ACTA STEREOL 1995; 14/2: 161-168
ORIGINAL SCIENTIFIC PAPER

SAMPLING OPERATORS FOR ACTIVE VISION

Eric DINET, Alain TREMEAU, Bernard LAGET

Institut d'Ingénierie de la Vision - URA CNRS 842
3, rue Javelin Pagnon - BP 505 - 42007 Saint-Etienne Cedex 1, France

ABSTRACT

A perceptual approach has been used to derive a sampling strategy. Such a point of view allows the merger of image analysis techniques based on computer vision and in terms of concepts relevant to human vision. From ideas arising in mathematical morphology used in conjunction with models of the physiological organization of the retina, sampling operators were derived for selective analysis of scenes. These operators can serve as a primitive strategy for guiding the focus of attention which initiates active vision.

Keywords : focus of attention, human vision, mathematical morphology, retina, sampling.

INTRODUCTION

In the last two decades a lot of mathematical techniques have been developed to solve problems met in computer vision. Unfortunately, the solutions offered by these techniques are often too restrictive and too poor in relation to the environment. For example, nowadays any artificial system is able to perform a pattern recognition operation, except into a peculiar framework. It is right that such an operation corresponds to a very difficult task essentially because of the three dimensional nature of the world.

However the ability of the human visual system to perceive our world is absolutely remarkable. Such a system can perform complex tasks apparently without the slightest effort. For example, we have any problem to analyze a natural scene, to separate and generally to recognize each of its components. A question then arises : what features yield a high performance of the human visual system ? This question not only concerns neurobiologists and psychophysiologists but also everyone facing problems linked to visual perception and particularly researchers working in computer vision (Marr, 1982).

In our opinion, if we are expected to build models for image processing, it is necessary to take into account researches realized in the field of human vision. Such a point of view allows to elaborate theories with a new approach and to look at problems from a perceptual angle. More precisely it allows the merger of image analysis techniques based on computer vision and in terms of concepts relevant to human vision.

In that way we have chosen to base our investigation on the organ that originally produces the visual perception : the retina. Indeed the retina is the first neuronal structure involved in the visual perception and it is largely studied by neurobiologists. The works and the

experimentation realized by neurobiologists provide an original basis to conceive new theories and tools for computer vision.
One goal of this paper is to show how knowledge of a peculiar organ implied in vision can be used to conceive new sampling operators for active vision.

THE RETINA AND THE VISUAL ACUITY

The retina is an extension of the brain which covers the back of the eye and receives the light from a source. More precisely the retina is a thin layer of neural tissue consisting of the association of different specialized cells such as a collection of light-sensitive neurones called photoreceptors (Pirenne, 1967). The photoreceptors transform light into neural signals which are communicated to the visual cortex. There are two fundamentally different types of photoreceptors in the human eye : the rods and the cones. The position of these two types of photoreceptors differs across the retina. Fig. 1 shows the distribution of cones and rods according to retinal eccentricity. The study of topographic variations of the retinal structure is essential for the analysis of the visual physiology which is ruled by the duality between cones and rods.

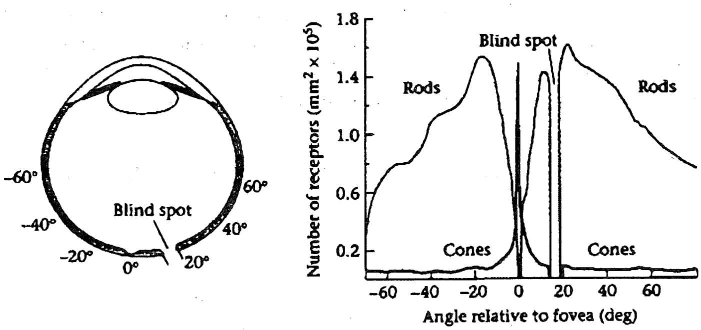


Fig. 1. The spatial repartition of cones and rods across the human retina (Wandell, 1995).

The small retinal area located in the visual axis (i.e. the visual angle is null) and called fovea is characterized by the greater density of cones. In the center of the fovea there are no rods but the highest concentration of cones. The density of cones decreases very rapidly with the eccentricity. Conversely the density of rods progressively increases and reaches a maximum around 20 degrees peripheral to the fovea.
The picture input is digitized and transformed into neural signals by the photoreceptors of retina to be treated by the visual system. This step greatly influences the so-called resolving power of human vision. Indeed the resolving power depends on the optical quality of the image provided by the ocular optical system and also depends on the anatomic and physiologic potentials of the retina to treat this image. The size of the elements in the retinal mosaic has a direct effect on this property.
Resolving power can be considered in terms of visual acuity. Visual acuity or acuity for short is a measure of the ability of the visual system to separate small details. It is conventionally defined as an inverse function of the minimal angle of resolution expressed in minutes of arc. Visual acuity of human eye plotted as a function of visual angle is shown in Fig. 2.
The maximal acuity is provided by the center of the fovea where the density of cones is the greatest. The acuity decreases rapidly within this area and decreases more slowly outside it.

Such a non-uniform resolving power leads to poor visual acuity in the peripheral areas of the retina. Then the location where the input data hits the retina is very important in terms of loss of information.

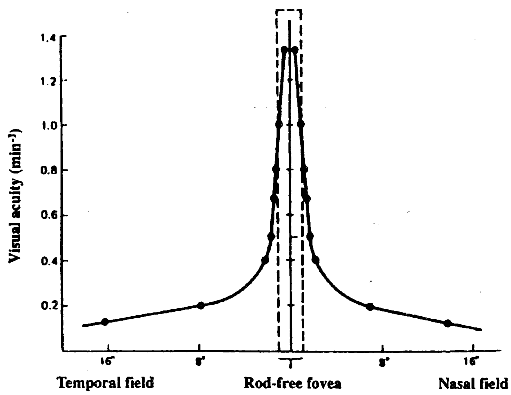


Fig. 2. The visual acuity of man plotted as a function of visual angle (Levine, 1985).

FOCUS OF ATTENTION

The variation of visual acuity can roughly be approached by an inverse function of the visual angle (Burt, 1988). Actually, such a distribution corresponds to a compromise between two conflicting requirements. It provides a wide field of view and a relatively high acuity. The price to pay is that a scene cannot be globally analyzed in detail : it will be sequentially viewed through eye movements.

The peripheral vision allows to detect events that may be important to the observer and allows to guide eye movements for foveal examination. It is then associated to preattentive functions such as detecting or alerting. On the other hand, foveal vision is commonly said to be associated with visual attention. With preattention mechanisms, an observer can focus its attention on a peculiar part of a scene. This notion of focusing of attention or focus of attention is extremely important in human vision since it allows a selective analysis of visual information.

NON-UNIFORM SAMPLING

In order to reproduce the resolving power of human vision we will simulate the spatial distribution of cones with a special sampling. In other words we are going to emulate a sensor whose the resolving power is not the same on its whole surface. More precisely, to approach at best the transition between foveal area and peripheral area, such a sensor has to have an acuity varying in function to the inverse of the visual angle.

Let F be an image defined in a discrete space $\mathscr{E} = \mathbb{Z}^2$. Let o be a point of F that we will call focusing point. This point can be seen as the location of F where a hypothetical observer would guide his gaze. The image of this point would then hit the center of his fovea.

Let $\mathscr{N}$ be a $n \times n$ points area of F centered at o. Keeping in mind that we work in $\mathbb{Z}^2$ we can say that $\mathscr{N}$ appears to a recover maximization of the foveal area of the observer by a simple geometric shape. The area $\mathscr{N}$ corresponds to a part of F where the resolution will be unchanged.

From the point o and outside the area $\mathcal{N}$, the resolution of F will be decreased in inverse proportion to the visual angle. To achieve this goal we will construct a sampling set $\mathcal{S}$. The points s of this set can be seen as the centers of several sensors like cone photoreceptors across the retina. Actually these sensors will be approached by sets $\mathcal{K}_l : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{E})$ of $l \times l$ point size. Each $\mathcal{K}_l$ is centered at a point s of $\mathcal{S}$ and $\mathcal{P}(\mathcal{E})$ is the space of all subsets of $\mathcal{E}$. Sets $\mathcal{K}_l\,(s)$ will be called integration sets. Fig. 3 sums all these notations up.

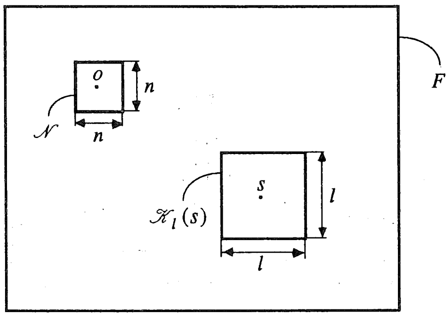


Fig. 3. Illustration of the main notations used in the text. o : focusing point, $\mathcal{N}$: $n \times n$ point area defining the part of the image F perceived with maximal acuity, s : point of the sampling set $\mathcal{S}$, $\mathcal{K}_l(s)$: integration $l \times l$ point size set centered at s.

The spatial repartition of integration sets $\mathcal{K}_l\,(s)$ on image F and consequently the construction of the sampling set $\mathcal{S}$ is made from the focusing point o. In order to well understand how this repartition is carried out, let $\mathcal{K}_l' : \mathcal{E} \rightarrow \mathcal{P}(\mathcal{E})$ be $l \times l$ point sets centered at distinct points x of $\mathcal{E}$. Suppose that these sets are arranged according to their size within different layers of an ordered stack. A layer of this stack consists of a finite number of sets $\mathcal{K}_l'(x)$ having the same size. The same sized $\mathcal{K}_l'(x)$ are juxtaposed and placed in the plane so as the layer they form be centered at o and have a $nl \times nl$ point global size.

From the top of the stack, the layers are ordered by increasing size. The width l of sets $\mathcal{K}_l'(x)$ increases of one unit from a layer to the next. The top of the stack consists of the area $\mathcal{N}$ of $n \times n$ point size. The layer just below the top consists of n^2 sets $\mathcal{K}_2'(x)$ (the notation means that each set $\mathcal{K}_2'(x)$ has a size of 2×2 points). The next layer is composed by n^2 sets $\mathcal{K}_3'(x)$ and so on. A transversal section of such a stack is shown Fig. 4 for an area $\mathcal{N}$ of 5×5 points.

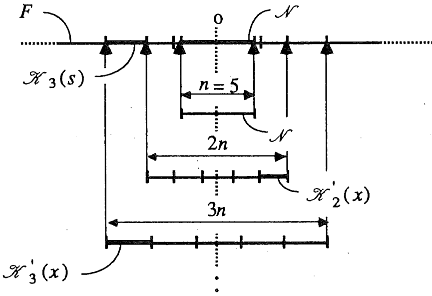


Fig. 4. Spatial repartition of the integration sets. This repartition is schematized by a multilayer model.

When looking at the stack above, the points s of $\mathcal{S}$ are given by the centers x of each $\mathcal{K}_l'(x)$ not completely recovered by a superior layer. In other words, the mosaic of the integration sets $\mathcal{K}_l(s)$ plotted in Fig. 5 is obtained by projecting the layers of the stack on the image under study just like shown by Fig. 4. The sampling set $\mathcal{S}$ consists then of the centers s of these integration sets.

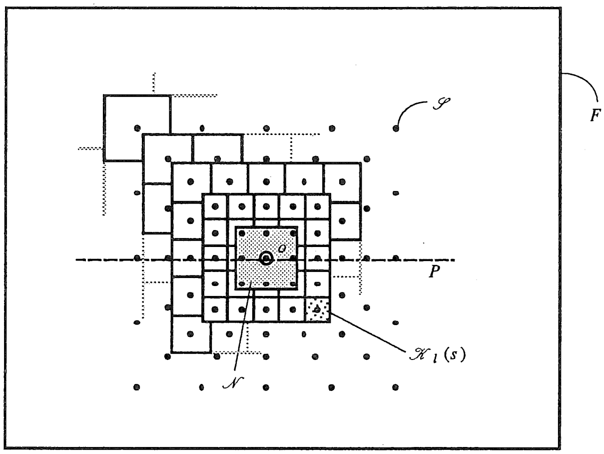


Fig. 5. Mosaic of the integration sets $\mathcal{K}_l(s)$ whose respective centers s form the sampling set $\mathcal{S}$. The straight line P shows the location of the cutting plane of Fig. 4.

Then, for a given focusing point o and a given width n of $\mathcal{N}$, the construction of a sampling set $\mathcal{S}$ can be expressed as :

$$\mathcal{S} = \bigcup \mathcal{S}_l \tag{1}$$

with

$l > 1$, l integer,

$$\mathcal{S}_l = \left\{ s = o - ql \mid q \in \mathcal{N} \text{ and } \mathcal{K}_l(s) \bigcap \mathcal{N}_{l-1}^{c} \neq \varnothing \right\}. \tag{2}$$

In statement (2), $\mathcal{N}_{l-1}^{c}$ corresponds to the complement of set $\mathcal{N}_{l-1}$. The set $\mathcal{N}_l$ is given by :

$$\begin{cases} \mathcal{N}_1 = \mathcal{N} \\ \mathcal{N}_l = \left[\bigcup_{s \in \mathcal{S}_l} \mathcal{K}_l(s) \right] \bigcup \mathcal{N}_{l-1} \end{cases} \tag{3}$$

Each point s of $\mathcal{S}$ corresponds to the center of an integration set $\mathcal{K}_l(s)$. Such a set delimits some points which will be gathered in order to reduce the resolution of the images under study. In that way we will present an integration method based on concepts of mathematical morphology.

MORPHOLOGICAL SAMPLING

In this section we will present a morphological sampling strategy. For a more complete exposition of the basic concepts of mathematical morphology we refer to (Serra, 1982; Ronse, 1990).

Let $\mathscr{E}$ be an euclidean or a discrete space and let $\mathscr{S}$ be a subgroup of $\mathscr{E}$ called the sampling set. Let $\mathscr{G}$ be the set of grey levels which we take here to be $\mathscr{G} = \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ or $\mathscr{G} = \overline{\mathbb{Z}} = \mathbb{Z} \cup \{-\infty, +\infty\}$. Let $\mathscr{K}_l : \mathscr{S} \to \mathscr{P}(\mathscr{E})$, where $\mathscr{P}(\mathscr{E})$ is the space of all subsets of $\mathscr{E}$. Each integration set $\mathscr{K}_l(s)$ corresponds to the support of a structuring function K_l (assume that the functions K_l are equal to $-\infty$ outside their support). From the method of construction of $\mathscr{S}$ we can obviously state that the union of all $\mathscr{K}_l$'s positioned at $s \in \mathscr{S}$ covers $\mathscr{E}$, that is

$$\mathscr{E} = \mathscr{S} \oplus \mathscr{K}_l = \bigcup_{s \in \mathscr{S}} \mathscr{K}_l(s). \tag{4}$$

Moreover each $\mathscr{K}_l(s)$ is symmetric around its center (i.e. $\mathscr{K}_l(s) = \check{\mathscr{K}}_l(s)$), and $\mathscr{K}_l(s) \cap \mathscr{S} = s$, and

$$s \in \mathscr{K}_l(t) \Rightarrow \mathscr{K}_l(s) \cap \mathscr{K}_l(t) \cap \mathscr{S} \neq \emptyset. \tag{5}$$

This condition which holds here is called the sampling condition (Haralick et al., 1985). It contrasts with the sampling theorem in linear signal processing from which only those frequencies below the Nyquist frequency can be reconstructed.
Let F be a function of the complete lattice $\mathscr{G}^{\mathscr{E}}$ (i.e. $F : \mathscr{E} \to \mathscr{G}$). We can then define a dilation $\delta : \mathscr{G}^{\mathscr{E}} \to \mathscr{G}^{\mathscr{S}}$ by :

$$\delta(F)(s) = \sup_{k \in \mathscr{K}_l(s)} \left(F(k) + K_l(k)\right). \tag{6}$$

This dilation realizes a morphological sampling of image F. This sampling depends both on the sampling set $\mathscr{S}$ and on the associated structuring functions K_l.
Now consider the dual mapping $\mathscr{K}_l^* : \mathscr{E} \to \mathscr{P}(\mathscr{S})$. It is defined as

$$\mathscr{K}_l^*(k) = \left\{ s \in \mathscr{S} \mid k \in \mathscr{K}_l(s) \right\}. \tag{7}$$

The theory of adjunctions (Heijmans, 1991) states that the dilation δ is uniquely related to the erosion ε such that for every $F \in \mathscr{G}^{\mathscr{E}}$ and $G \in \mathscr{G}^{\mathscr{S}}$ we have :

$$\delta(F) \le G \Leftrightarrow F \le \varepsilon(G). \tag{8}$$

Then we can prove that the adjoint erosion $\varepsilon : \mathscr{G}^{\mathscr{S}} \to \mathscr{G}^{\mathscr{E}}$ can be written for any function G of the complete lattice $\mathscr{G}^{\mathscr{S}}$ as

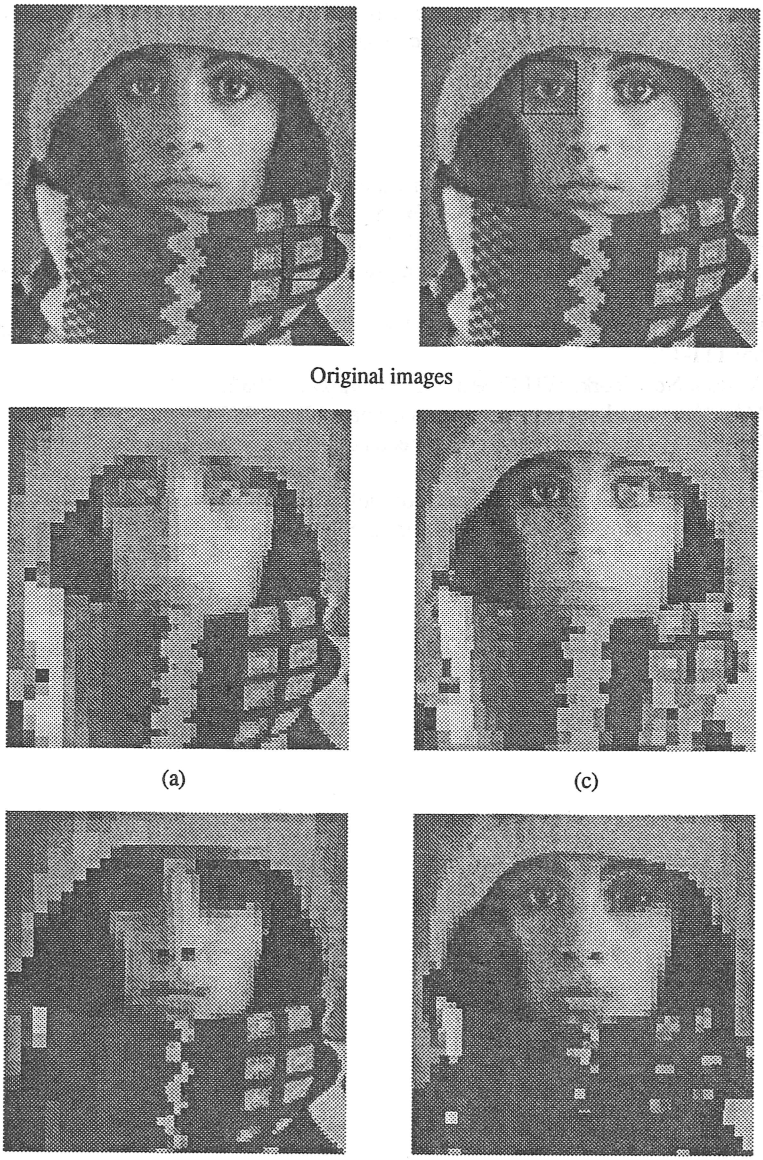
$$\varepsilon(G)(k) = \inf_{s \in \mathscr{K}_l^*(k)} \left(G(s) - K_l(k)\right). \tag{9}$$

The composition $\varphi = \varepsilon\delta$ is a morphological closing on $\mathscr{G}^{\mathscr{E}}$. It consists of the combination of a sampling operator ($\delta : \mathscr{G}^{\mathscr{E}} \to \mathscr{G}^{\mathscr{S}}$) and a reconstruction operator ($\varepsilon : \mathscr{G}^{\mathscr{S}} \to \mathscr{G}^{\mathscr{E}}$).

RESULTS AND DISCUSSION

The sampling presented in this paper is based on the spatial distribution of cones across the human retina as well as concepts of mathematical morphology. We have previously seen that such a distribution is at the basis of the notion of focus of attention, a powerful mechanism for selective analysis.
The sampling naturally provides a loss of information which is all the more important that the distance to the focusing point increases. This loss of information is not the same for the dilation operator and the erosion operator. The erosion exhibits the lowest grey levels of an image while contrary the dilation stresses the highest grey levels as shown by Fig. 6 for a flat structuring function (i.e. the structuring function K_l only attains the value zero on its support). The "blocking effect" observed in the sampled images is all the more important that the sampling brings out a punctual information at the expense of its spatial neighboring context.

Such a configuration provides an accentuation of the contrasts between the intensities of the grey levels conveyed by neighboring integration sets.



Original images

(a) (c)

(b) (d)

Fig. 6. The dark frame on both original images delimits the peculiar area $\mathcal{N}$. (a) and (b), morphological sampling of the left original image respectively by dilation and by erosion. (c) and (d), morphological sampling of the right original image respectively by dilation and by erosion.

The sampling operators allow us to retain in background some areas of study for which local information tends to raise up their neighboring context and may focus the attention. Then the sampling operators can be useful for guiding other visual processes and will simplify their tasks. In other words our sampling operators can serve as a primitive strategy for guiding the focus of attention. This suggested approach can be used when addressing the problem of active vision.

REFERENCES

Burt PJ. Attention mechanisms for vision in a dynamic world. Proc 9th ICPR, 1988: 977-87.

Haralick RM, Zhuang X, Lin C, Lee JSJ. The digital morphological sampling theorem. IEEE Trans Acoustics Speech Signal Process 1989; 37 (12): 2067-90.

Heijmans HJAM. Theoretical aspects of gray-level morphology. IEEE Trans Pattern Anal Machine Intell 1991; 13 (6): 568-82.

Levine MD. Vision in man and machine. New York: McGraw-Hill Publishing Company, 1985: 111-13.

Marr D. Vision. New York: WH Freeman and Company, 1982.

Pirenne MH. Vision and the eye. London: Associated Book Publishers, 1967: 25-36.

Ronse C. Why mathematical morphology needs complete lattices. Signal Process 1990; 21: 129-54.

Serra J. Image analysis and mathematical morphology. London: Academic Press, 1982.

Wandell BA. Foundations of vision. Sunderland: Sinauer Associates, 1995: 45-50.

Presented at the 9th International Congress for Stereology, Copenhagen, August 20th to 25th, 1995.