Adaptive Vision Leveraging Digital Retinas: Extracting Meaningful Segments

Nicolas Burrus, Thierry Bernard

Adaptive Vision Leveraging Digital Retinas: Extracting Meaningful Segments

Nicolas Burrus and Thierry M. Bernard

ENSTA / UEI
32 Boulevard Victor, 75015 Paris, France
firstname.lastname@ensta.fr

Abstract. In general, the less probable an event, the more attention we pay to it. Likewise, considering visual perception, it is interesting to regard important image features as those that most depart from randomness. This statistical approach has recently led to the development of adaptive and parameterless algorithms for image analysis. However, they require computer-intensive statistical measurements. Digital retinas, with their massively parallel and collective computing capababilities, seem adapted to such computational tasks. These principles and opportunities are investigated here through a case study: extracting meaningful segments from an image.

1 Introduction

Designing robust vision algorithms is a serious challenge. The infinite variability of natural images and the difficulty to find precise rules specifying how to solve a -generally trivial for a child- vision problem are greatly contributing to this complexity.

Dynamically adapting algorithms to images they encounter is certainly a way to overcome part of this complexity. Being robust also requires a strict management of algorithm parameters, by relating them to a physical quantity, deducing them from image properties, learning them, selecting them to optimize further treatments through a closed loop process, etc.

Recently, an almost parameterless statistical framework has been proposed [6], relying on the so-called Helmholtz principle, which states that meaningful events are events whose probability to appear in a purely random environment is very low. It seems that human perception follows this rule to some extent, and this framework has been applied notably to gestalts detection with some success [3, 5]. The absence of parameters mainly comes from the fact that no generative probability model of events has to be defined, but only a rarity measure in a well-defined random environment. In addition, some properties of the image can be taken into account to define the random model in which the rarity of other properties will be evaluated, enabling adaptation to the analyzed image.

Such kind of approaches generally requires a considerable number of potentially meaningful events to be evaluated, making real-time processing difficult or impossible. Adaptivity also requires the computation of global quantities, such as probability distributions over the image, which are time- and power-consuming

to obtain using a standard computer. The reason is that pixel data have to be transfered many times, for each pixel, from memory to processor.

To ease such global computations, less standard architectures are needed, that better combine processor and memory, in a more distributed fashion. The latter issue is addressed for more general reasons by the computer architecture community (e.g. [14]). But, in this paper, we focus on artificial retinas (also known as vision chips [11]), which mix processor and memory in an extreme way. Indeed, these are smart imaging sensors, with processing ressources in each pixel, thus making massively parallel image array processors without input bottleneck.

On the output side, many retinas are fitted with a global adder (analog as [1] or digital as [10]), able to quickly provide the sum of pixel data over the whole image. The global adder has been used to measure image moments, e.g. for extracting the position of a target. More powerfully, it has been used in a feedback scheme to allow image capture with automatic histogram equalization [12]. We believe that this feedback scheme is worth being systematically extended to image processing : sums provided by a global adder can be surely taken advantage of to better control the way in which images are processed. In particular, it can provide at low cost statistical measures about images in order to make algorithms more adaptive, therefore more robust, as we are looking for. Of course, this only makes sense with programmable retinas, that is retinas with a programmable processor in each pixel - thus allowing versatile image processing - such as the digital retinas we design in our lab [13, 9].

In the present paper, our goal is two-fold:

- show the potential of these general principles through a case study: meaningful segment detection;
- use this experience to improve algorithm/architecture adequation, by motivating future evolutions of both vision system architectures and statistical methods.

In the following, we deal with segment detection in natural images, a standard primitive which can be interesting in artificial environments and which drastically reduces the information contained in an image, while keeping important features. We are looking for an adaptive, parameterless algorithm taking advantage of retina capabilities.

After a global overview of the proposed algorithm in Section 2, Section 3 focuses on segment candidates extraction on digital retinas, then Section 4 gives statistical criteria to decide whether the candidates are meaningful or not. Finally, quantitative results are given in Section 5 and questions raised by this study are discussed in Section 6.

2 Overview of the algorithm

2.1 What is a segment?

Definition 1. *A segment in a cone C is a one-pixel thick connected set of pixels, such that:*

- *each pixel has a local gradient direction in C;*
- *for each non-extremity pixel, the direction of the vector formed by its two neighbors is also in C.*

Gradient vectors are computed using a Sobel operator. To keep cone belonging easy to check, only eight cones are considered, corresponding to the possible angles in a 5x5 discrete neighborhood, as shown in Figure 1. The main steps of the algorithm are as follows:

1. groups of pixels conforming to the definition of segments are extracted as briefly described in Section 3 and their properties (length, mean of pixels gradient magnitude, etc.) are attached to one of their extremities;
2. the extremities are selected by an *a contrario* statistical criterion, as will be detailed in Section 4. The criterion takes into account global image measures and for each segment answers the question: "could a segment with the same properties possibly be observed in a purely random environment?";
3. segments for which the answer is "no" are reconstructed from their representative extremity, resulting in a binary image of meaningful segments.

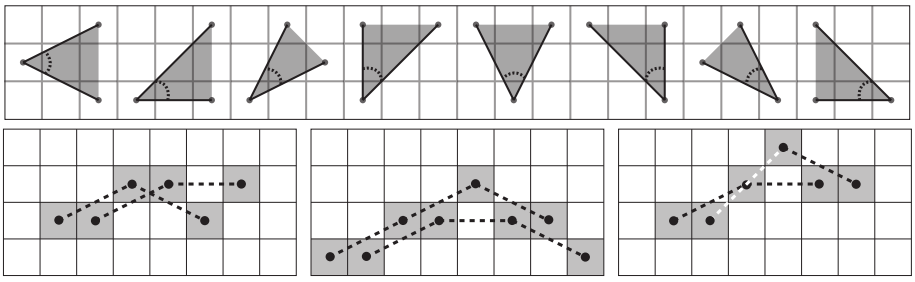


Fig. 1. Top row: the eight overlapping direction cones. Bottom row: illustration of Definition 1 for the first cone (horizontal). Dark pixels must have a local gradient direction in the cone. Dashed lines show the direction induced by the neighbors of each point. Among the three sets of dark pixels, only the left and central ones are segments according to Definition 1, since all dashed lines are within the cone. The right one is not a valid segment since the white dashed line lies outside the cone (too large angle).

3 Candidates extraction

Segment extraction, the first step of the overall algorithm (see Section 2) is itself performed in three steps, as illustrated in Figure 2:

1. Eight binary images are computed, each representing a direction cone in which segments will be looked for. Any pixel with a determined gradient direction is marked as white in the direction image(s) of which it matches the direction.
2. In each direction image, connected sets of white pixels are made one-pixel thick, such that the remaining pixels lie where the image gradient magnitude is maximal in the orthogonal direction.
3. In each direction image, independently, connected groups of white pixels which match our segment definition are reduced to one of their extremities, to which segments properties are attached. This step is performed by iterative segment erosion: for e.g. horizontal segments, left extremities are removed at each iteration. Before removal, extremities transfer all the information gathered so far to their right neighbor. At the end of this step, extremities support the needed properties of their associated segment.

   The extremities are now ready to go through the selection step.

4 Candidates selection

4.1 About the Number of False Alarms (NFA)

The question addressed in this section is: observing a segment with some properties, how to decide whether this segment is meaningful or whether it is just an artefact or coincidence? Two segment properties will be considered, the mean of the gradient magnitude of the segment pixels in Section 4.2 and the length of the segment in Section 4.3. In the spirit of [6], we chose to reason *a contrario*, i.e. instead of computing the probability to observe such a segment in a natural image, we try to answer the question: could the observed segment possibly appear in a noise image? If not, it must be due to a real world phenomenon: object, shadow, etc. To quantify this *a contrario* likelihood, we recall the definition of the number of false alarms of an event.

Definition 2. *The number of false alarms of an event E is the expected number of occurrences of E in a random environment.*

Using the NFA, the notion of ϵ-meaningfulness may be defined, with ϵ a strictly positive (possibly $\ll 1$) real number.

Definition 3. *An ϵ-meaningful event E is an event such as $NFA(E) < \epsilon$. A 1-meaningful event is simply called a meaningful event.*

In practice, choosing $\epsilon = 1$ means that the event is expected to appear less than once in a random context. It is a sound choice as the NFA generally has a

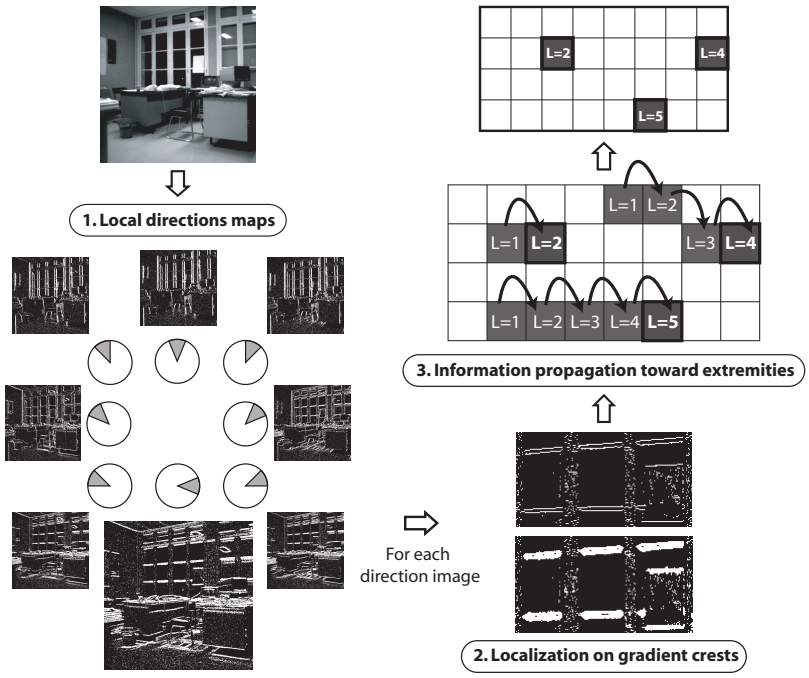


Fig. 2. Overview of the segment extraction algorithm. Step 1 produces eight binary direction images by analyzing local gradient directions in the original image. Each image corresponds to a direction cone and white pixels are pixels whose local direction fits in the cone. Step 2 and 3 are illustrated on portions of the horizontal direction image. Step 2 makes connected sets of white pixels one-pixel thick by only keeping pixels lying where the image gradient magnitude is maximal. Step 3 iteratively propagates segment information (length in the example) from the left extremity to the right one, following the rules of Definition 1. Finally, only right extremities will be fed to the selection step.

exponential behavior w.r.t. event properties, so the dependence on ϵ is rather a log-dependence.

An example is provided in Figure 3 to illustrate how deviations from a random model make events perceptually meaningful.

Finally, to decide whether an event is meaningful in this framework, we need three elements, chosen *a priori*:

1. What kind of events are we looking for?
2. Which event's property should be analyzed?
3. What is the relevant *a contrario* random model?

For segments detection, we have already answered question 1 with Definition 1. Question 2 and 3 will be shortly addressed in Section 4.2 and Section 4.3.

Fig. 3. Illustration of the number of false alarms. Unlike the more complex segments matching Definition 1, the events considered here are simply "perfectly horizontal segments". The property associated to the event is the length and the *a contrario* random model is an image whose black pixel density is the same as in the original image but where points are spatially independent and uniformly distributed. A segment of 6-pixel length has been artificially added to each image, which otherwise follows the *a contrario* model. As the density of black pixels gets smaller, the number of false alarms of the 6-pixel segment decreases and it becomes an increasingly important deviation from the *a contrario* model. Our perception seems to follow similar rules: the segment becomes detectable with a NFA of 0.5 and obvious for a NFA of 10^{-3}.

4.2 Selection based on a gradient magnitude criterion

A natural criterion to start with is based on the contrast between the segment and its neighborhood, an approach comparable to [4]. Whereas [4] was interested in the minimal gradient intensity value along level lines, here we consider the mean gradient value along the segment to be less sensitive to outliers. The gradient magnitude is computed using 2x2 finite differences to avoid creating artificial correlation between pixels (see [2] for detailed explanations).

The *a contrario* model we choose is a white uniform noise with the same gradient magnitude distribution as in the original image, but where pixels are spatially independent and uniformly distributed. This makes the model adapted to the global gradient properties of the image, while making pixel spatial organization in the original image the source of large deviations. The rarity of a segment will not come from the fact of observing pixels with high gradients in itself, but from the fact that a group of adjacent pixels contains many high gradient values.

Under these assumptions, we can compute a number of false alarms for segments.

Definition 4. *Let μ_g and σ_g be respectively the gradient magnitude mean and standard deviation on the whole image. Let $N_{segments}$ be the number of candidate segments detected in the image. Let $\mu(S)$ be the mean gradient value of a segment S and $L(S)$ its number of pixels. Then*

$$NFA(S) = N_{segments} \times (1 - normcdf(\mu(S), \mu_g, \frac{\sigma_g}{\sqrt{L(S)}}))$$

where $normcdf(x, \mu, \sigma)$ is the normal cumulative distribution function with mean μ and deviation σ applied to x.

This definition comes from the central limit theorem. Under the *a contrario* assumption, a segment can be seen as a collection of $L(S)$ independent and identically distributed samples of the image, thus the mean of their gradient magnitude should follow a normal law if $L(S)$ is big enough, according to the central limit theorem. Since we have $N_{segments}$ candidates, the expected number of segments having a deviation from the random model at least as large as the one observed for S is the $NFA(S)$ of Definition 4.

We have implemented this selection criterion on a standard computer, but not on digital retinas because of some limitations of the current generation. This is discussed in more details in Section 6. This has led us to consider a different criterion as defined in Section 4.3, enabling a fast implementation on our retina.

4.3 Selection based on segment length

Instead of considering the gradient values along the segment, one might wonder what minimal length is required for a segment to be meaningful, whatever its contrast. The question becomes: in a direction image I_d, how many chains of pixels of length l matching Definition 1 would be expected under an *a contrario* random model?

The choice of the *a contrario* model is somewhat similar to the one of Section 4.2. Taking I_d, the *a contrario* image is an image whose white pixels density is the same as I_d, but where pixels are spatially independent and uniformly distributed. This way, the selection adapts to the global density of pixels sharing the same local directions, and large deviations correspond to large groups of adjacent white pixels.

Unfortunately, even under these fairly simple assumptions, a NFA is analytically difficult to compute. This complexity comes from the rather particular connectivity induced by our definition of segments, which makes the number of candidates difficult to count. Still, we have to find the minimal length above which the NFA will be less than one, depending on the direction image white pixel density. This can be evaluated by stochastic Monte Carlo simulation, that is, by analyzing the actual statistical distribution of the lengths of segments occurring in randomly generated images.

Let I_d be a direction image of size $N \times N$, and p_{white} its white pixel density $\frac{\#whitepixels}{N \times N}$. The following procedure is repeated M times:

1. Generate a random black and white image of size $N \times N$ by drawing independently for each pixel the value white or black according to p_{white} ;
2. Apply on it the segment extraction algorithm of Section 3 ;
3. Store the histogram of the segment lengths.

This results, for one p_{white} value, into a collection of M samples of segment lengths histograms, as depicted in Figure 4. We are looking for the length threshold L_{min} which ensures $NFA(L(S)) < 1$ whenever $L(S) \geq L_{min}$. $NFA(L(S))$ is the expectation of the number of occurrences of segments with length greater than $L(S)$ in random images. It can be estimated from the simulations. Having a $NFA < 1$ means the expected maximal segment length in a random image must be less than L_{min}.

From the M simulations above, one can compute a confidence bound on the expected maximal length. Let X_i the maximal length observed in random image i. The empirical mean μ and empirical deviation σ of the maximal length are then:

$$\mu = \frac{1}{M} \sum_{i=1}^{M} X_i \qquad \sigma^2 = \frac{1}{M-1} \sum_{i=1}^{M} (X_i - \mu)^2$$

Let μ_{true} be the real expectation of the maximal segment length. When M is big enough, the random variable $Y = \frac{(\mu - \mu_{true})\sqrt{M}}{\sigma}$ follows a Student law with $M-1$ degrees of freedom. We construct a bound on μ_{true} such that:

$$P(Y < t) = \alpha$$

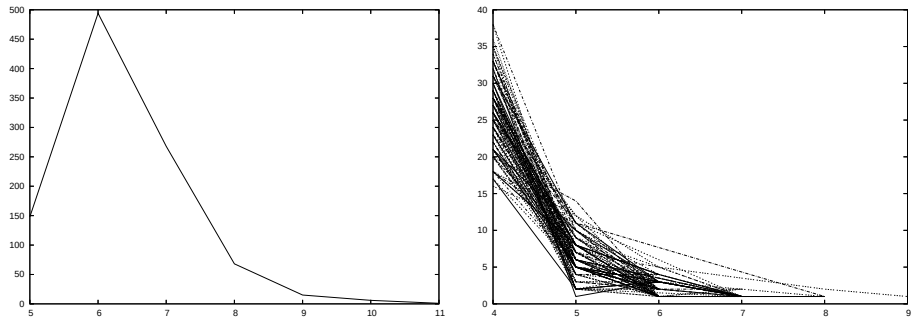
with α the confidence we want. We can get α arbitrarily close to 1 by increasing M and t. Note that α gets exponentially closer to 1 with respect to t, so for $M = 1000$ and $t = 3.1$ the Student law gives $\alpha = 0.999$, leading to:

$$P(\mu_{true} < \mu + 3.1 \frac{\sigma}{\sqrt{1000}}) = 0.999$$

Thus, choosing $L_{min} = \mu + 3.1 \frac{\sigma}{\sqrt{1000}}$ ensures $P(NFA(L(S)) < 1) = 0.999$ whenever $L(S) \geq L_{min}$. Figure 4 shows the typical exponentially decreasing distribution of maximal lengths values.

Finally, running simulations for different p_{white} gives a table of minimal lengths thresholds. Then, the selection algorithm becomes, for each direction image:

1. Estimate p_{white} from the direction image using the global adder;
2. Lookup in a table the corresponding minimal length threshold;
3. Remove extremities associated to segments having a too small length.



(a) (b)

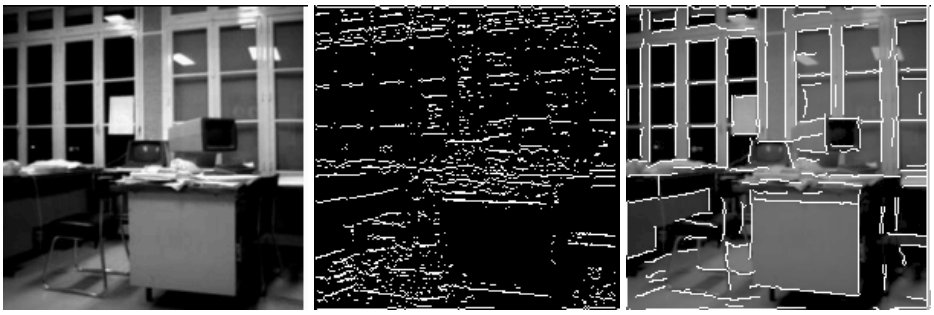
Fig. 4. Horizontal axes are segments length. Vertical axes are the number of occurrences. (a) Histogram of maximal segments lengths for 1000 different uniform noise images with $p_{white} = 0.1$ (b) Histogram of segments lengths for 100 different uniform noise images with $p_{white} = 0.1$, only segments of length greater than four are considered.

5 Quantitative results

Meaningful segment extraction based on the segment length criterion (see Section 4.3) has been successfully implemented on Pvlsar34, a home-made digital retina of 200x200 pixels, each containing a pixellic Boolean processor with 42 bits of local memory, under SIMD control. To evaluate our algorithm, we bypassed Pvlsar34 capture abilities by transferring standard images into retina memory, scaled to 200x200 pixels and reduced to 64 gray levels to save retina memory. Figure 5 shows an example of segment extraction on an interior scene. Note that there is no free parameter to set in the method, since the segment length thresholds are automatically derived from the direction images densities.

Figure 6 clearly illustrates the benefits of context adaptation. If meaningful segments had been selected on the "house" image with the same threshold as the one derived for the "desk" image, a lot of false alarms would have been obtained, as shown on Figure 6(c).

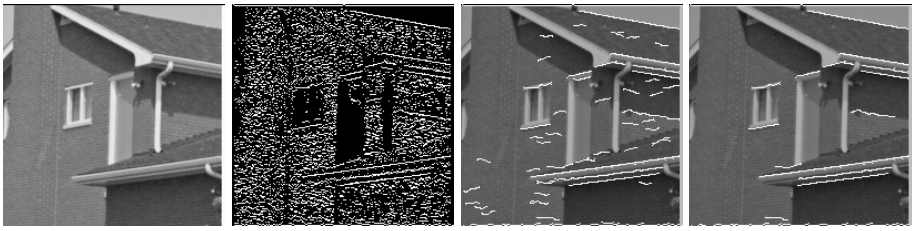
The meaningful segment extraction algorithm runs in real-time on our digital retina Pvlsar34, at video rate. It runs 10 times slower on an up-to-date personal computer, with 2 images processed per second. This factor of ten seems ridiculously small considering the massive parallelism (40k processors operating together) available in Pvlsar34. One of the reasons is that Pvlsar34 is operated



(a) (b) (c)

Fig. 5. (a) "Desk" image (b) Horizontal direction image, localized on gradient maxima, $p_{white} = 0.09$ (c) In white: segments which have a meaningful length.

at a low frequency of 5MHz, which ensures a very low power consumption of a few tens milliwatts, 3 to 4 orders of magnitude as small as that of a PC! Another reason is examined in the next section.



(a) (b) (c) (d)

Fig. 6. (a) "House" image (b) Horizontal direction image, localized on gradient maxima, $p_{white} = 0.2$ (c) In white: horizontal segments which would have a meaningful length if the same threshold as for the desk image was used (d) In white: meaningful horizontal segments according to the new threshold.

6 Discussion

6.1 About statistical criteria

We try to avoid ad-hoc parameters. But there are still a number of choices which are subject to discussion: the definition of segments, the considered properties, the chosen *a contrario* models, etc. Of course, ideally, those assumptions should be replaced by objective measurements or justifications. What should be noted however is that the nature of the remaining *a priori* is never quantitative but

only qualitative. This means the *a priori* decisions always rely on reasoning, never on numerical, empirical values.

Another limitation of current *a contrario* approaches is the global nature of the statistics from which large deviations are measured. There is an underlying assumption of image spatial stationarity, which is obviously not true in the general case. A direct consequence is the so-called blue sky effect, where a very flat zone in the image influences detection in other, shaky, parts. In [2], relying on level lines nesting properties, local meaningful level lines are detected using the statistics of the region associated to their closest surrounding meaningful level line. However, this is not easily applicable in our case.

Finally, we notice that our meaningful segment extraction algorithm does not use so much the retina abilities to compute global statistics though global summations. Using it much more intensively could provide interesting algorithmic and statistical innovations in the future.

6.2 About candidates and properties extraction

Whereas using a digital retina fitted with a fast global adder is a source of algorithmic inspiration, implementing algorithms on it suggests architectural improvements. Here, what are lessons to draw? Whereas gradient and direction-related computations are fast, information propagation toward extremities takes most of the computation time. Indeed, these propagations are done synchronously in Pvlsar34, and only a few pixels (the extremities) are actually performing useful computations at each iteration. This is clearly under-exploiting massive parallelism. To drastically reduce propagation delays and energy consumption, asynchronous retinas (e.g [8, 7]) have been proposed, allowing efficient computing of regional quantities, which are typical of middle level vision. For example, the computation of a gradient magnitude mean over a segment would become tractable, thereby enabling more complex properties to be statistically analyzed. More generally, to cope with statistical detection of big groups of pixels, we believe asynchronism will play an important role and we are currently working on the realization of the model described in [9].

6.3 Conclusion

This paper shows a first step towards more adaptive, parameterless and statistically founded algorithms taking advantage of digital retinas philosophy and capabilities. We have developed an original algorithm for the detection of meaningful segments. On the presented images, detected segments indeed seem to be the important ones. These encouraging results are obtained in spite of the relative simplicity of the statistical segment model we have chosen. Implementation on our home-made digital retina has allowed real-time operation but has recalled the limitations of its synchronous SIMD character for middle level vision.

References

1. T. M. Bernard and P. E. Nguyen. Vision through the power supply of the NCP retina. In *SPIE, Charge Coupled Devices and Solid State Sensors V*, volume 2415, pages 159–163, 1995.
2. F. Cao, P. Musé, and F. Sur. Extracting Meaningful Curves from Images. *Journal of Mathematical Imaging and Vision*, 22(2):159–181, 2005.
3. A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful Alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.
4. A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3):271–284, 2001.
5. A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, 2003.
6. A. Desolneux, L. Moisan, and J.-M. Morel. Maximal meaningful events and applications to image analysis. *Annals of Statistics*, 31(6):1822–1851, 2003.
7. B. Ducourthial and A. Mérigot. Parallel asynchronous computation for image analysis. *Proceedings of the IEEE*, 90(7):1218–1229, July 2002.
8. V. Gies, T. M. Bernard, and A. Mérigot. Convergent micro-pipelines: a versatile operator for mixed asynchronous-synchronous computations. In *IEEE International Symposium on Circuits and Systems*, pages 5242–5245, 2005.
9. V. Gies, T. M. Bernard, and A. Mérigot. Asynchronous regional computation capabilities for digital retinas. In *IEEE Workshop on Computer Architecture for Machine Perception and Sensing, Submitted*, 2006.
10. T. Komuro, S. Kagami, and M. Ishikawa. A dynamically reconfigurable SIMD processor for a vision chip. *IEEE Journal Of Solid State Circuits*, 39(1):265–268, 2004.
11. A. Moini. *Vision Chips*. Kluwer Academic Publishers, 2000.
12. Y. Ni, F. Devos, M. Boujrad, and J. H. Guan. Histogram-equalization-based adaptive image sensor for real-time vision. *IEEE Journal Of Solid State Circuits*, 32(7):1027–1036, 1997.
13. F. Paillet, D. Mercier, and T. M. Bernard. Second generation programmable artificial retina. In *IEEE ASIC/SOC Conference*, pages 304–309, 1999.
14. M. B. Taylor et al. Evaluation of the raw microprocessor: An exposed-wire-delay architecture for ILP and streams. In *International Symposium on Computer Architecture*, pages 2–13, 2004.