units of kT/q.

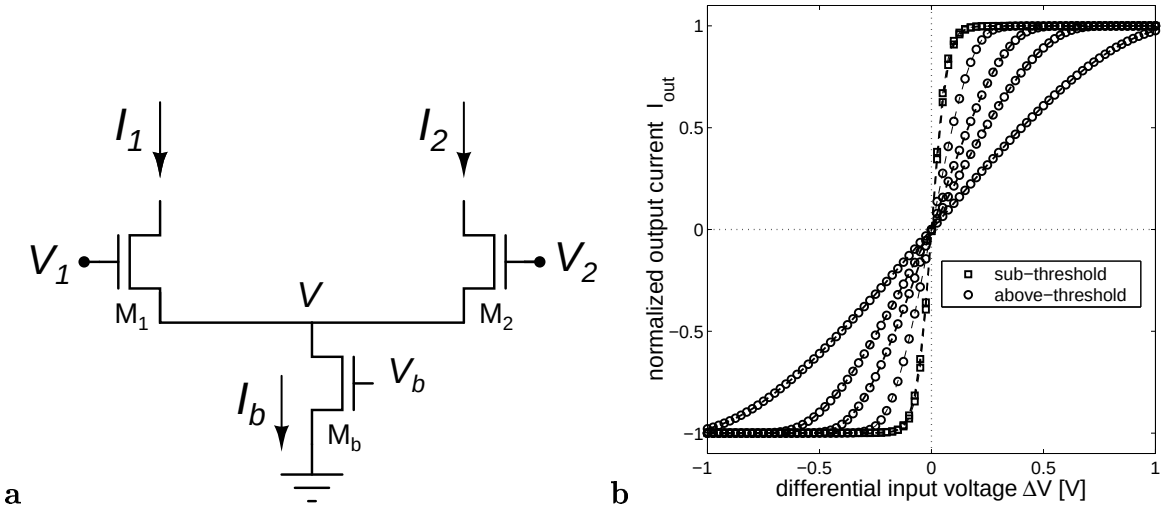


Figure D.1: *Differential pair circuit.* (a) The differential pair circuit. (b) The measured normalized output currents $I_{out} = (I_1 - I_2)/(I_1 + I_2)$ as a function of the differential input voltage $\Delta V = V_1 - V_2$ for sub- and above-threshold bias voltages V_b. For voltages below the threshold voltage ($V_{th} = 0.79$ V) the normalized output curves perfectly overlap because the linear-range is independent of the bias current (square markers). For increasing bias voltages above threshold (V_b = [1, 1.25, 1.5, 2 V]), however, the linear range increases continuously.

The sum of the currents in the two legs of the differential pair equals the bias current provided by M_b, thus

$$I_1 + I_2 = I_b \qquad \text{with} \qquad I_b = I_0 \exp(\kappa V_b). \tag{D.1}$$

They are given as

$$I_1 = I_0 \exp(\kappa V_1 - V) \qquad \text{and} \qquad I_2 = I_0 \exp(\kappa V_2 - V). \tag{D.2}$$

Applying (D.2) in (D.1) we can solve for V and reformulate (D.2) as

$$I_1 = I_b \frac{\exp(\kappa V_1)}{\exp(\kappa V_1) + \exp(\kappa V_2)} \qquad \text{and} \qquad I_2 = I_b \frac{\exp(\kappa V_2)}{\exp(\kappa V_1) + \exp(\kappa V_2)}. \tag{D.3}$$

Multiplying denominator and numerator with $\exp\left(-\kappa(V_1 + V_2)/2\right)$ for both expression in (D.3), the differential output current simplifies to the well known

$$I_{out} = I_1 - I_2 = I_b \tanh\left(\frac{\kappa(V_1 - V_2)}{2}\right). \tag{D.4}$$

The differential pair is a very useful transconductance that provides a nicely behaving relation between its bi-directional output current I_{out} and the difference of the two input voltages, which is ideally independent of the DC-component of the input. The linear range[1], however, is limited to $\pm 1.1kT/q$ which is roughly ± 30 mV at room temperature. Since I_b appears only as a scaling factor in (D.4) the linear range is **independent** of the bias current and stays constant. Therefore, the transconductance of the differential pair

$$g_m = \left.\frac{\partial I_{out}}{\partial \Delta V}\right|_{\Delta V = 0} = \frac{I_b \kappa}{2} \quad \text{with} \quad \Delta V = V_1 - V_2 \tag{D.5}$$

also is linear in the bias current.

Equation (D.4) is valid as long as the bias transistor is in saturation and I_b is approximately constant. If we assume the extreme case $V_1 \gg V_2$, then all the bias current flows though M_1. To keep M_b in saturation ($V \geq 4kT/q$), we find the following input condition to hold:

$$\max(V_1, V_2) > V_b + \frac{4kT}{q} \tag{D.6}$$

Above-threshold regime

We now consider the differential pair circuit operating in strong inversion such that the currents through M_1, M_2 and M_b are accurately described by Equation (C.10). To make

[1] characterized as the voltage range within which the output current I_{out} does not deviate more than 5% from linearity.

things easier we assume β and κ to be equal and constant for all transistors. The currents through the two legs of the differential pair are

$$I_1 = \frac{\beta}{2\kappa}(\kappa(V_1 - V_{T0}) - V)^2 \quad \text{and} \quad I_2 = \frac{\beta}{2\kappa}(\kappa(V_2 - V_{T0}) - V)^2. \tag{D.7}$$

Applying Equations (D.7) to the current equilibrium

$$I_b \;=\; I_1 + I_2 \quad \text{with} \quad I_b = \frac{\beta\kappa}{2}(V_b - V_{T0})^2 \tag{D.8}$$

leads to a quadratic expression in V which has the following solutions:

$$V \;=\; \frac{\kappa}{2}\left((V_1 - V_{T0}) + (V_2 - V_{T0}) \pm \sqrt{\frac{4I_b}{\beta\kappa} - \Delta V^2} \right) \tag{D.9}$$

with $\Delta V = V_1 - V_2$. Since V is preferably lower than V_1 or V_2 we consider the smaller of the two solutions. Substitution of V in Equations (D.7) then gives

$$I_1 = \frac{\beta\kappa}{8}\left(\Delta V + \sqrt{\frac{4I_b}{\beta\kappa} - \Delta V^2} \right)^2 \quad \text{and} \quad I_2 = \frac{\beta\kappa}{8}\left(-\Delta V + \sqrt{\frac{4I_b}{\beta\kappa} - \Delta V^2} \right)^2 \tag{D.10}$$

Finally, the differential output current becomes

$$I_{out} \;=\; \frac{\beta\kappa}{2}\Delta V \sqrt{\frac{4I_b}{\beta\kappa} - \Delta V^2} \quad . \tag{D.11}$$

Under which conditions for the input voltages V_1 and V_2 is Equation (D.11) a good approximation of the above-threshold behavior of the differential pair?

Firstly, we have to ensure that M_b is in saturation. According to (C.10) this is the case if $V > \kappa(V_b - V_{T0})$. Since V is lowest if the complete bias current flows through either one of the two legs, the lower *saturation limit* for one of the input voltages follows as

$$\max(V_1, V_2) > V_b + (V_b - V_{T0}). \tag{D.12}$$

Secondly, we also have to make sure that the transistors M_1 and M_2 stay in above-threshold operation. Thus we can determine the upper limit of the input voltage $|\Delta V|$. Assume that M_2 is right at the threshold voltage[2] which is defined as $\kappa(V_2 - V_{T0}) - V = 0$. Furthermore, we neglect the residual sub-threshold current in M_2. Then, substitution of

[2]Correctly spoken, at threshold we are in the region of *moderate inversion* which spans roughly $200mV$ around the threshold voltage. Thus, the above assumption overestimates the maximal input range that ensures above-threshold operation.

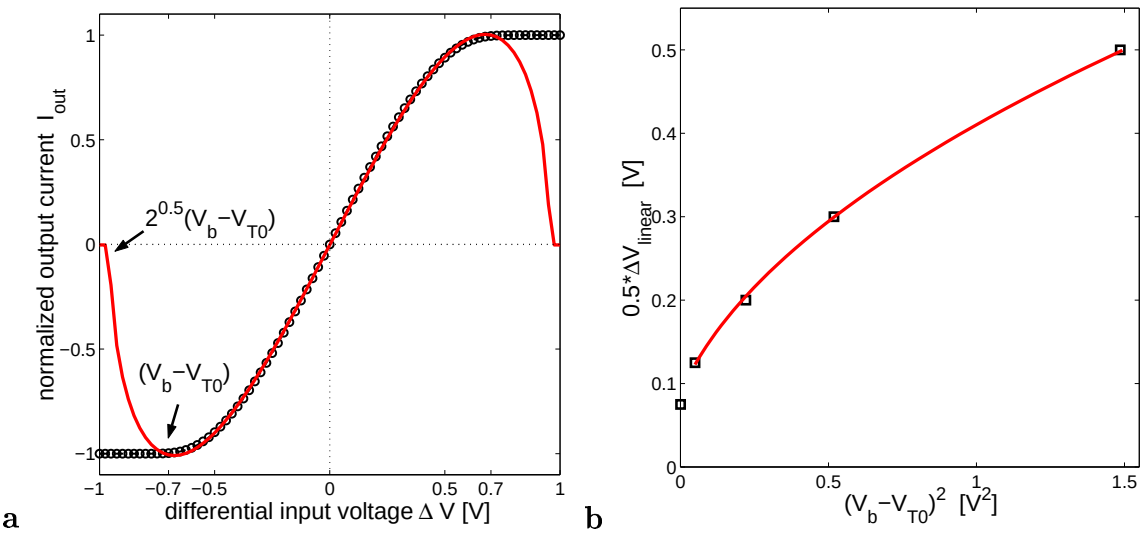


Figure D.2: *Above-threshold operation characteristics.* (a) The measured normalized output current I_{out} is shown as a function of the differential input voltage $\Delta V = V_1 - V_2$ for $V_b = 1.5$ V. According to Equation (D.11), a least-square-fit is performed (bold curve). The range for which the fit is accurate closely matches the theoretically found input limit $V = (V_b - V_{T0})$ (with $V_{T0} = 0.79$ V). When the square-root in Equation (D.11) becomes negative, the real part of the fitting curve is zero. This is observed at ± 0.98 V which is approximately $\sqrt{2}(V_b - V_{T0})$. (b) In above-threshold operation, the linear range (measured 5% limit) increases inversely proportionally to the normalized transconductance g_m/I_b, thus proportionally to the square-root of the bias current. The measured values match closely an appropriate fitting function (bold line).

$V = \kappa(V_2 - V_{T0})$ in I_1 (D.7) and the approximation $I_1 = I_b$ leads to the second condition, the *input limit*

$$\Delta V \leq (V_b - V_{T0}). \tag{D.13}$$

We might have expected that Equation (D.11) is valid as long its square-root is positive. As it turned out now, the valid input range is smaller by a factor of $\sqrt{2}$.

Figure D.2 shows the measured output current and a least-square error fit based on Equation (D.11). As we see, the input limit is not of much importance because the real circuit, unlike the fitting function, behaves very politely: Going beyond the input limit barely changes the output current anymore because the differential pair is already close to its saturation condition where I_b is flowing either to one or the other leg. Note, how the fitting function slightly overestimates the output current around the derived input limit (D.13). As mentioned, this is mainly because we assumed strong inversion to hold right down to the threshold voltage and neglected the region of moderate inversion. As a consequence, the current through the transistor that falls out of strong inversion does not

decrease as strongly and thus, the differential output current is larger than predicted.

In above-threshold operation, the transconductance of the differential pair is

$$g_m = \sqrt{I_b \beta \kappa} \quad . \tag{D.14}$$

Clearly, the transconductance is no longer proportional to the bias current I_b, so the linear range **increases** as I_b gets larger. Under the assumption that the characteristic shape of the normalized output currents remains similar for different bias current levels, the linearity limits will occur at constant current levels. Then, the linear range is inversely proportional to the normalized transconductance:

$$\Delta V_{\text{linear}} \propto \left(\frac{g_m}{I_b} \right)^{-1} \propto \sqrt{I_b}. \tag{D.15}$$

Figure D.2b shows the measured linear range and a least-square fit according to (D.15).

Appendix E

Process Parameters

The **Austria Mikro Systeme AMS** 0.8 μm BiCMOS process is a mixed analog/digital process and offers the implementation of generic npn-bipolar transistor. It provides two layers of metal and two layers of low ohmic polysilicon (poly) to form poly/poly capacitors. The substrate is p- doped and the gates consist of n+ doped polysilicon. The most important process parameters are listed below in Table E.1.

MOSFET Transistor	*NMOS*	*PMOS*	units
Effective substrate doping	74x10^{15}	28x10^{15}	cm^{-3}
Minimum drawn gate length	0.8	0.8	μm
Oxide thickness	16	16	nm
Threshold voltage 20/20	0.80	-0.84	V
Transconductance	100	35	$\mu A/V^2$
Body factor 20/20	0.74	0.45	$\sqrt{V}$
Bipolar Transistor	*NPN*		
Beta	100		-
Early-voltage V_{Early}	32		V
Transit frequency F_t	12		GHz
Thin Oxide Poly/Poly Capacitor	1.8		$fF/\,\mu\text{m}^2$
Operating Voltage	2.5 - 5.5		V

Table E.1: AMS 0.8 μm BiCMOS process parameters.

Both, the optical flow chip and the motion segmentation chip are fabricated in this AMS process. Fabrication costs could be kept relatively low by accessing the process through the multi-project-wafer (MPW) service provided by Europractice[1].

[1] www.europractice.com

References

Adelson, E. and J. Bergen (1985, February). Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America 2*(2), 284–299.

Adelson, E. and J. Movshon (1982, December). Phenomenal coherence of moving visual patterns. *Nature 300*(9), 523–525.

Allman, J., F. Miezin, and E. McGuinness (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Reviews in Neuroscience 8*, 407–430.

Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. *Intl. Journal of Computer Vision 2*, 283–310.

Andreou, A. and K. Strohbehn (1990). Analog VLSI implementation of the Hassenstein-Reichardt-Poggio models for vision computation. *Procs. of the 1990 Intl. Conference on Systems, Man and Cybernetics*.

Andreou, A., K. Strohbein, and R. Jenkins (1991). Silicon retina for motion computation. In *Procs. IEEE Intl. Symposium on Circuits and Systems*, pp. 1373–1376. IEEE.

Arreguit, X., A. van Schaik, F. Bauduin, M. Bidiville, and E. Raeber (1996, December). A CMOS motion detector system for pointing devices. *IEEE Journal of Solid-States Circuits 31*(12), 1916–1921.

Aubert, G., R. Deriche, and P. Kornprobst (1999). Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics 60*(1), 156–182.

Barlow, H. and W. Levick (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology 178*, 477–504.

Barron, J., D. Fleet, and S. Beauchemin (1994). Performance of optical flow techniques. *Intl. Journal of Computer Vision 12*(1), 43–77.

Benson, R. and T. Delbruck (1992). Direction-selective silicon retina that uses null inhibition. In D. Touretzky (Ed.), *Neural Information Processing Systems 4*, pp. 756–763. MIT Press.

Bertero, M., T. Poggio, and V. Torre (1987, May). Ill-posed problems in early vision. Technical Report 924, MIT AI Lab.

Braddick, O. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society London B 290*, 137–151.

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neuroscience 16*(7), 263–268.

Bradley, D. and R. Andersen (1998, September). Center-surround antagonism based on disparity in primate area MT. *The Journal of Neuroscience 15*, 7552–7565.

Bradley, D., N. Qian, and R. Andersen (1995, February). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature 373*, 609–611.

Bremmer, F. and M. Lappe (1999, July). The use of optical velocites for distance discrimination and reproduction during visually simulated self-motion. *Experimental Brain Research 127*(1), 33–42.

Bronstein, I. and K. Smendjajew (1996). *Teubner-Taschenbuch der Mathematik*. Stuttgart, Leibzig: B.G. Teubner.

Bülthoff, H., J. Little, and T. Poggio (1989, February). A parallel algorithm for real-time computation of optical flow. *Nature 337*(9), 549–553.

Burke, D. and P. Wenderoth (1993). The effect of interactions between one-dimensional component gratings on two-dimensional motion perception. *Vision Research 33*(3), 343–350.

Camus, T. (1994, September). *Real-Time Optical Flow*. Ph. D. thesis, Brown University, Dep. Computer Science.

Cesmeli, E. and D. Wang (2000, July). Motion segmentation based on motion/brightness integration and oscillatory correlation. *IEEE Trans. on Neural Networks 11*(4), 935–947.

Chahine, M. and J. Konrad (1995, November). Esimtation and compensation of accelerated motion for temporal sequence interpolation. *Signal Processing, Image Communication 7*, 503–527.

Chang, M., A. Tekalp, and M. Sezan (1997, September). Simultaneous motion estimation and segmentation. *IEEE Trans. on Image Processing 9*(6), 1326–1333.

Cichocki, A. and R. Unbehauen (1993). *Neural Networks for Optimization and Signal Processing* (3rd ed.). Wiley and Sons.

Delbruck, T. (1993a, May). Bump circuits. Technical Report CNS Memo 26, Caltech, Pasadena, California 91125.

Delbruck, T. (1993b). *Investigations of analog VLSI visual transduction and motion processing*. Ph. D. thesis, Department of Computational and Neural Systems, California Institute of Technology, Pasadena, CA.

Delbruck, T. (1993c, May). Silicon retina with correlation-based velocity-tuned pixels. *IEEE Trans. on Neural Networks 4*(3), 529–541.

Delbruck, T. and C. Mead (1994). Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor cirucuits. Technical Report 30, Caltech Computation and Neural Systems Program.

Deutschmann, R. and C. Koch (1998a). An analog VLSI velocity sensor using the gradient method. In *Procs. IEEE Intl. Symposium on Circuits and Systems*, pp. 649–652. IEEE.

Deutschmann, R. and C. Koch (1998b). Compact real-time 2D gradient-based analog VLSI motion sensor. In *Intl. Conference on Advanced Focal Plane Arrays and Electronic Cameras*. AFPAEC.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In A. Dold and B. Eckmann (Eds.), *Constructive Theory of Functions of Several Variables*, pp. 85–100. Springer Verlag Berlin.

Duffy, C. (2000). Optic flow analysis for self-movement perception. In M. Lappe (Ed.), *Neural Processing of Optic Flow*, pp. 199–218. Academic Press.

Duffy, C. and R. Wurtz (1993). An illusory transformation of optic flow fields. *Vision Research 33*(11), 1481–1490.

Etienne-Cummings, R. (1993). *Biologically Motivated Analog VLSI Systems for Optomotor Tasks*. Ph. D. thesis, University of Pennsylvania.

Etienne-Cummings, R., S. Fernando, N. Takahashi, and J. Van der Spiegel (1993, December). A new temporal domain optical flow measurement technique for focal plane VLSI implementation. In *CAMP*.

Etienne-Cummings, R., J. Van der Spiegel, and P. Mueller (1999, September). Hardware implementation of a visual-motion pixel using oriented spatiotemporal neural filters. *IEEE Trans. on Circuits and Systems 2 46*(9), 1121–1136.

Etienne-Cummings, R., J. Van der Spiegel, N. Takahashi, and P. Mueller (1996, November). VLSI implementation of cortical visual motion detection using an analog neural computer. In *Advances in Neural Information Processing Systems 8*.

Fennema, C. and W. Thompson (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing 9*, 301–315.

Ferrera, V. and H. Wilson (1990). Perceived direction of moving two-dimensional patterns. *Vision Research 30*(2), 273–287.

Fleet, D. and A. Jepson (1990). Computation of component image velocity from local phase information. *Intl. Journal of Computer Vision 5*(1), 77–104.

Fletcher, R. (1980). *Practical Methods of Optimization*. John Wiley and Sons. Volume 1.

Fletcher, R. (1981). *Practical Methods of Optimization*. John Wiley and Sons. Volume 2.

Fukushima, K., Y. Yamaguchi, M. Yasuda, and S. Nagata (1970, December). An electronic model of the retina. *Proceedings of the IEEE 58*(12), 1950–1951.

Gamble, E. and T. Poggio (1987, October). Visual integration and detection of discontinuities: The key role of intensity edges. Technical Report 970, MIT AI Lab.

Gee, A. and R. Prager (1995, January). Limitations of neural networks for solving Traveling Salesman problems. *IEEE Trans. on Neural Networks 6*(1), 280–282.

Geman, S. and D. Geman (1984, November). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 6*(6), 721–741.

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston, MA: Houghton Mifflin.

Gray, P. and R. Meyer (1993). *Analysis and Design of Analog Integrated Circuits* (3 ed.). New York: Wiley and sons.

Grossberg, S. (1978). Competition, decision and consensus. *Journal of Mathematical Analysis and Applications 66*, 470–493.

Grzywacz, N. and A. Yuille (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Procs. of the Royal Society of London B*(239), 129–161.

Grzywacz, N. and A. Yuille (1991). Theories for the visual perception of local velocity and coherent motion. In M. S. Landy and J. A. Movshon (Eds.), *Computational Models of Visual Processing*, pp. 231–252. Cambridge, Massachusetts: MIT Press.

Hahnloser, R., R. Douglas, M. Mahowald, and K. Hepp (1999, August). Feedback interactions between neuronal pointers and maps for attentional processing. *Nature Neuroscience 2*(8), 746–752.

Harris, J. (1991). *Analog Models for Early Vision*. Ph. D. thesis, California Institute of Technology, Pasadena.

Harris, J. and C. Koch (1989, April). Resistive fuses: circuit implementations of line discontinuities in vision. Snowbird Neural Network Workshop.

Harris, J., C. Koch, and J. Luo (1990, June). A two-dimensional analog VLSI circuit for detecting discontinuities in early vision. *Science 248*, 1209–1211.

Harris, J., C. Koch, E. Staats, and J. Luo (1990). Analog hardware for detecting discontinuities in early vision. *Intl. Journal of Computer Vision 4*, 211–223.

Harrison, R. and C. Koch (1998). An analog VLSI model of the fly elementary motion detector. In M. Kearns and S. Solla (Eds.), *Advances in Neural Information Processing Systems 10*, pp. 880–886. MIT Press.

Hassenstein, B. and W. Reichardt (1956). Systemtheoretische Analyse der Zeitreihenfolgen und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforschung 11b*, 513–524.

Heeger, D. (1987a). A model for the extraction of image flow. *Optical Society of America*, 151–154.

Heeger, D. (1987b, August). Model for the extraction of image flow. *Journal of the Optical Society of America 4* (8), 1455–1471.

Hertz, J., A. Krogh, and R. Palmer (1991). *Introduction to the theory of neural computation*. Lecture Notes - Santa Fe Institute. Perseus Books.

Higgins, C. and C. Koch (1997, Februar). Analog CMOS velocity sensors. *Procs. of Electronic Imaging SPIE 3019*.

Higgins, C. and C. Koch (1999). An integrated vision sensor for the computation of optical flow singular points. In M. S. Kearns, S. A. Solla, and D. A. Cohn (Eds.), *Advances in Neural Information Processing Systems 11*, Cambridge M. MIT Press.

Hildreth, E. C. (1983). *The Measurment of Visual Motion*. MIT Press.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks 2*, 331–349.

Hopfield, J. (1982, April). Neural networks and physical systems with emergent collective computational abilities. *Procs. National Academic Sciences U.S.A. 79*, 2554–2558.

Hopfield, J. (1984, May). Neurons with graded response have collective computational properties like those of two-state neurons. *Procs. National Academic Sciences U.S.A. 81*, 3088–3092.

Hopfield, J. and D. Tank (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics* (52), 141–152.

Horiuchi, T., B. Bishofberger, and C. Koch (1994). An analog VLSI saccadic eye movement system. In J. Cowan, G. Tesauro, and J. Alspector (Eds.), *Advances in Neural Information Processing Systems 6*, pp. 5821–5889. San Mateo, CA: Morgan Kaufmann.

Horiuchi, T., J. P. Lazzaro, A. Moore, and C. Koch (1991). A delay-line based motion detection chip. In R. Lippman, J. Moody, and D. Touretzky (Eds.), *Advances in Neural Information Processing Systems 3*, Volume 3, pp. 406–412. San Mateo, CA: Morgan Kaufmann.

Horn, B. and B. Schunck (1981). Determining optical flow. *Artificial Intelligence 17*, 185–203.

Horn, B. K. (1988, December). Parallel networks for machine vision. Technical Report 1071, MIT AI Lab.

Huang, L. and Y. Aloimonos (1991, October). Relative depth from motion using normal flow: an active and purposive solution. In *IEEE Workshop on Visual Motion*, pp. 196–203. IEEE Computer Society: IEEE Computer Society Press.

Hubel, D. and A. Wiesel (1962). Receptive fields, binocular interaction and functional architecture in the cats visual cortex. *Journal of Physiology 160*, 106–154.

Hutchinson, J., C. Koch, J. Luo, and C. Mead (1988, March). Computing motion using analog and binary resistive networks. *Computer 21*, 52–64.

Indiveri, G., J. Kramer, and C. Koch (1996). Parallel analog VLSI architectures for computation of heading direction and time-to-contact. In *Advances in Neural Information Processing Systems 8*, pp. 720–726.

Jagota, A. (1995, May). Approximating maximum clique with a Hopfield network. *IEEE Trans. on Neural Networks 6*(3), 724–735.

Kamgar-Parsi, B. and B. Kamgar-Parsi (1990). On probelm solving with Hopfield neural networks. *Biological Cybernetics 62*, 415–423.

Kaski, S. and T. Kohonen (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks 7*(6/7), 973–984.

Kastner, S. and L. Ungerleider (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neurosciences 23*, 315–341.

Koch, C., H. Wang, R. Battiti, B. Mathur, and C. Ziomkowski (1991, October). An adaptive multi-scale approach for estimating optical flow: Computational theory and physiological implementation. *IEEE Workshop on Visual Motion, Princeton, New Jersey*, 111–122.

Koch, C., H. Wang, B. Mathur, A. Hsu, and H. Suarez (1989, March). Computing optical flow in resistive networks and in the primate visual system. In *Workshop on Visual Motion*, pp. 62–72. IEEE Computer Society: IEEE Computer Society Press.

Kramer, J. (1996, April). Compact integrated motion sensor with three-pixel interaction. *IEEE Trans. on Pattern Analysis and Machine Intelligence 18*(4), 455–460.

Kramer, J., R. Sarpeshkar, and C. Koch (1995). An analog VLSI velocity sensor. In *Intl. Symposium on Circuits and Systems*, pp. 413–416. IEEE.

Kramer, J., R. Sarpeshkar, and C. Koch (1996). Analog VLSI motion discontinuity detectors for image segmentation. In *Intl. Symposium on Circuits and Systems*, pp. 620–623. IEEE.

Kramer, J., R. Sarpeshkar, and C. Koch (1997, February). Pulse-based analog VLSI velocity sensors. *IEEE Trans. on Circuits and Systems 2 44*(2), 86–101. 86.

Lamme, V., B. van Dijk, and H. Spekreijse (1993, June). Contour from motion processing occurs in primary visual cortex. *Nature 363*, 541–543.

Lappe, M. (2000). Computational mechnisms for optic flow analysis in primate cortex. In M. Lappe (Ed.), *Neural Processing of Optic Flow*, pp. 235–268. Academic Press.

Lappe, M., F. Bremmer, M. Pekel, A. Thiele, and K.-P. Hoffmann (1996, October). Optic flow processing in monkey STS: A theoretical and experimental approach. *Journal of Neuroscience*, 6265–6285.

Liang, X.-B. and J. Wang (2000, November). A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints. *IEEE Trans. on Neural Networks 11*(6), 1251–1262.

Limb, J. and J. Murphy (1975). Estimating the velocity of moving images in television signals. *Computer Graphics Image Processing 4*, 311–327.

Little, J. and A. Verri (1989, March). Analysis of differential and matching methods for optical flow. In *Workshop on Visual Motion*, pp. 173–180. IEEE Computer Society: IEEE Computer Society Press.

Little, W. and G. Shaw (1975). A statistical theory of short and long term memory. *Behavioral Biology 14*, 115–133.

Liu, S.-C. (1996, June). Silicon model of motion adaptation in the fly visual system. *Procs. 3rd. Joint Symposium on Neural Computation*.

Liu, S.-C. (1998, November). Silicon retina with adaptive filtering properties. In *Advances in Neural Information Processing Systems 10*, pp. 712–718. MIT Press.

Liu, S.-C. (2000, December). A neuromorphic aVLSI model of global motion processing in the fly. *IEEE Trans. on Circuits and Systems 2 47*(12), 1458–1467.

Liu, S.-C. and J. Harris (1992). Dynamic wires: An analog VLSI model for object-based processing. *Intl. Journal of Computer Vision 8:3*, 231–239.

Lorenceau, J., M. Shiffrar, N. Wells, and E. Castet (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research 33*(9), 1207–1217.

Lucas, B. and T. Kanade (1981). An iterative image registration technique with an application to stereo vision. In *Procs. Image Understanding Workshop*, pp. 121–130. DARPA.

Mahowald, M. (1994). *An Analog VLSI System for Stereoscopic Vision*. Boston: Kluwer.

Mahowald, M. and C. Mead (1991, May). The silicon retina. *Scientific American*, 76–82.

Marroquin, J. (1985, April). Optimal Bayesian estimators for image segmentation and surface reconstruction. Technical Report 839, MIT AI Lab.

Marroquin, J., S. Mitter, and T. Poggio (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association 82*(397), 76–89.

McCulloch, S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics 5*, 115–133.

Mead, C. (1989). *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.

Mead, C. (1990, October). Neuromorphic electronic systems. *Proceedings of the IEEE 78*(10), 1629–1636.

Memin, E. and P. Perez (1998, May). Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. on Image Processing 7*(5), 703–719.

Moini, A., A. Bouzerdoum, K. Eshraghian, A. Yakovleff, X. Nguyen, A. Blanksby, R. Beare, D. Abbott, and R. Bogner (1997, February). An insect vision-based motion detection chip. *IEEE Journal of Solid-State Circuits 32*(2), 279–283.

Moore, A. and C. Koch (1991). A multiplication based analog motion detection chip. In B. Mathur and C. Koch (Eds.), *SPIE Visual Information Processing: From Neurons to Chips*, Volume 1473, pp. 66–75. SPIE.

Movshon, J., E. Adelson, M. Gizzi, and W. Newsome (1985). The analysis of moving visual patterns. *Experimental Brain Research Supplementum 11*, 117–151.

Murray, D. and B. Buxton (1987, March). Scene segmentation from visual motion using global optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence 9*(2), 220–228.

Nakayama, K. and G. Silverman (1988a). The aperture problem - 1. perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research 28*(6), 739–746.

Nakayama, K. and G. Silverman (1988b). The aperture problem - 2. spatial integration of velocity information along contours. *Vision Research 28*(6), 747–753.

Nowlan, S. J. and T. J. Sejnowski (1994). Filter selection model for motion segmentation and velocity integration. *Journal of Optical Society of America 11*, 3177–3200.

Pack, C. and R. Born (2001, February). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature 409*, 1040–1042.

Perrone, J. and A. Thiele (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience 4*(5), 526–532.

Pesavento, A. and C. Koch (1999). Feature detection in analog VLSI. In *Procs. 33rd Asiolmar Conference on Signals, Systems and Computers*, Monterey USA.

Platt, J. (1989, July). *Constraint Methods for Neural Networks and Computer Graphics*. Ph. D. thesis, California Institute of Technology, Dept. of Computer Science.

Poggio, T., V. Torre, and C. Koch (1985, September). Computational vision and regularization theory. *Nature 317*(26), 314–319.

Rao, R. and D. Ballard (1996). The visual cortex as a hierachical predictor. Technical Report 96.4, Dept. of Computer Science, University of Rochester.

Rao, R. and D. Ballard (1999, January). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature 2*(1), 79–87.

Reppas, J., S. Niyogi, A. Dale, M. Sereno, and R. Tootell (1997, July). Representation of motion boundaries in retinotopic human visual coritcal areas. *Nature 388*, 175–179.

Sarpeshkar, R., W. Bair, and C. Koch (1993). Visual motion computation in analog VLSI using pulses. In *Advances in Neural Information Processing Systems 5*, pp. 781–788.

Sawaji, T., T. Sakai, H. Nagai, and T. Matsumoto (1998). A floating-gate MOS implementation of resistive fuse. *Neural Computation 10*(2), 486–498.

Schrater, P., D. Knill, and E. Simoncelli (2000, January). Mechanisms of visual motion detection. *Nature Neuroscience 3*(1), 64–68.

Schunck, B. (1989, October). Image flow segmentation and estimation by constraint line clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence 11*(10), 1010–1027.

Singh, A. (1991). *Optic Flow Computation: A Unified Perspective*. Los Alamitos CA: IEEE Computer Society Press.

Stiller, C. and J. Konrad (1999, July). Estimating motion in image sequences. *IEEE Signal Processing 16*(4), 71–91.

Stocker, A. (1998, January). Smooth optical flow computation by a constraint solving circuit. *Newsletter in Neural Networks 11*(1).

Stocker, A. and R. J. Douglas (1999). Computation of smooth optical flow in a feedback connected analog network. In M. S. Kearns, S. A. Solla, and D. A. Cohn (Eds.), *Advances in Neural Information Processing Systems 11*, Cambridge, MA, pp. 706–712. MIT Press.

Stone, L., A. Watson, and J. Mulligan (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Research 30*(7), 1049–1067.

Sundareswaran, V. (1991, October). Egomotion from global flow field data. In *IEEE Workshop on Visual Motion*, pp. 140–145. IEEE Computer Society: IEEE Computer Society Press.

Sze, S. (1981). *Physics of Semiconductor Devices* (2 ed.). New York: John Wiley and Sons.

Szeliski, R. (1996). Regularization in neural nets. In P. Smolensky, M. Mozer, and D. Rumelhart (Eds.), *Mathematical Perspectives on Neural Networks*, pp. 497–532. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Tank, D. and J. Hopfield (1986, May). Simpel "neural" optimization networks: An A/D converter, signal decision circuit and a linear programming circuit. *IEEE Trans. of Circuits and Systems 33*(5), 533–541.

Tanner, J. (1986). *Integrated Optical Motion Detection.* Ph. D. thesis, California Institute of Technology. 5223:TR:86.

Tanner, J. and C. Mead (1986). An integrated analog optical motion sensor. In S.-Y. Kung, R. Owen, and G. Nash (Eds.), *VLSI Signal Processing, 2*, pp. 59 ff. IEEE Press.

Terzopolous, D. (1986, July). Regularization of inverse visual problems involving discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence 8*(4), 413–423.

Thompson, W. (1980, November). Combining motion and contrast for segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence 2*(6), 543–549.

Tikhonov, A. and V. Arsenin (1977). *Solutions of Ill-posed Problems.* Washington D.C. 20005: Winston and sons, Scripta Technica.

Treue, S. and J. Maunsell (1996, August). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature 382*, 539–541.

Treue, S. and J. Trujillo (1999, June). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature 399*, 575–579.

Ullman, S. (1979). *The Interpretation of Visual Motion.* Cambridge MA: MIT Press.

Ullman, S. and A. Yuille (1987, November). Rigidity and smoothness of motion. Technical Report 989, MIT AI Lab.

Van Santen, J. and G. Sperling (1984). Temporal covariance model of human motion perception. *Journal of Optical Society of America A1*, 451–473.

Van Santen, J. and G. Sperling (1985, February). Elaborated Reichardt detectors. *Journal of Optical Society of America 2*(2), 300–320.

Verri, A., F. Girosi, and V. Torre (1990). Differential techniques for optical flow. *Journal of Optical Society of America 7*(5), 912–922.

Verri, A. and T. Poggio (1989). Motion field and optical flow: Qualitative properties. *IEEE Trans. Pattern Analysis and Machine Intelligence 11*(5), 490–498.

Vittoz, E. (1989). Analog VLSI implementation of neural networks. In *Procs. of Journees d'Electronique et Artificial Neural Networks*, pp. 223–250. press politechnique.

von Neumann, J. (1982/1945). First draft of a report on the EDVAC. In B. Randall (Ed.), *The Origins of Digital Computers: Selected Papers*. Berlin: Springer.

Watson, A. and A. Ahumada (1985). Model of human visual motion sensing. *Journal of the Optical Society of America A 2*, 322–342.

Weiss, Y. and E. Adelson (1998, February). Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision. Technical Report 1624, MIT AI Lab.

Wilson, G. and G. Pawley (1988). On the stability of the Traveling Salesman Problem algorithm of Hopfield and Tank. *Biological Cybernetics 58*, 63–70.

Wu, S. and J. Kittler (1993, March). A gradient-based method for general motion estimation and segmentation. *Journal of Visual Communication and Image Representation 4*(1), 25–38.

Yo, C. and H. Wilson (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research 32*(1), 135–147.

Yu, P., S. Decker, H.-S. Lee, C. Sodini, and J. Wyatt (1992, April). CMOS resistive fuses for image smoothing and segmentation. *IEEE Journal of Solid State Circuits 27*(4), 545–553.

Yuille, A. (1989). Energy functions for early vision and analog networks. *Biological Cybernetics 61*, 115–123.

Yuille, A. and N. Grzywacz (1988, May). A computational theory for the perception of coherent visual motion. *Nature 333*(5), 71–74.

Yuille, A. and N. Grzywacz (1989). A mathematical analysis of the motion coherence theory. *Intl. Journal of Computer Vision 3*, 155–175.

Index

Curriculum Vitae

Alan Stocker

Institute of Neuroinformatics
University and ETH Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland

Born 12.1.1970 in Lachen SZ, Switzerland
Swiss

Education

Doctoral student at the Institute of Neuroinformatics	1996 - 2001
Undergraduate studies at ETH Zürich, Dipl. Masch.-Ing. ETH (M. Sc.)	1990 - 1995
Majors: Biomedical Engineering and Material Science	
Secondary school at Kantonsschule Pfäffikon SZ, Matura Typus C	1984 - 1989
Primary school at Primarschule Lachen SZ	1976 - 1984

Research Experience

Research and teaching assistant at the Institute of Neuroinformatics	1996 - 2001
Supervisor of semester and diploma theses	
Telluride workshop on 'Neuromorphic Engineering'	1997, 1998
Member IEEE	since 1998