Computational Attention for Defect Localisation

Matei Mancas[1], Devrim Unay[1], Bernard Gosselin[1], Benoît Macq[2]

[1] Faculty of Engineering, Mons (FPMs), TCTS Lab
31, Bd. Dolez, 7000, Mons, Belgium
{matei.mancas, bernard.gosselin}@fpms.ac.be
devrim.unay@philips.com
[2] Catholic University of Louvain (UcL), TELE Lab
2, Place du Levant, 1348 Louvain-la-Neuve, Belgium
Macq@tele.ucl.ac.be

Abstract. This article deals with a biologically-motivated three-level computational attention model architecture based on the rarity and the information theory framework. It mainly focuses on a low-level step and its application in pre-attentive defect localisation for apple quality grading and tumour localisation for medical images.

Keywords: computational attention, saliency, rarity, apple defect, tumour

1 Introduction

The human visual system (HVS) is a topic of increasing importance in computer vision research since Hubel's work [1] and the comprehension of the basics of biological vision. Mimicking some of the processes done by our visual system may help to improve the existing computer vision systems.

In this article, we describe a biologically-motivated three-level visual attention and we apply the first low-level step in defect localisation on apples. An extension on tumour localisation in head and neck will also be done. An important result is that the use of an "atlas" (set of healthy -not defected- images) can highly improve the results. This atlas models the knowledge already acquired about the analysed images.

The general idea of our visual attention model is described in the next section. Part three provides the description of the defect localisation mechanism. The final section will conclude the work and discuss our approach.

2 Visual Attention (VA)

In this article, we shall only address the low-level pre-attentive processes of visual attention. Pre-attentive visual attention is reflex-based and it occurs faster than an eye saccade (eye movement) corresponding to 200 milliseconds for humans. The pre-attentive interest areas detection is a "parallel" fast process by opposition with the saccade-based image analysis which is a "serial" and slower process [2].

2.1 Biological background

The Superior Colliculus (SC) is the brain structure which directly communicates with the eye motor command in charge of eye's orientation. One of its tasks is to direct the eyes onto the "important" areas of the surrounding space: studying the SC afferent paths can provide important clues about visual attention.

There are two afferent pathways for the SC, one direct path from the retina, and another indirect path crossing the Lateral Geniculate Nucleus (LGN) and the primary cortex area (V1) before coming back to the SC.

Studies on afferent SC pathways [3] showed that the direct path from the retina is responsible for spatial (W cells) and temporal (Y cells) analysis and the indirect pathway is mainly responsible for spatial and motion direction and colour analysis.

2.2 Attention modelling

Many methods may be found in the literature about visual attention and saliency. They are mainly divided into two categories. The first one [4][5] uses locally computed salient features and the second one [6] [7] [8] [9] [10] uses similarity and comparisons all over the image in a global processing.

Our definition will be based on the **rarity** concept which is necessarily a global concept integrating the local processing of different cells and which could be situated within the second saliency computation method category. We noticed that our vision is not attracted by specific features, but by features which are in minority in an image. Based on a global rarity idea, we propose a three-level approach of visual attention which is divided into three parts: a low-level approach which is exclusively pre-attentive, a high-level one which is exclusively attentive and a medium-level approach which can be either pre-attentive or attentive depending on the number of medium-level features. The low-level approach could be directly carried inside the SC where only luminance and motion cells are available.

2.3 Rarity quantification

A pre-attentive analysis is achieved by humans in less than 200 milliseconds, so the pre-attentive model should also be very fast. The fastest and most basic operation is to count similar areas in the image, hence to use the histogram. Within the context of information theory, this approach based on the histogram is close to the so-called self-information. Let us note m_i a message containing an amount of information. This message is part of a message set M. A message self-information $I(m_i)$ is defined as:

$$I(m_i) = -\log(p(m_i)) \quad \textbf{(1)}$$

where $p(m_i)$ is the probability that a message m_i is chosen from all possible choices in the message set M or the occurrence likelihood. We obtain an attention map by replacing each message m_i by its corresponding self-information $I(m_i)$. We define $p(m_i)$ as a two-terms combination:

$$p(m_i) = \left(\frac{H(m_i)}{Card(M)} \right) \times \left(1 - \frac{\sum_{j=1}^{Card(M)} \left| m_i - m_j \right|}{Card(M) \times Max(M)} \right) \quad \textbf{(2)}$$

The first term is the direct use of the histogram: *H(m$_i$)* is the value of the histogram *H* for message m_i and *Card(M)* is the cardinality of the message set M.

The second term quantifies the distance between a message and all the others or its global contrast. If a message is very different from all the others, this term will be lower and the message attention will be higher.

3 Low-level Spatial Visual Attention

In an image we can consider in a first approximation that a message m_i is the grey-level of a pixel at a given space location and the message set *M* is the entire image. Nevertheless, comparing only isolated pixels is not efficient. In order to introduce a spatial relationship, areas surrounding each pixel should be considered.

Stanford [11] showed that the W-cells which are responsible of the spatial analysis inside the SC may be separated into two classes: the tonic W-cells (sustained response all over the stimulus) and the phasic W-cells (high responses at stimulus variations).

Our approach uses the mean and the variance of a pixel neighbourhood in order to describe its statistics and to model the action of tonic and phasic W-cells.

We compute the local mean and variance on a 3x3 sliding window as our experience showed that this parameter is not of primary importance. To find similar pixel neighbourhoods we count the neighbourhoods which have the same mean and variance (first term of **Eq. 2**). Than we compute the distance between the pixel neighbourhood mean and the others to get the second term of **Eq. 2**.

Contours and statistically smaller areas get higher attention scores on the VA map. If we consider only local computations as, for example, the local standard deviation or the local entropy, contours are also highlighted but the textured areas have a too high score. In our method, more regular a texture is, less surprising it is and less important the attention score will be [12]. Achieved observations prove the importance of a global integration of the local processing made by the cells. Rarity or surprise, which obviously attracts our attention, cannot be computed only locally.

4 Defect Localisation

4.1 The fruit grading problem

Automatic quality inspection of fresh fruits by machine vision is a challenge not only due to their largely varying physical appearances, but also because we need to decrease the cost, time and error of inspection introduced by human experts. Apple fruits have numerous kinds of defects and highly varying skin colour that complicate their inspection by machine vision.

Jonagold apples have bi-coloured skin causing problems at inspection due to the colour transition areas. Database used in this work is composed of 280 healthy and 246 defected Jonagold apples that are injured by various natural (russets, bruises, rots, flesh damages…) and artificially created (some bruises) defects. Image acquisition is achieved by a multispectral system consisting of a high-resolution monochrome camera and four band-pass filters centred at 450, 500, 750 and 800nm. Defected skin on the images is manually segmented by an expert and these manual segmentations are used as ground truth in this work.

4.2 Low-level attention and Jonagold Apples

In our tests we used 750 and 800 nm apple images because most of the defects are highly visible on these two modalities. A pre-processing step is needed in order to eliminate areas where apple defects are never located (as the air surrounding the apple…) and areas where one have many shadow and uneven illumination (as apple borders). These shadows due to apple curvature may introduce a lot of confusion between real defects and healthy skin which have the same grey-level characteristics. We can see the result of the pre-processing step in **Fig. 1**, second column where the apple is eroded and any variation in its background is suppressed.

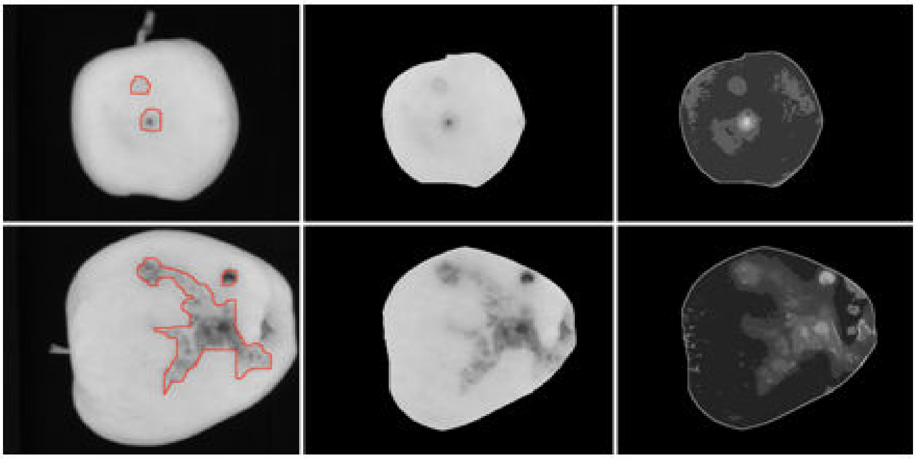

Fig. 1. From left to right: initial image with segmented defect, pre-processing step, our VA map

The third column shows two examples of results obtained after applying our low-level attention map. As we saw in part 3, apple main contours have a high attention score as edges are rare in this image. We also can see that the other areas, which are well highlighted, are quite well correlated with the defect segmentations in column one. Here, rare areas are anomalous, hence defected.

Nevertheless, some regions that are neither contours nor defected have also high attention scores. These “false positives” are mainly due to illumination artefacts or to the presence of stem or calyx regions which are quite similar to defects.

4.3 Building and using an apple “Atlas”

In the previous section, pixel neighbourhood rarity was computed on the initial images (**Fig. 1**, first column). As already emphasized, defect areas were highlighted but a major problem is in illumination and shadows which are also rare within the initial image but which occurs frequently in this kind of images. In order to avoid most of these problems we use an atlas which is simply a volume containing the test image concatenated with a set of healthy images.

If illuminations and shadows often occur on healthy images, we will find this kind of information in the atlas, thus even if it is rare within the initial image it will be less rare if the entire atlas is taken into account. On the contrary, the defected skin will be even rarest as it never occurs within the atlas, but only on the test image.

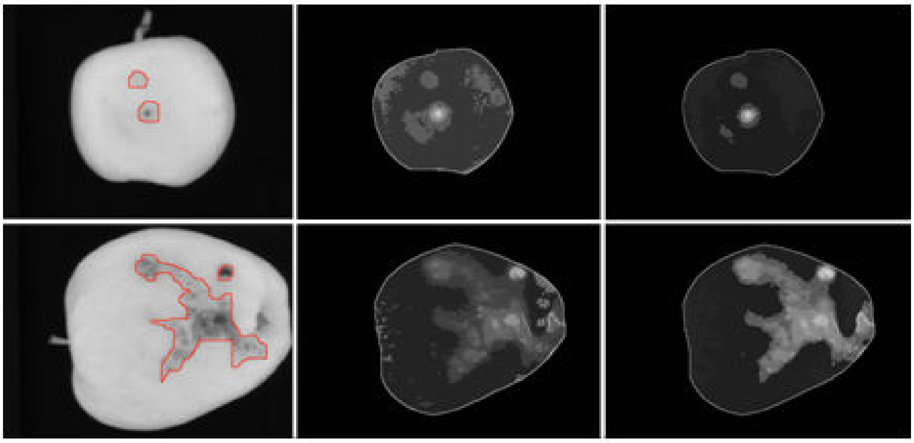

Fig. 2. From left to right: initial image with segmented defects, VA map using initial image only, VA map using initial image and an atlas

The second column of **Fig. 2** shows the results obtained using only the test image. The third column shows the results obtained by adding the test image to the atlas which was a set of twenty images of healthy apples from the same modality. We can see the two improvements already announced: first, defected skin has a higher attention score and second, most of the noise due to illumination is suppressed.

The atlas lets rarity to be computed not only on the presently seen image but also on previously seen images. It acts like a memory providing a priori knowledge and the possibility to learn from previous experiences.

4.4 Results and discussion

In order to evaluate efficiency of VA map with/without the atlas on apple defects, we compared densities of the attention values from defected and healthy areas of the skin using the ground truths. Here, comparisons are based on two different formulations where the first uses the means of the densities, while the second employs both the mean and the standard deviation as in **Eq. 3** and **4**, respectively.

$$\Delta_{mean} = \sum_{fruit} \mu_{defected} - \mu_{healthy} \qquad \mathbf{(3)}$$

$$\Delta_{\mathrm{min\,max}} = \sum_{fruit} \left(\mu_{defected} - \sigma_{defected}\right) - \left(\mu_{healthy} + \sigma_{healthy}\right) \qquad (4)$$

A typical comparison in graphical representation is displayed in **Fig. 3** for limb rub and bruise types of defect. Our visual analysis revealed that Atlas-based VA maps provided a more distinct separation between healthy and defected skins for most of the defect types (e.g. limb rub, frost damage, hail damage, scald…), whereas for others it was difficult to notice that separation (e.g. bruise, russet and flesh damage). Hence a quantitative analysis was necessary.

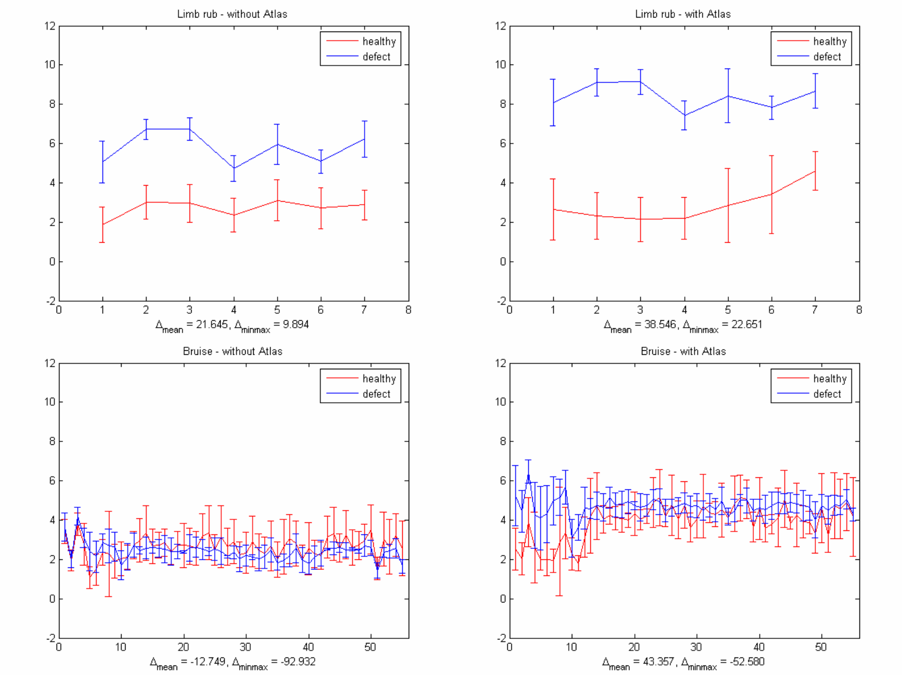

Fig. 3. Comparison of with/without atlas visual attention based inspection. Top row refers to limb rub defect, while the bottom one is for bruised fruits. Left column displays results without Atlas, whereas those on the right are with Atlas.

The following equation (**Eq. 5**) provides a formulation of the improvement between the results of VA with and without the Atlas. **Table 1** displays numerical results of VA based apple inspection for each defect type as well as the whole database.

$$improvement = \frac{\Delta_{Atlas} - \Delta_{noAtlas}}{\left|\Delta_{noAtlas}\right|} \qquad \mathbf{(5)}$$

These numerical results reveal significant improvements (in favour of Atlas usage) for most of the defect types with the exception of flesh damage and russet defects when evaluation is based on *minmax*.

		without Atlas		with Atlas		Improvement	
defect type	#fruits	Δ_{mean}	Δ_{minmax}	Δ_{mean}	Δ_{minmax}	mean	minmax
Bruise	55	-12.75	-92.93	43.36	-52.58	4.4	0.4
Flesh damage	24	25.43	-16.88	37.69	-15.22	0.5	0.1
Frost damage	11	24.70	2.30	48.11	11.96	1.0	4.2
Hail damage	16	35.96	3.97	53.81	9.00	0.5	1.3
Hail damage perf	31	114.46	56.25	184.18	102.66	0.6	0.8
Limb rub	7	21.64	9.89	38.55	22.65	0.8	1.3
Other	20	42.88	5.82	76.76	25.17	0.8	3.3
Rot	23	43.26	-5.90	84.83	12.94	1.0	3.2
Russet	42	23.09	-51.44	50.37	-51.79	1.2	0.0
Scald	17	29.68	-4.16	50.54	4.46	0.7	2.1
All database	246	348.36	-93.09	668.20	69.24	0.9	1.7

Table 1. Results of visual attention based apple inspection.

Our detailed analysis on individual results revealed that errors of the VA based inspection were mostly due to two reasons. Firstly, stem and calyx areas, which are natural parts of apples, were regarded as rare events by VA. These false alarms can be removed using a separate technique dedicated to their identification based on support vector machines [13]. The second reason is that some defects are difficult to detect by VA, because they are either very complex (e.g. russet) or not clearly visible in the selected filter image (e.g. flesh damage and some bruises). Furthermore, some of the erroneous defects provide better visibility at 450 or 500 nm images. Therefore, a more robust inspection system should combine the results of VA from several filter images, which is one of our future works.

4.5 From defected apples to head and neck tumours

Defect localisation is one of the main challenges of machine vision and its applications are widespread. Fruit quality inspection or defects in industrial manufacturing are obvious applications, but defect localisation may also find applications to pathology localisation on medical images. Abnormalities are also rare therefore, the same ideas may be used in pathologies localisation (as tumours).

The test set images are head and neck computed tomography (CT scans) and the purpose is a coarse localisation of possibly pathological areas. CT scan images are very noisy and it appears that using the neighbourhood variance bring more confusion to the final images. Hence, we used only the mean of pixel neighbourhood in order to compute the VA maps.

As for apples, in a first pre-processing step we eliminated the regions in the image where tumours could not be located: the air and bones. On **Fig. 4**, first row, one has four different tumours with segmented tumour active areas. The second row presents the direct single-image VA maps. The higher responses correspond to contours as in apples case. High responses correspond to tumours or other rare grey-levels within the image. If there are blood vessels for example, they are very few and compact so they will be very well highlighted.

On **Fig. 4**, third row we can see results using an atlas of 39 healthy head and neck CT scan images. As in the case of apples, the grey-level which is mainly responsible of the tumour has smaller occurrence within the atlas, so it has a highest response in most of the cases. Some grey-levels which are quite common within the atlas will have lower VA scores. If atlas knowledge is used, the highly salient areas which will be first inspected are smaller. This result is consistent with the tests done in [14] where an experienced specialist (atlas-based) inspected smaller and more precise areas than a non-specialist (single image-based).

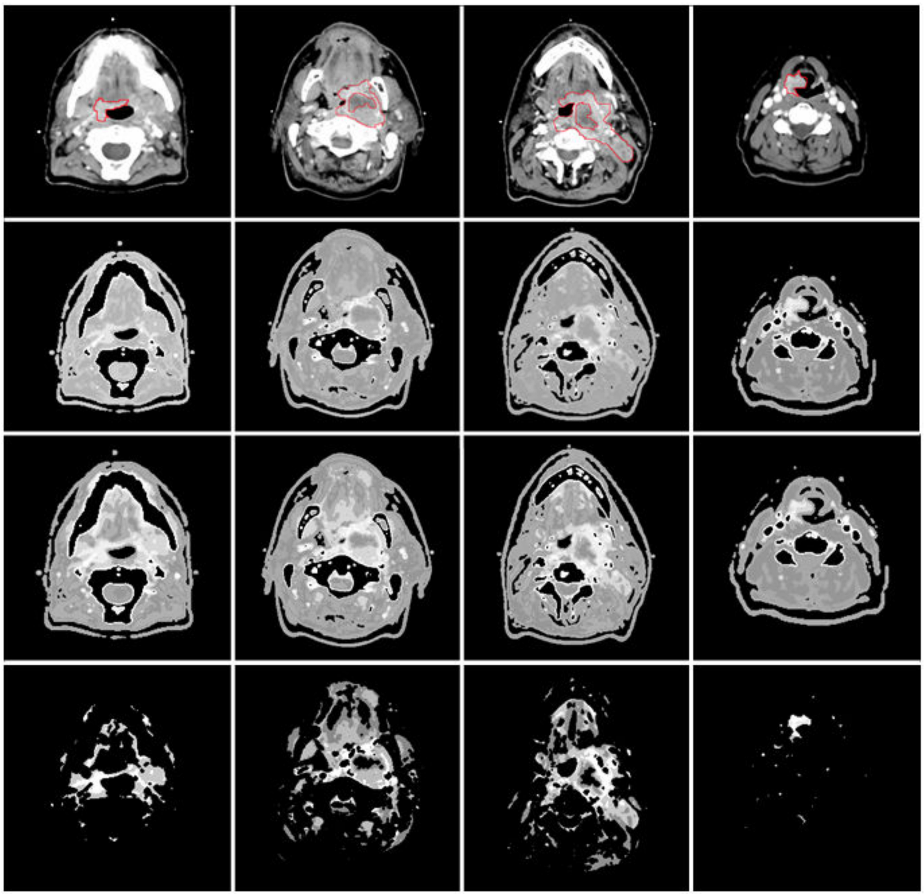

Fig. 4. From top to down: first row: Four CT scan images with segmented active tumour regions, second row: single-image VA maps, third row: atlas-based VA maps, fourth row: atlas-based VA maps using attention variation information between rows two and three

In order to find these areas which would be inspected by experienced radiologists, we may compare the atlas-based and the single image results. If there is no variation between the attention score of the single-image or atlas-based methods, this means that there is no change in the grey-level occurrence between healthy and pathological images, hence concerned areas should be healthy.

In the fourth row of **Fig. 4** we suppressed from the atlas-based VA maps the regions which have no variation and negative attention variation (VA_{Atlas} - $VA_{Single-image} \geq 0$) for the first three columns and the null and negative attention variation for the fourth column (VA_{Atlas} - $VA_{Single-image} \leq 0$). The decision is based on keeping the region with the higher attention density.

Fig. 4, fourth row shows that the remaining areas contain the tumours, and moreover, some other healthy regions with high attention score as blood vessels were removed. Even if they are rare in the test image but also in the atlas, there is no attention variation for blood vessels, therefore they are eliminated. This is very interesting as blood vessels are healthy areas but they have very high VA scores.

Our study showed the feasibility of a fully automatic pathology areas localisation in medical images. Nevertheless tumours can either be localised by rare grey-levels or by rare shapes. Our low-level approach described here only handles rare grey-levels. This method works on organs where pathologies are mainly visible because they have different grey-levels from the rest of the image or from the rest of the already seen images. To detect rare shapes (asymmetry, irregular shapes) when these features are important in tumour localisation, one should use a complimentary high-level approach.

5 Discussion and Conclusion

In this article we described a low-level visual attention approach based on grey-level rarity within an image or an atlas which is a set of images. This low-level method was applied to defect localisation in two different domains: automatic fruit grading with apple defects and medical imaging with automatic localisation of tumours inside CT scan images.

The low-level approach provided interesting results for “pre-attentive” defect localisation which means that defects can be visually located using grey-level rarity.

Concerning apple defect localisation, our hope is to build a system which can automatically provide defected skin to a feature extractor in order to automatically train a set of supervised classifiers. These classifiers are already able to provide good results in fruit grading [15]. At each season defects of apples vary depending on many factors as weather, storage… and a supervised classifier have to be re-trained. This means that one needs an important database manually segmented which is very irksome for a human specialist. The use of low-level attention may let us build a self-trainable fruit grading system capable of learning alone each season’s apple defects, and than automatically grade the fruits.

For medical imaging, the low-level attention approach could work for organs where tumours can be visually located by using only grey-levels and no symmetry or shape criteria as the liver, etc… Using a smartly chosen atlas may greatly improve the results. For organs where shape could be important in locating tumours, other features should be used and a high-level approach may be more appropriate.

References

1. Hubel, D.H. "Eye, brain and vision", New York: Scientific American Library, N°22, 1989
2. Treisman, A. M., and Gelade, G. "A feature-integration theory of attention", Cognitive Psychology, 12(1): 97-136, 1980
3. Crabtree, J.W., Spear, P.D., McCall, M.A., Jones, K.R., and Kornguth, S.E. "Contributions of Y- and W-cell pathways to response properties of cat superior colliculus neurons: comparison of antibody- and deprivation-induced alterations", J Neurophysiol., 56(4):1157-1173, 1986
4. Itti, L., and Koch, C. "A saliency-based search mechanism for overt and covert shifts of visual attention", Vision Research, 40:1489-1506, 2000
5. Le Meur, O., Le Callet, P., Barba, D., and Thoreau, D. "A coherent computational approach to model bottom-up visual attention", IEEE PAMI, 2005
6. Walker, K.N., Cootes, T.F. and Taylor, C.J., "Locating salient object features", Proc. of British Machine Vision Conference, 2:557-566, 1998
7. Mudge, T.N., Turney, J.L., and Volz, R.A., "Automatic generation of salient features for the recognition of partially occluded parts", Robotica, 5:117-127, 1987
8. Stentiford, F.W.M., "An estimator for visual attention through competitive novelty with application to image compression", Picture Coding Symposium, pp. 25-27, 2001
9. Boiman, O., and Irani, M. "Detecting irregularities in images and in video", Proceedings of Int. Conference on Computer Vision, 2005
10. Itti, L., and Baldi, P. "A principled approach to detecting surprising events in video", Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 631-637, 2005
11. Stanford, L.R. "W-cells in the cat retina: correlated morphological and physiological evidence for two distinct classes", J Neurophysiol., 57(1):218-244, 1987
12. Mancas, M., Mancas-Thillou, C., Gosselin, B. and, Macq, B. "A rarity-based visual attention map -application to texture description -", Proc. IEEE ICIP, 2006
13. Unay, D., and Gosselin, B. "Stem and calyx recognition on 'Jonagold' apples by pattern recognition", J Food Eng, 78 (2): 597-605, 2007.
14. Hu, X-P., Dempere-Marco, L., and Yang, G-Z. "Hot Spot Detection Based on Feature Space Representation of Visual Search", IEEE Transactions on Medical Imaging, 22(9), pp. 1152-1162, 2003
15. Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M.-F., and Debeir, O. "Automatic Grading of Bi-Colored Apples by Multispectral Machine Vision", Pattern Analysis and Applications, in review.