

Periodic Motion Detection and Segmentation via Approximate Sequence Alignment

Ivan Laptev[†], Serge J. Belongie[‡], Patrick Pérez[†] and Josh Wills[‡]

[†] IRISA / INRIA Rennes
Campus universitaire de Beaulieu
35042 Rennes Cedex France
{ilaptev | perez}@irisa.fr

[‡] Dept. of Computer Science and Engineering
UC San Diego
La Jolla, CA 92093-0404
{sjb | jwills}@cs.ucsd.edu

Abstract

A method for detecting and segmenting periodic motion is presented. We exploit periodicity as a cue and detect periodic motion in complex scenes where common methods for motion segmentation are likely to fail. We note that periodic motion detection can be seen as an approximate case of sequence alignment where an image sequence is matched to itself over one or more periods of time. To use this observation, we first consider alignment of two video sequences obtained by independently moving cameras. Under assumption of constant translation, the fundamental matrices and the homographies are shown to be time-linear matrix functions. These dynamic quantities can be estimated by matching corresponding space-time points with similar local motion and shape. For periodic motion, we match corresponding points across periods and develop a RANSAC procedure to simultaneously estimate the period and the dynamic geometric transformations between periodic views. Using this method, we demonstrate detection and segmentation of human periodic motion in complex scenes with non-rigid backgrounds, moving camera and motion parallax.

1. Introduction

Periodic motion is ubiquitous in the natural world, with instances ranging from the simple harmonic motion of a pendulum to the complex movements of an Olympic runner. It has been widely recognized as a powerful cue for detecting salient objects or gestures in video; see for example [20, 16, 12, 5, 13, 21]. The focus of this paper is on leveraging the periodic motion of objects for a different, and thus far unexplored, purpose: motion segmentation.

In the conventional motion segmentation problem, one considers a scene containing moving rigid objects. Then,

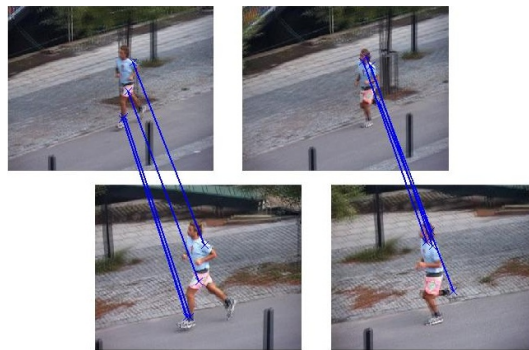


Figure 1. Illustration of periodic motion for a runner. The pose of the person is similar over the periods while the position relative to the camera is different. Hence, period-separated views can approximately be treated as stereo pairs. Matching of corresponding space-time points in video can be used to derive object-centered alignment of a video sequence to itself over one or more cycles of periodic motion.

according to a two- or multi-view geometric relation such as planar homography H or a fundamental matrix F , each object emerges as a group by virtue of adhering to a relation with certain parameters. Some early examples of this for the case of approximately planar motion layers include [27, 6]; more recent examples addressing the case of independently moving 3D objects include [26, 22]. Central in all of these approaches is the assumption that the motion of the object relative to the camera be rigid. Without rigidity, the classical point-to-point and point-to-line relations are rendered invalid. While a number of recent works have appeared that attempt to extend motion segmentation and structure from motion algorithms to the non-rigid case, e.g., by using rigid basis parts [23, 29], such approaches do not avail of the special properties of periodic motion when it is present. The special property in particular that we ex-

plot in this paper is the following: given a monocular video sequence of a periodic moving object, any set of period-separated frames represents a collection of snapshots of a particular pose of the moving object from a variety of view-points. This is illustrated in Figure 1. Thus each complete period in time yields one view of each pose assumed by the moving object, and by finding correspondences in frames across neighboring periods in time, one can apply standard techniques of multi-view geometry, with the caveat that in practice such periodicity is only approximate.

With this observation in hand, we address two key problems in this paper: (1) periodic motion detection, and (2) periodic motion segmentation. Our novelty in both of these cases is that we do not assume the object has been segmented, tracked and/or stabilized with respect to the camera. By relaxing these assumptions we are able to detect and segment periodic motion in complex scenes with motion parallax and non-rigid motion of the background.

The organization of this paper is as follows. In the remainder of this Section, we discuss related work. In Section 2, we outline the underlying geometry of 2D views of 3D periodic motion. We describe our approach to solving the correspondence problem using spatio-temporal interest points in Section 3. Our experimental results on the problems of periodic motion detection and segmentation are presented in Sections 4 and 5. We conclude in Section 6.

1.1. Related work

Periodic motion has been addressed in several works. Time-frequency analysis of image sequences was explored in [16, 12, 5, 21]. These methods find the period and classify periodic motion given the tracks of objects over time. While the methods in [16, 12, 5] rely on object segmentation, tracking and/or camera stabilization, the method in [21] does not assume segmentation and is more close to ours. It, however, has not been applied to scenes with moving camera and dynamic backgrounds. A drawback of all these methods is the assumption of periodicity in the image space which can be violated by projective transformations.

Multi-view constraints on periodic views of an object were explored in [20, 1]. In [20] cyclic motion under affine camera model was analyzed using pre-segmented trajectories of corresponding points. In [1] multi-view constraints were explicitly used to reconstruct periodic motion. Wide base-line matching of periodic views of an object was made under assumptions of a simple background and a given value of the period. In our method we borrow inspiration from [1] and use periodicity cue for detecting and segmenting periodic motion in complex scenes with non-rigid backgrounds and motion parallax.

Closely related to our work is the problem of sequence alignment of unsynchronized video from different view-

points in the case of stabilized cameras [4, 17, 24, 3] or cameras undergoing arbitrary motion [25, 28]. These approaches aim to discover the relative temporal offset and the spatial alignment between video streams by leveraging the epipolar constraint, which holds for corresponding points once the correct offset and the valid multi-view relation is found. Our work differs in two main respects. Firstly, due to deviations from perfect periodic motion in the moving target (e.g., a pedestrian), the video from one period to the next is only approximately the same, in contrast to the conventional problem in which each video depicts an identical event. This factor complicates both the correspondence problem and the satisfaction of the multi-view geometric relations. Secondly, while most other methods assume that both cameras are stationary or stabilized with respect to the background, here we do not make such an assumption and consider spatial and temporal alignment of video sequences obtained by two cameras with independent motion which is restricted to a constant translation. To our knowledge there is no prior work addressing spatial alignment of video for independently moving cameras.

2. Geometry of periodic motion

Periodic motion in space with the period p can be treated as a *collection* of rigid 3D configurations \mathcal{C} reoccurring with frequency $1/p$. Such a collection is generally influenced by the rigid geometric transformations originating from the gross motion of a periodic pattern. When observed by a camera, *periodic views* of \mathcal{C} are also influenced by the internal and external camera parameters that may be dynamic.

It follows that the views of a periodic object at times t and $t + p$ can be approximately treated as a stereo pair. Sub-sequences of periodic views

$$\begin{aligned} S_a &= \{s_t, s_{t+1}, \dots\}, & S_a &\subset S \\ S_b &= \{s_{t+p}, s_{t+p+1}, \dots\}, & S_b &\subset S \end{aligned} \quad (1)$$

of the same image sequence S can then be treated as two separate sequences of the same dynamic scene observed by one real and one virtual camera simultaneously. Hence, the problem of detecting periodic motion can be addressed in the framework of sequence-to-sequence alignment.

The general problem of sequence alignment [4, 3, 25] is a difficult one given that both cameras may move independently. In this work we make a simplifying assumption that both the motion of the camera and the gross motion of the periodic object is a constant translation. Such an assumption is often satisfied in practice since both the motion of observers and the motion of periodic objects such as animals and people often has a translative nature. We also currently assume static internal camera parameters.

To address the problem of matching periodic motion, in Section 2.1, we consider alignment of two image sequences

of a dynamic scene observed by two independently translating cameras. Specific constraints for aligning sequences with periodic motion are derived in Section 2.2.

2.1. Sequence alignment

Consider two cameras with the time-dependent relative translation $\mathbf{a} = \mathbf{a}_0 + t\mathbf{a}_1$, $\mathbf{a} \in \mathbb{R}^3$. By associating the origin with one of the cameras, the canonical form of camera matrices can be written as $P_a = K_a[A|\mathbf{a}]$ and $P_b = K_b[I|\mathbf{0}]$. Projection of a 3-d point using P_a, P_b results in image points with homogeneous coordinates $\mathbf{x}_a, \mathbf{x}_b$ satisfying the well-known epipolar constraint $\mathbf{x}_a^\top F \mathbf{x}_b = 0$ with the fundamental matrix $F = K_a^{-\top} A K_b^\top [K_b A^\top \mathbf{a}]_\times$ [7]. By expanding \mathbf{a} , it follows that the time-corresponding views obtained with P_a, P_b are related by the *dynamic fundamental matrix* that is a time-linear matrix function of the form

$$F(t) = K_a^{-\top} A K_b^\top [K_b A^\top (\mathbf{a}_0 + t\mathbf{a}_1)]_\times = F_0 + tF_1. \quad (2)$$

Given 3-d points on a plane $(\mathbf{n}, d)^2$, their projections satisfy a stronger relation given by $\mathbf{x}_a = H\mathbf{x}_b$ with the homography H decomposed as $H = K_a(A - \mathbf{a}\mathbf{n}^\top/d)K_b^{-1}$ [7]. For the case when the plane distance $d = d_0 + td_1$ is changing with respect to both cameras, we note that the homography as a homogeneous quantity is defined up to a scalar. Hence, we cancel the time-dependent term d in the denominator of H and obtain a time-linear expression³

$$H(t) = K_a(d_0A - \mathbf{a}_0\mathbf{n}^\top + t(d_1A - \mathbf{a}_1\mathbf{n}^\top))K_b^{-1}. \quad (3)$$

This proves the following proposition:

Proposition 2.1 *The homography induced by a plane and relating time-corresponding views of two constantly translating cameras is a time-linear matrix function of the form $H(t) = H_0 + tH_1$.*

The idea of time-linear homography is illustrated in Figure 2. Note that the alignment of point trajectories using static homography in Figure 2(d) fails while the time-dependent homography gives robust alignment disregarding noise in Figure 2(e).

2.2. Periodic motion alignment

To detect periodic motion in a sequence S , we wish to find a period p and a transformation $F(t)$ or $H(t)$ that enable alignment of subsequences S_a, S_b (1). By assuming

¹ Here K_a, K_b are 3×3 matrices with internal, static camera parameters, A is a static rotation matrix of the first camera and I is an identity.

² Here $\mathbf{n} \in \mathbb{R}^3$ is the plane normal and d is the distance between the plane and the origin in the 3-d space.

³ The time-dependent term d can be canceled since for any corresponding points $\mathbf{x}_a, \mathbf{x}_b$ at time t $H(t)$ is a constant matrix.

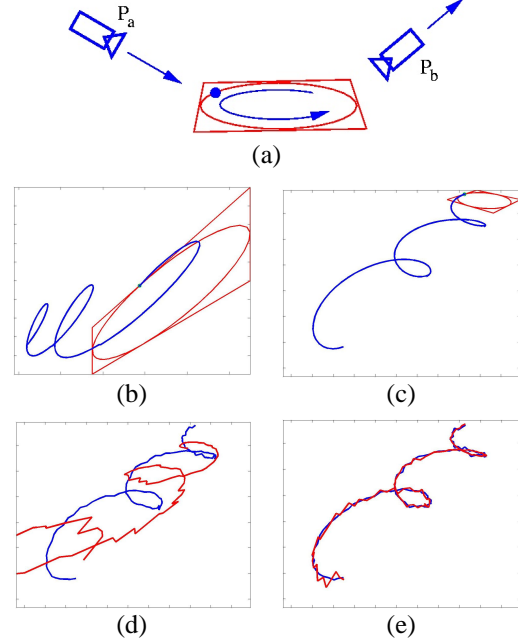


Figure 2. (a): A point (blue) moves on a planar circle (red) and is observed by two independently translating cameras P_a and P_b . (b)-(c): Trajectories of a point observed by the moving cameras P_a and P_b . Trajectories of an observed point depend on the camera motion. Given a few space-time point correspondences, the trajectories in both sequences can be aligned using an SVD estimation of the time-linear homography $H(t)$ of the proposition 2.1. Using this method, the alignment of the noise-corrupted trajectory in (b) to the noise-corrupted trajectory in (c) is shown in (e). If using constant homography, the alignment of trajectories fails as illustrated in (d).

the constant relative velocity \mathbf{v} between the periodic object and the camera, subsequences S_a and S_b correspond to one real and one virtual camera with identity rotation matrices I , the same internal camera parameters K and a constant relative offset $p\mathbf{v}$. By associating the origin with one of the cameras, the projection matrices become $P_a = K[I|p\mathbf{v}]$ and $P_b = K[I|0]$. From (2) it follows that the fundamental matrix $F(t)$ is constant for all periodic views of P_a, P_b and can be further simplified as

$$F_{per} = [Kp\mathbf{v}]_\times. \quad (4)$$

Since F_{per} is a homogeneous quantity, it has only two degrees of freedom and can be estimated from only two space-time point correspondences in S_a, S_b . This makes the estimation of F_{per} highly over-constrained as will be demonstrated in Section 4. Given the constancy of F and the fact that $\mathbf{x}^\top [Kp\mathbf{v}]_\times \mathbf{x} = 0$ for any \mathbf{x} , all corresponding periodic points will be on the same epipolar line as illustrated in Figure 3(b) and Figures 4(b),(d),(f).

For periodic motion on the plane, the time-linear homography in proposition 2.1 can be simplified as follows. By projecting the relative velocity \mathbf{v} on the plane normal \mathbf{n} , we note that the dynamic distance d between the plane (\mathbf{n}, d) and the origin can be expressed as $d = d_0 + td_1 = d_0 - t\mathbf{n}^\top \mathbf{v}$. Using projection matrices $P_a = K[I|p\mathbf{v}]$, $P_b = K[I|0]$ the expression in (3) can be re-written as

$$H_{per}(t) = I - Kp\mathbf{v}\mathbf{n}^\top K^{-1}/d_0 - t\mathbf{n}^\top \mathbf{v}I/d_0 \quad (5)$$

with only one time-dependent parameter in the last term. Note, that if camera is translating parallel to the plane, i.e., $\mathbf{n}^\top \mathbf{v} = 0$, the time-dependent term vanishes and the homography is reduced to a static one with the form $H_{per} = I - Kp\mathbf{v}\mathbf{n}^\top K^{-1}/d_0$. If we further assume the plane (\mathbf{n}, d) is perpendicular to the camera, i.e., $\mathbf{n} = (0, 0, 1)^\top$, $H_{per}(t)$ in (5) is reduced to a matrix function of the form

$$\tilde{H}_{per} = \begin{pmatrix} 1 & h_1 \\ & 1 & h_2 \\ & & h_3 \end{pmatrix} + t \begin{pmatrix} h_4 & & \\ & h_4 & \\ & & h_4 \end{pmatrix} \quad (6)$$

with four unknowns h_1, \dots, h_4 that can efficiently be estimated using two space-time point correspondences only. We use (6) for segmenting periodic motion in Section 5.

We note that the alignment of periodic views according to F_{per} (4) and $H_{per}(t)$ (5) is a necessary but not a sufficient condition for periodic motion in video sequences. Indeed, subsequences S_a, S_b (1) obtained for any constantly translating rigid object could be aligned using some F_{per} (4) and any value of period p . Sufficient conditions for the presence of periodic motion will be discussed in the next Section.

3. Space-time image features

To estimate the dynamic $F(t)$ and $H(t)$ matrices of Section 2, we can take advantage of time linearity and apply SVD-based methods that are commonly used for estimating static F, H from two views [7]. Unlike the static case, however, estimation of $F(t), H(t)$ requires correspondences of space-time points in two image sequences. We find these correspondences by *directly* matching points in space-time. The advantage of this method is that it is independent of tracking and segmentation used in previous methods for sequence alignment [4, 17].

To estimate corresponding points in two sequences, we consider space-time interest points with significant variation of local motion and shape. Such points or *Local Space-Time Features* (LSTF) can be detected by maximizing the local variation of the image function over space and time [9]. Given the distinctive spatio-temporal properties of such points, correspondence can be estimated from the similarity of their local spatio-temporal neighborhoods. Due to the space limitations we refer the reader to [9, 10] for the

details of this method. Figure 3 illustrates LSTF points detected for a sequence containing a jogging person. Close similarity of spatio-temporal neighborhoods of matching periodic points can be confirmed in Figure 3(c).

The detector in [9] delivers a rather sparse set of points that is sufficient for the detection of periodic motion described in Section 4. Segmentation of periodic motion in Section 5, however, requires a denser set of points that enable more accurate alignment of periodic views. To detect such points, we relax the assumption of local extrema of the image variation over time and detect *Weak Local Space-Time Features* (WLSTF) by applying a standard static interest point detector [14] restricted to the regions of non-constant motion [11]. For each detected point we then compute a local spatio-temporal descriptor according to [10]. Examples of WLSTF points detected for pairs of periodic frames are illustrated in Figures 5(a)-(b) and Figure 1.

The advantage of LSTF and WLSTF points in the context of this work is twofold. Firstly, they contain distinctive motion information and greatly disambiguate the search for space-time correspondences required for estimating F_{per} (4) and $H_{per}(t)$ (5). Secondly, LSTF and WLSTF are insensitive to constant translation in images and disable periodic matching of sequences with trivial translative motion. Hence, the alignment of such points by F_{per} and $H_{per}(t)$ provides a sufficient condition for the presence of periodic change of appearance in image sequences.

4. Periodic motion detection

We formulate a RANSAC procedure for simultaneous estimation of (i) the period of periodic motion p , (ii) the fundamental matrix F_{per} (4) and (iii) the space-time points of periodic motion. We begin with the detection of pairs of corresponding LSTF points based on the similarity of their local descriptors. In each RANSAC iteration we randomly select two pairs of matching LSTF points with similar temporal delay Δt and estimate fundamental matrix F_{per} (4) using a standard SVD approach [7]. We then score the estimates of F_{per} and Δt by the number of point pairs $(\mathbf{x}_1, \mathbf{x}_2)$ that (i) have similar descriptors, (ii) satisfy the epipolar constraint $\mathbf{x}_2^\top F_{per} \mathbf{x}_1 \approx 0$ and (iii) have temporal offset $|t_2 - t_1| \approx \Delta t$. A large number of consistent point pairs indicates correctness of the estimated motion period $p = \Delta t$ and the fundamental matrix F_{per} . The best estimates p, F_{per} are chosen after a fixed number of iterations.

Application of this method to the detection of periodic motion of people is illustrated in Figures 3-4. For a jogging person, in Figure 3(b) we show a few groups of periodic LSTF points found by the algorithm. Points within each group have similar local neighborhoods (Figure 3(c)) and correspond to the similar body points of the person at period-separated moments of time. Each group of points

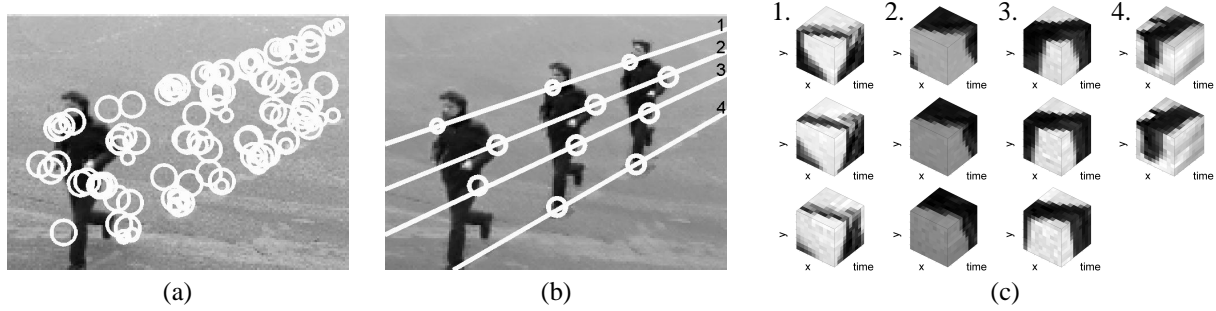


Figure 3. Detection of periodic motion. (a): Circles indicate spatial position of LSTF points detected for a sequence with a jogging person. The features are time-projected onto one frame of a sequence. (b): Result for periodic motion detection. Four selected groups of corresponding periodic points are connected by epipolar lines 1, ..., 4. (c): Space-time image neighborhoods corresponding to the points in (b). Note the similarity of neighborhoods corresponding to the features on the same epipolar line.

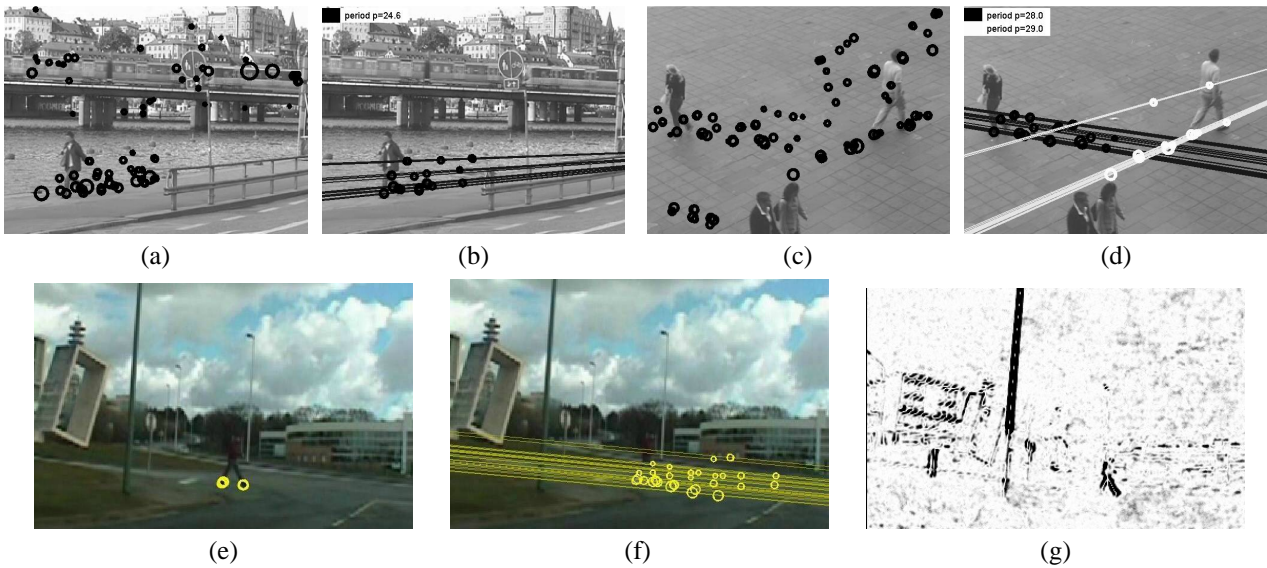


Figure 4. Detection of periodic motion. (a),(c),(e): Detected LSTF points. (b),(d),(f): Results for periodic motion detection. Period-corresponding features of periodic motion are located on similar epipolar lines. LSTF points originating from non-periodic motion in the background are suppressed. (g): Results for dominant motion estimation for a sequence with a moving camera in (e) using quadratic motion model [15]. The outlier regions of dominant motion (black) correspond to a moving person as well as to the motion of other static structures in the background due to motion parallax.

is located on the same epipolar which verifies the correct estimation of F_{per} .

Periodic motion detection for more complex scenes is shown in Figure 4. In Figures 4(a)-(b) we observe that the method converged to a correct estimate of inlier features corresponding to the periodic motion of a walking person while all outliers originating from motion in the background (e.g. from the train) were correctly rejected. Epipolar lines correctly indicate the direction of motion of a person, while the estimated value of the period $p = 24.6$ has been manually verified as correct.

Using RANSAC we can easily detect multiple instances of periodic motion in the same sequence. For this purpose we ignore periodic features found in the previous runs of the algorithm and check the presence of additional instances

of periodic motion within the remaining features.⁴ In Figure 4(c)-(d) this method separates periodic motion of two persons with different motion directions and slightly different periods of the gait.

While the camera in previous examples was stationary, in the example of Figures 4(e)-(f) the camera was inside a car driving toward a walking person. As illustrates in Figure 4(f), our algorithm correctly detected a person and her motion direction. The result of dominant motion estimation in this example (see Figure 4(g)) indicates the potential problem for motion-based segmentation methods due to motion parallax. Unlike previous methods for periodic

⁴ Using this method, multiple instances of periodic motion can be separated provided their difference in the motion direction given by F_{per} , period value p or both.

motion detection [5, 16], our approach does not rely on segmentation. This property can be valuable in applications such as automatic car navigation and driving assistance.

Regarding other applications, the presented method could be combined with the recognition of human actions and used to extend [19] with the recognition of multiple periodic actions performed simultaneously in scenes with possibly complex background motion.

5. Periodic motion segmentation

We consider the problem of segmenting periodic motion in image sequences. Motion is commonly recognized as a strong cue for separating objects from the background (e.g. using camera stabilization followed by the estimation of residual motion). Complex scenes with motion parallax and multiple independent motions in the background, however, often make this approach problematic.

Given the difficulty, it is still possible to use motion as an informative cue for segmentation if the *type* of motion of particular objects is different from the *type* of motion in the rest of the scene. Here we use this idea and aim to separate objects with a specific periodic motion from complex backgrounds. To simplify the problem, we assume that the periodic motion is on the plane. We further assume that the object, observer or both are translating constantly with respect to each other and use the time-linear homography $\tilde{H}_{per}(t)$ (6) to align periodic views of the object. For this purpose we detect and match WLSTF points introduced in Section 3 and apply RANSAC procedure to estimate $\tilde{H}_{per}(t)$ and p simultaneously.

As an example sequence we considered a dynamic scene from the movie “Run Lola Run” illustrated in Figure 5. The sequence has substantial motion parallax since the camera is translating backward. The person is approaching the camera with periodic running motion, hence, her periodic views are affected by translation and linear scale changes; see Figures 5(a)-(b). The detected WLSTF points cover the body of the person rather densely and allow for accurate estimation of the homography between corresponding views. Feature matches that are consistent with the estimated homography are illustrated in Figure 5(c). The result of alignment in Figure 5(d) visually confirms the correct estimation of the homography and of the period of motion despite only partial and approximate correspondence of periodic views.

Using period-aligned views of the object we obtain an *object-centered camera stabilization* (as opposed to background camera stabilization used in other methods). With such a representation, the segmentation of periodic motion can be approached by evaluating the quality of alignment in different image regions. Evaluating alignment in terms of image values directly, however, has shown to be problematic due to many homogeneous image regions. As an

alternative approach, we note that two object-aligned periodic frames will usually have low disparity for the object and large disparity for the background. We use this observation and apply stereo-matching to estimate the disparity for each point of the sequence. Using graph-cut implementation of stereo-matching [8] we estimated disparity for time-separated views with the time delays $\Delta t = p, 2p, 3p$. The average disparity map obtained with such a method for a frame in Figure 5(b) is illustrated in Figure 5(e). Finally, using disparity estimates as a likelihood measure, we applied a binary graph-cut segmentation [2] to obtain the segmentation mask of the periodic object illustrated in Figure 5(f).

We also considered a scene with non-rigid background motion caused by the motion of cars (see Figure 6). For comparison, we first show results of dominant motion estimation in Figure 6(b). As can be seen, this approach is not able to separate the motion of the person from the motion of cars in the background. Using our method, the periodic motion of a person in this sequence was successfully segmented as illustrated in Figure 6(d).

Although the results of segmentation in this example are not perfect, the resulting regions could be useful for example to initiate or to re-initialize a tracker or to recognize the class of periodic motion in a complex scene. We believe that more accurate segmentation can be obtained by applying additional methods such as in [18].

6. Conclusion

We presented a method for detecting and segmenting periodic motion in video sequences. The particular advantage of the proposed method is that it can be applied to complex scenes, but does not rely on camera stabilization, on segmentation nor on tracking. Our solution is formulated in the framework of sequence alignment. In this respect we (i) investigated a general case of sequence-to-sequence alignment for independently translating cameras and (ii) showed how this approach applies to the detection and segmentation of periodic motion in complex video sequences with motion parallax and non-rigid motion of the background.

One limitation of our approach is the assumption of constant translation. To address the general class of motion while preserving linear estimation of $F(t)$ and $H(t)$, one could consider piecewise linear or polynomial approximations of $F(t), H(t)$. Another direction for future investigation concerns alignment of non-periodic motion in different video sequences using the framework of point-wise sequence alignment developed here.

7. Acknowledgments

J. Wills and S. Belongie are supported by NSF-CAREER #0448615, DOE/LLNL contract no. W-7405-ENG-48 (sub-contracts B542001 and B547328), and the Sloan Fellowship.

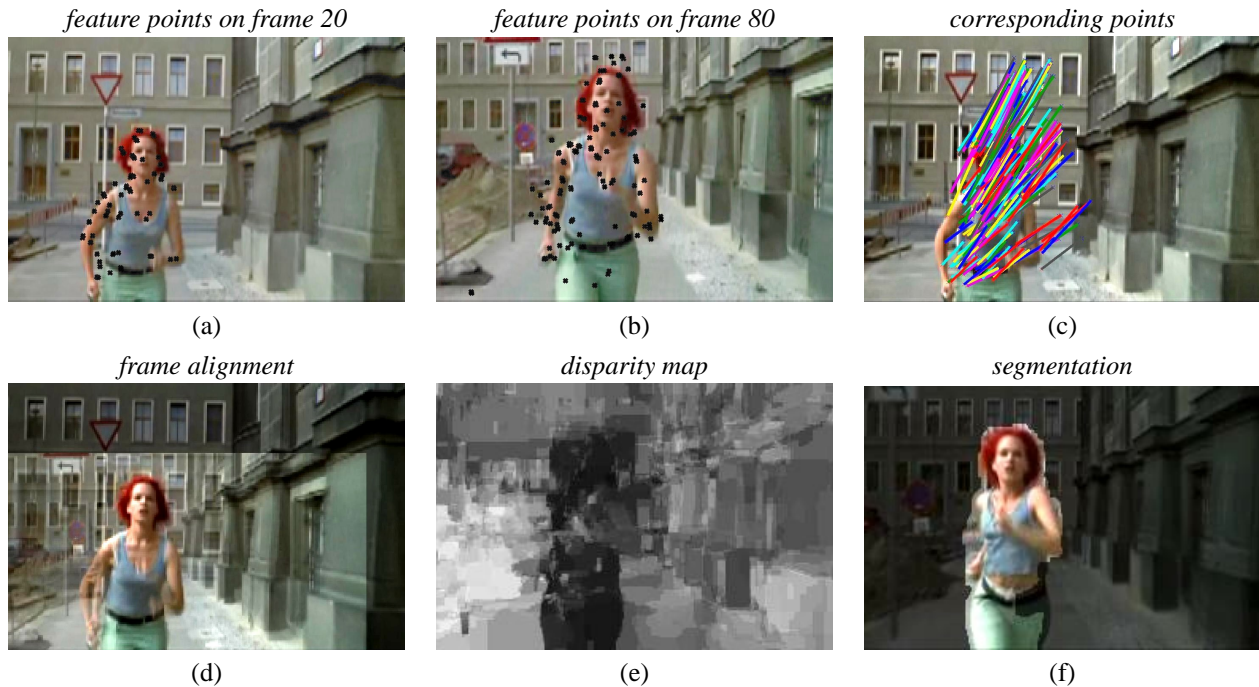


Figure 5. Matching, alignment and segmentation of periodic motion. (a)-(b): Two periodic views selected from the same video sequence and separated in time by three values of a period. Black dots show locations of detected WLSTF points for each frame. (c): Matches of WLSTF points in (a)-(b) that are consistent with the estimated dynamic homography \hat{H}_{per} . (d): Result of alignment of the frame (b) to the frame (a). The sum of the frame (a) and the transformed frame (b) is shown. (e): Estimated disparity map with low values corresponding to regions with good alignment of periodic frames. (f): Final result of segmentation.

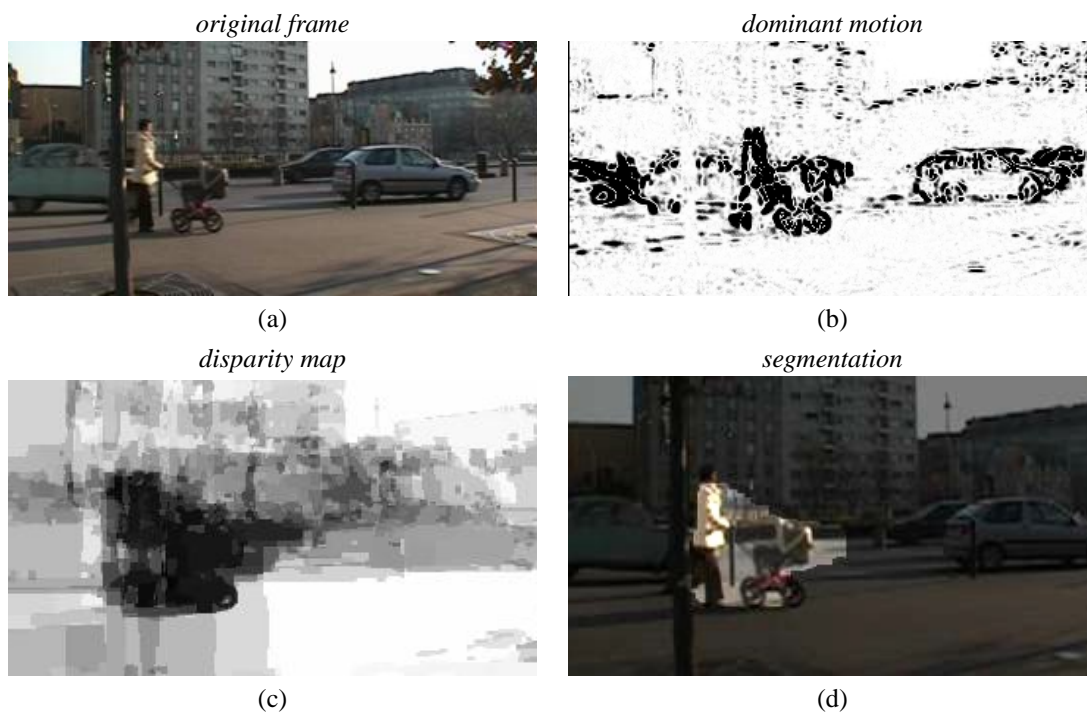


Figure 6. Segmentation of periodic motion for a scene with non-rigid background. (a): Original frame. (b): Dominant motion estimation using quadratic motion model [15]. (c): Estimated disparity for the frame in (a). (d): Final result of segmentation.

References

- [1] S. Belongie and J. Wills. Structure from periodic motion. In *First International Workshop on Spatial Coherence for Visual Motion Analysis*, Prague, Czech Republic, May 2004.
- [2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE-PAMI*, 26(9):1124–1137, September 2004.
- [3] R.L. Carceroni, F.L.C. Padua, G.A.M.R. Santos, and K.N. Kutulakos. Linear sequence-to-sequence alignment. In *Proc. CVPR*, pages I:746–753, 2004.
- [4] Y. Caspi and M. Irani. Spatio-temporal alignment of sequences. *IEEE-PAMI*, 24(11):1409–1424, November 2002.
- [5] R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE-PAMI*, 22(8):781–796, August 2000.
- [6] T. Darrell, S. Sclaroff, and A. Pentland. Segmentation by minimal description. In *Proc. ICCV*, pages 112–116, 1990.
- [7] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [8] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. ECCV*, page III: 82 ff., 2002.
- [9] I. Laptev and T. Lindeberg. Space-time interest points. In *Proc. ICCV*, pages 432–439, 2003.
- [10] I. Laptev and T. Lindeberg. Local descriptors for spatio-temporal recognition. In *First International Workshop on Spatial Coherence for Visual Motion Analysis*, 2004.
- [11] T. Lindeberg, A. Akbarzadeh, and I. Laptev. Galilean-corrected spatio-temporal interest operators. In *Proc. ICPR*, 2004.
- [12] H. Liu, T.H. Hong, M. Herman, and R. Chellappa. Motion-model-based boundary extraction and a real-time implementation. *CVIU*, 70(1):87–100, April 1998.
- [13] Y. Liu, R.T. Collins, and Y. Tsin. Gait sequence analysis using Frieze patterns. In *Proc. ECCV*, 2002.
- [14] D.G. Lowe. Object recognition from local scale-invariant features. In *Proc. ICCV*, pages 1150–1157, 1999.
- [15] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *JVCIR*, 6(4):348–365, December 1995.
- [16] R. Polana and R.C. Nelson. Detection and recognition of periodic, nonrigid motion. *IJCV*, 23(3):261–282, June 1997.
- [17] C. Rao, A. Gritai, M. Shah, and T. Syeda-Mahmood. View-invariant alignment and matching of video sequences. In *Proc. ICCV*, pages 939–945, 2003.
- [18] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. In *Proc. SIGGRAPH*, 2004.
- [19] C. Schödl, I. Laptev, and B. Caputo. Recognizing human actions: a local SVM approach. In *Proc. ICPR*, 2004.
- [20] S.M. Seitz and C.R. Dyer. View invariant analysis of cyclic motion. *IJCV*, 25(3):231–251, December 1997.
- [21] A. Thangali and S. Sclaroff. Periodic motion detection and estimation via space-time sampling. In *IEEE Workshop on Motion and Video Computing*, pages II:176–182, 2005.
- [22] P. H. S. Torr. Geometric motion segmentation and model selection. In J. Lasenby, A. Zisserman, R. Cipolla, and H. Longuet-Higgins, editors, *Philosophical Transactions of the Royal Society A*, pages 1321–1340. Roy Soc, 1998.
- [23] L. Torresani, C. Bregler, and A. Hertzmann. Learning non-rigid 3D shape from 2D motion. In *Proc. NIPS*, 2003.
- [24] P.A. Tresadern and I. Reid. Synchronizing image sequences of non-rigid objects. In *Proc. BMVC*, 2003.
- [25] T. Tuytelaars and L.J. Van Gool. Synchronizing video sequences. In *Proc. CVPR*, pages I:762–768, 2004.
- [26] R. Vidal and Y. Ma. A unified algebraic approach to 2-d and 3-d motion segmentation. In *Proc. ECCV*, pages I:1–15, 2004.
- [27] J. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proc. CVPR*, pages 361–366, 1993.
- [28] A. Whitehead, R. Laganiere, and P. Bose. Temporal synchronization of video sequences in theory and in practice. In *IEEE Workshop on Motion and Video Computing*, pages II:132–137, 2005.
- [29] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. In *Proc. ECCV*, pages IV:573–587, 2004.