Local Visual Pyramids for Image Segmentation

Hubert Konik, Eric Dinet and Bernard Laget
Université Jean Monnet
lnstitut d'Ingénierie de la Vision
Saint-Etienne, Cedex, France

Abstract

A new concept that uses a pyramidal approach in image segmentation is proposed. The method introduces local visual pyramids in reply to inherent limitations of the classical pyramidal structure (e.g. concerning small or elongated objects). The aim of local visual pyramids is to simulate the human vision in its attention focusing processes through an individual and a contextual analysis. Actually, a local visual pyramid is a hierarchy of fine to coarse resolution versions of an image where the resolution decreases twofold between consecutive levels. The kernel of such a pyramid is defined using the distribution of human visual acuity i.e. the kernel weights visual information roughly by an inverse function of the visual angle. Such an approach allows to improve the precision and the robustness of pyramid-based segmentation processes.

Introduction

On seeing an image even for the first time, a human observer is able to perform its segmentation apparently without any knowledge. Nevertheless, when the image is very unfamiliar and very textured, the requested time is more considerable and the boundaries of the segmented regions are rougher. In this context, the approach used by the human visual system seems to consider the image as a whole before being interested by the details. To emulate this process, over the last few years, multiresolution techniques have been introduced. By definition, a pyramid is more precisely a hierarchy of fine to coarse resolution versions of an image, where the details progressively disappear. Unfortunately, the classical structure suffers from a lack of adaptability, especially concerning small or elongated objects. Then, we will first briefly review the pyramidal approach and its application in image processing. Then, we will develop a new concept for segmentation, using local visual pyramids in order to simulate the human vision in its attention focusing processes through an individual and a contextual analysis.

Using a Multiresolution Approach

On seeing a scene, an observer usually goes away to have a global view while she gets nearer to delineate each entity. Then, first introduced for improving computation times of image processing algorithms[l], image pyramids have played a more and more important role during the last decade,[2] notably because of this possible computational "up-and-down movements".

The basic idea of the pyramid structure is to produce a stack of interrelated images with progressively reduced resolution. The sampling rate of these lower-resolution images is reduced in accordance with the elimination of the higher frequencies. Generally speaking, it contains the image at different levels decreasing twofold from one to another. The stack of levels is produced by filtering the successive images and by a reduction of the sampling rate. In this way, it can simulate the human vision during a segmentation step: far to extract relevant parts and close to delineate precisely each of them.

The Construction

It is noteworthy that many different functions have been introduced to realize such a representation. Typical functions used are convolutions with different kernels filters, morphological operations or model fitting.[3] Generally speaking, filters are iteratively applied to generate the sequence of images with progressively decreasing resolution. Another classical scheme is based on alternating sequential morphological filters. Before sampling images must be in this particular case morphologically filtered by an opening or a closing,[4] giving morphological representations.[5] With a similar approach, a higher level can also be produced by assigning the extremum of the lower level window to the representative pixel.[6]

The most useful scheme, well matched to the human visual encoding,[7] is the Burt’s one.[8] The principle is to repeatedly apply a low-pass filter, for example a Gaussian one, that generates reduced-resolution versions of the input image. The low-pass filter is implemented by convolution of a kernel w with the image. Using a kernel of size K × K, each pixel value in the current level is a weighted sum of the pixel values—the sons— in a K × K neighborhood in the previous one centered around that pixel. Considering an even kernel, that leads to a less sensitive to rotations structure, each element (i,j) at level h is computed as follows:

$$f_h(i,j) = \sum_{u=0}^{K-1}\sum_{v=0}^{K-1} w(u,v) \cdot f_{h-1}\left(2i+u-\left\lfloor \tfrac{K-1}{2} \right\rfloor, 2j+v-\left\lfloor \tfrac{K-1}{2} \right\rfloor\right)$$

To be Focused to Improve the Treatment

Whatever kernel you take, the conventional structure is always fixed and too rigid to be adapted for any application. Actually, a critical view of the classical pyramid implies that it has to be rejected as a general segmentation tool, due particularly to the sub-sampling introduced at higher levels.[9] More particularly, the case of non-isolated, countless or elongated objects constitute some prohibitive unsolvable restrictions with such a global tool. The idea is then to simulate the human vision in its attention focusing processes through an individual and a contextual analysis inside a (or several in the elongated case) local pyramid roughly containing it. In fact, not only one global pyramid but one for each relevant part is constructed, as shown in Figure 1.

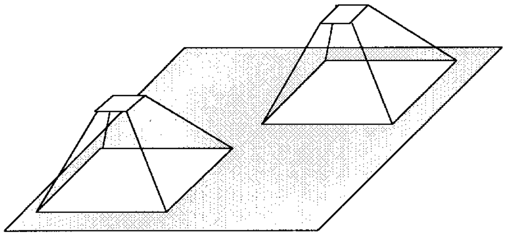

Figure 1: Local pyramids used in the visual multiresolution approach.

Now, the problem consists in choosing the best kernel of size K × K in a human visual approach.

Definition of the Kernel Using the Visual Acuity

From a mathematical point of view, the even kernel used in the construction with a 4-decimation must verify different constraints:[8]

- $\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} w(i,j) = 1$
- $$\begin{aligned} w(i,j) &= w(K-1-i,j) \\ &= w(i,K-1-j) \qquad \forall i,j \in [0,K-1] \\ &= w(K-1-i,K-1-j) \end{aligned}$$
- $$\begin{aligned} |w(i,j)| \le |w(k,l)| \quad & \forall 0 \le i \le k \le \left\lfloor \tfrac{K-1}{2} \right\rfloor, \\ & 0 \le j \le l \le \left\lfloor \tfrac{K-1}{2} \right\rfloor \\ |w(i,j)| \ge |w(k,l)| \quad & \forall \left\lfloor \tfrac{K-1}{2} \right\rfloor \le i \le k \le K-1, \\ & \left\lfloor \tfrac{K-1}{2} \right\rfloor \le j \le l \le K-1 \end{aligned}$$
- $\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} w(x+2i, y+2j) = \frac{1}{4} \quad \forall x,y = 0,1$
- $w(i,j) = h(i) \cdot h(j)$

In fact, this kernel can be determined, when K = 4, in terms of two free variables a and b:

$$\begin{bmatrix} b^2 & ab & ab & b^2 \\ ab & a^2 & a^2 & ab \\ ab & a^2 & a^2 & ab \\ b^2 & ab & ab & b^2 \end{bmatrix}$$

The size is appropriate considering the size of the bases of the different local pyramids (if this size is 64 × 64, the third level is already of size 8 × 8).

Now, our main feature is to "mimic" as far as possible the human visual system. Its sharpness of sight is closely linked to the number and to the density of visual cells. And the greatest visual acuity is provided by the fovea. Beyond this particular area, the acuity falls roughly in inverse proportion to the visual angle[11], as shown in Figure 2.

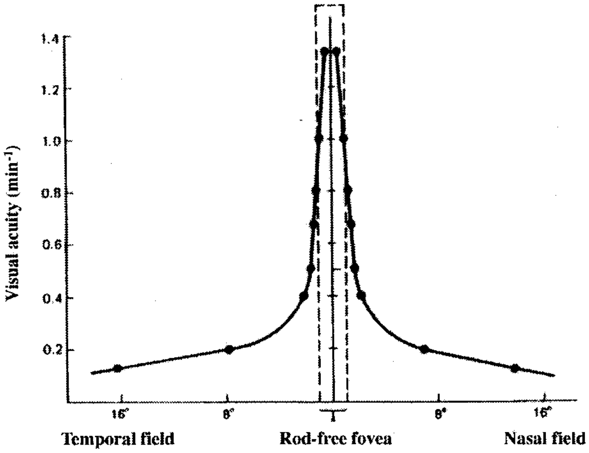

Figure 2: The visual acuity of man plotted as a function of visual angle.

Thus, the fovea is used to examine a selected area of a scene in detail whereas the peripheral vision only allows to detect the elements that may interest the observer. Then, in order to achieve a selective analysis of an image, foveal mechanisms can be simulated in a local visual pyramid.

First of all, let state a significant result for the central elements of a pyramid: they only generate precisely four fathers at each level through the pyramid. Figure 3 shows the central elements at each level of the local visual pyramid until the level of size 2 × 2.

In practice, two different kernels are used: one with a strong weight a and a negative variable b for the central elements, and another one with two equal weights for the other convolutions.

The Construction for Color Image Processing

Presently, local visual pyramids can be used for color image processing, that is still limited by the very significant amount of data. And as color processing lacks effective fast algorithms, pyramidal techniques are well-adapted to minimize the information in accordance with the human vision

So, we propose to generalize our new tool for color images. First of all, the problem of linear color mixing is important: it allows one to compute, thanks to a real spatiocolor approach, a new set of representative colors that is more relevant with regard to the full-resolution image.

Regarding the construction of a pyramid—a convolution and a subsampling operation— it must be done in a linear color slpace as RGB (the gamma-corrected digital color space). Then, we propose to compute three local color visual pyramids for each focusing point, one per color component.

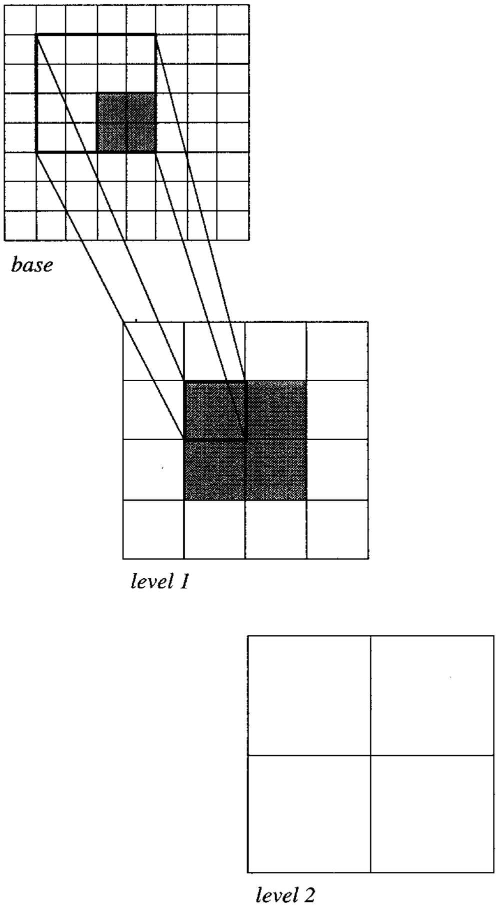

Figure 3: The evolution of the central elements through the pyramid

The Segmentation Method

First of all, let us suppose the focusing points have been extracted. A focusing point corresponds actually to the center of a focusing area, where the vision system must focus its resources to a detailed analysis. We will present an application where this extraction is automatic, but without any knowledge on the image, this process is still restricting. Now, let consider the process inside one local visual pyramid.

The Bottom-up and Top-down Processes

The segmentation scheme is divided into two steps: the detection via a bottom-up process and the delineation via a top-down one. Figure 4 sums up this multiresolution process.

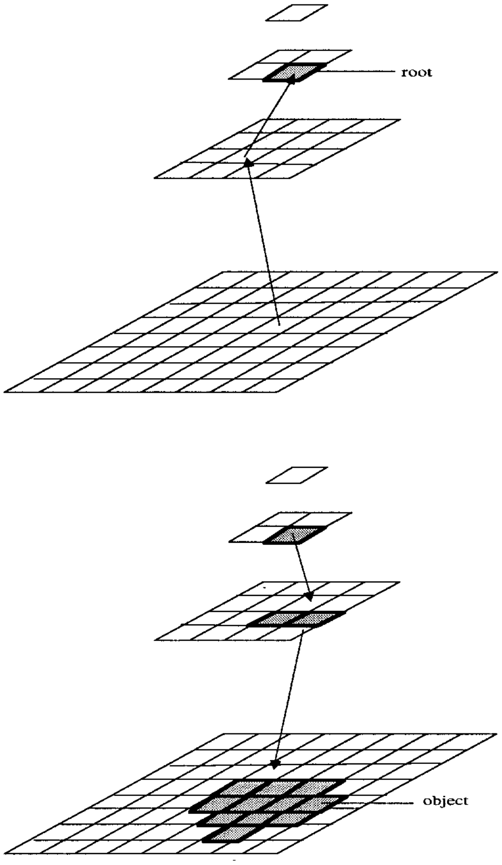

Figure 4: The multiresolution processes.

More precisely, a directional bottom-up process is used to derive roots which define starting points of a top-down process. While the classical extraction[13], generating only one root, seems to be too restrictive with the objects we are concerned with, we introduce the detection of four roots. Each corresponds in fact in the following through each father presented in Figure 3. For every “orientation”, a directional contrast is computed at each level. By comparing its evolution between consecutive levels, the best element representing the object is chosen. Roots correspond to the elements in the upper levels where the contrast is maximum —begins to decrease for the first time.

Then, assuming that four roots are obtained, they define starting points of the top-down process. Among all the sons, only the similar enough elements are agregated to the object. This adjustement process is iterated until the base of the local visual pyramid.

Experimental Results

The applicative aim of our segmentation tool is to extract tufts due to abrasion. They appear in an image as sharp emergent regions, as shown in Figure 5. The results are presented in Figure 6, where the influence zone of each object (Voronoi cells) is given.

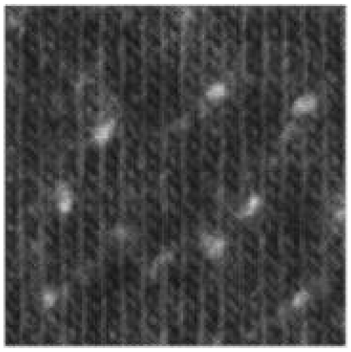

Figure 5: The textile image.

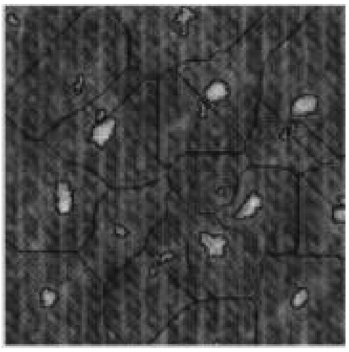

Figure 6: Results of the segmentation.

Conclusion

The pyramid structure corresponds to a hierarchy of fine to coarse resolution versions of a digital image. Such a structure is classically used in segmentation processes through "up-and-down movements". In that way segmentation methods are divided into two steps: the detection via a bottom-up process and the delineation via a top-down process.

Nevertheless the classical pyramid structure is often too rigid especially for elongated or small objects. Then we have introduced a new pyramidal approach: the local visual pyramid. The local visual pyramid does not provide a global analysis and simulates human vision through an individual and a contextual analysis. Then the local visual pyramid suppresses some prohibitive intrinsic limitations of the classical pyramid structure. Moreover the use of the distribution of the human visual acuity improves the detection of objects.

References

1. S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing", *Computer Graphics and Image Processing*, **4**: 104 (1975).
2. J. M. Jolion and A. Rosenfeld, "Pyramid framework for early vision", Kluwer (1994).
3. F. Chin, A. Choi and Y. Luo, "Optimal generating kernels for image pyramids by piecewise fitting", *IEEE Trans. on PAMI*, **14:** 1190 (1992).
4. R. M. Haralick, C. Lin, J. S. J. Lee and X. Zhuang, "Multiresolution morphology", *Proceed. Internal. Conf. on Computer Vision*, 516 (1987).
5. A. Toet, "A morphological pyramidal image decomposition", *Pattern Recognition Letters*, **9**: 245 (1989).
6. A. Meisels and R. Versano, Token-textured object delineation by pyramids, Image and Vision Computing, 10:55 (1992).
7. B. A. Wandell, "Foundations of Vision", Sinauer Associates, *Sunderland*, chap. 8 (1995).
8. P. J. Burt," Fast filter transforms for image processing", *Computer Graphics and Image Processing*, **14**: 1190 (1992) .
9. M. Bister, J. Cornelis and A. Rosenfeld, "A critical view of pyramid segmentation algorithms", *Pattern Recognition Letters*, **11**: 605 (1990).
10. H. Konik and B. Laget, "Using local pyramids for more robust object delineation", *IEEE Southwest Symp. on Image Analysis and Interpretation*, San Antonio, 1(1996).
11. M.D. Levine, "Vision in man and machine", McGrawHill Publishing Company, New York (1985).
12. H. Konik, V. Lozano and B. Laget, "Color pyramids for image processing", *Journal of Imaging Science and Technology*, **40**: 535 (1996).
13. A. Rosenfeld and A.C. Sher, "Detection and delineation of compact objects using intensity pyramids", *Pattern Recognition*, **21**: 147 (1988).