

Adaptive Image Sampling Based on the Human Visual System

Frédérique Robert[‡], Eric Dinet[†], Bernard Laget^{†§}*

**CPE Lyon - Laboratoire LISA, Villeurbanne Cedex, France*

†Institut d'Ingénierie de la Vision, Saint-Etienne Cedex 1, France

‡ISEM, Maison des Technologies, Toulon, France

Abstract

Everyone agrees that the human visual system is able to perform complex tasks apparently without the slightest effort. The sharpness of sight of the human visual system is closely linked to the number and the density of visual cells. The greatest visual acuity is provided by the fovea which is a small retinal area located in the axis of the human eye pupil. Beyond this particular area, the acuity falls roughly in inverse proportion to the visual angle.

A simulation of foveal mechanisms can be derived from a space-variant image sampling. In that way, the main problem to solve is to find out a relevant model for a retinal cell description. The whole field of view is covered by varying sized disks, in order to modelize the retinal cells. The basic idea of our method is to modify the resolution of the analysed image starting from a given point. This starting point or focusing point corresponds to the center of a focusing area, where the vision system focuses its resources to a detailed analysis.

Introduction

Computer vision systems are confronted with prodigious amounts of visual information. To work in real time, a computer vision system must analyze just that information of interest to the current visual task. Such a selective analysis of a scene is continually used in human vision. A digital image captured by a camera or by almost all two-dimensional analog-to-digital conversion devices conveys information about depth, color, pattern, etc. Generally such an image contains far more data than can be analyzed by a computer vision system in real time. The idea of directing computational resources to locations where they are mostly required is obviously based on the behavior of living organisms. While most current computer vision applications work with a constant spatial resolution, biological systems and more precisely visual systems of primates are based on a space-variant architecture.^{1,2} Such an architecture allows the visual system to focus its resources only on the information of interest while the remaining information is analyzed with a lower precision. We have access to visual data which represent a field of view of approximately 200°. We possess movable space-variant visual sensors (our eyes) which com-

bine a high-resolution central fovea with decreasing resolution in the periphery. Thus high-resolution processing is applied only where necessary. Since most data available in an image is irrelevant to a given visual task, it seems then useful to realize a selective filtering. This perspective implies a remarkable reduction of data and computational resources.

Attention Mechanisms

The human visual system makes extensive use of multiple resolutions. Visual acuity is greatest at the center of gaze (the fovea) and it falls monotonically but rapidly with increasing visual angle into peripheral vision. Another characteristic of human foveated vision is that the visual system gathers partial information about the surrounding environment by moving the focus-of-attention and then determining the next step for the specific task. To gather detailed information about the visual world, a human observer has to move his fovea. Foveal vision is then generally said to be associated with visual attention while the role of peripheral vision is to guide gaze movements. Indeed, to deal with our environment, we must be able to detect in the tremendous amount of visual information the most important elements for the current task to grant a sufficient attention. The rapidity of detection of relevant information is often crucial. It is for example the case when driving a car. In such a context, a danger or an unexpected event can appear anywhere. The observer does not know in advance where to look in his environment.

There are clear advantages when using a space-variant system. It can provide a compromise between two conflicting requirements: a wide field of view and a relatively high resolution. Actually, the ratio of the sample points needed for a sensor with space-variance (roughly equivalent to the human retina) to the sample points needed for providing uniformly high resolution is approximately 1:1000-10,000. While a single image is clearly represented in an « iconic » (though spatially transformed) manner on the retina and in the primary visual cortex. The image of a scene is mainly formed on the outside of the fovea since the fovea corresponds to a small area of the retina (roughly 1°).

Radial Sampling

The human monocular field of view is 208° .³ The spatial sampling that covers relatively wide angles of view is not well defined, since the visual system cannot be readily described in terms of pixels. Foveal techniques seem powerful in computer vision just as they are in human vision. To simulate the resolving power of the human visual system in the machine vision context, we define a computational model of image sampling. This model reproduces the two main retinal areas: a foveal area with constant sampling rate and a peripheral area where the sampling rate decreases roughly in function of the inverse of the visual angle.

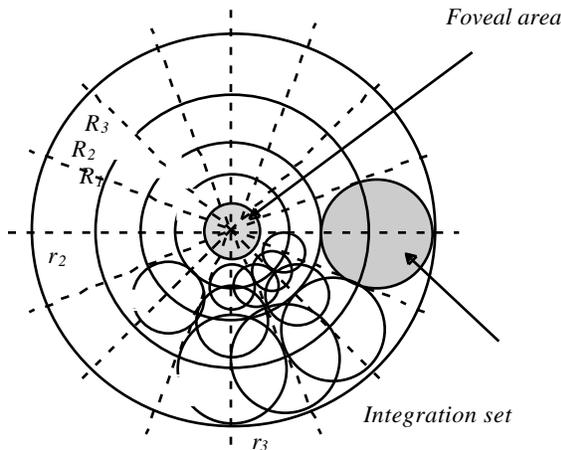


Figure 1. Sampling scheme based on the resolving power of the human visual system.

We are going to follow the sampling scheme of Figure 1 to describe these two retinal areas. From the focusing point, center of the sampling rings and of the fovea, to the fovea boundary, the resolution will be unchanged. Then, outside the fovea, the resolution will be decreased as a linear function of eccentricity. This can be achieved by determining a sampling set made of centers of disks (integration sets). These disks are assumed to be circular sensors of increasing radii r_i . This way of sampling enables to define a ring set. Each ring includes the same number of disks with equal radii. The disk radius increases according to the ring radius (radial distance to the focusing point).

The sensors (that we also call integration sets) of our sampling scheme are spatially distributed in a similar fashion to Yamamoto, Yeshurun and Levine.³ Centers of disks are uniformly spaced on each ring according to an overlap factor v_θ . The radius R_l of the l th ring depends also on an overlap factor denoted by v_r . This overlap factor corresponds to the ratio of the diameter $2r_i$ of an integration set to the radius R_l of the ring where its center is located. The radius of the l th ring R_l can be expressed by

$$R_l = R_0 \left(1 + \frac{2v_r(1 - v_\theta)}{2 - v_r(1 - v_\theta)} \right)$$

The radius of the integration set on the l th ring is

$$r_l = \frac{v_r R_l}{2}$$

Each integration set delimits some points of the initial image under study. These points will be gathered to merge the visual information and reduce the resolution in order to obtain an encoded image. The overall result of this process is a space-variant filtering.

Image Encoding by Radial Sampling

The encoding process enables to set up a way of gathering information on each disk previously defined. The plane is cut up into a finite number of equal angular sectors, whose common summit is the focusing point. An increasing sequence of radii is computed and, on each ring centered at the focusing point, a disk per sector is considered, but the sampling step for the sequence of radii is not constant.

Then the sampled image can be encoded as follows: on the first line the values corresponding to the integration sets of the first ring (whose radius R_l is the minimum of the sequence) represent the average values of the sensor areas. On the second line there are the values of second ring disk centers and so on. This process enables to represent information to be shown out. Though the density of the observed scene is regular in both axis directions, it yields that this sampling scheme gives an homogeneous radial density. The processed image is then a developed image of the original one around the focusing point. On the encoded image appears radial information integrated on each part of the sectors. These parts have the original feature of being equal-weighted according to the distance to the focusing point. Figure 2 shows the initial image under study in the paper.

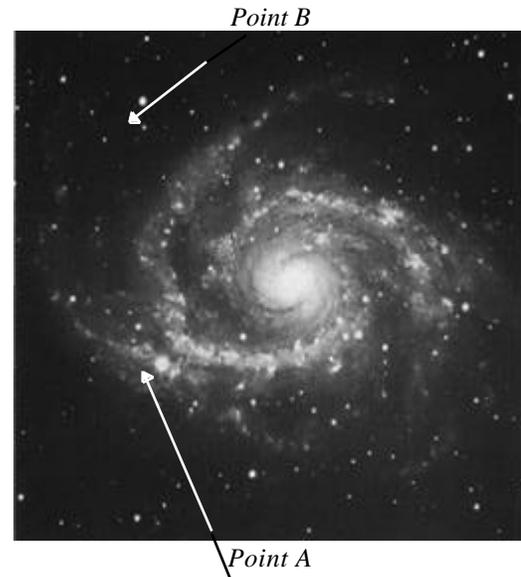


Figure 2. Original image and two different focusing points.

Image Encoding by Average Integration

Gathering information by average integration means computing the average value on each integration set. Then, this value is associated with the disk center.

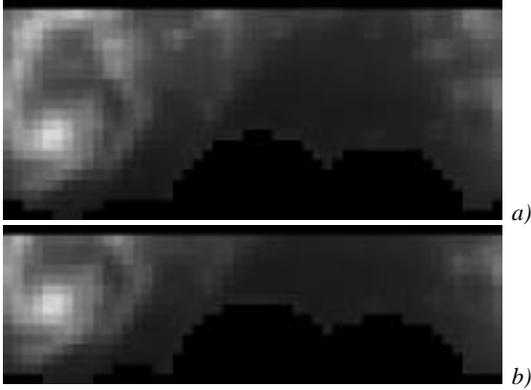


Figure 3. Encoded image by average sampling for a foveal radius of a) $R_0 = 15$ and b) $R_0 = 30$ with A as focusing point.

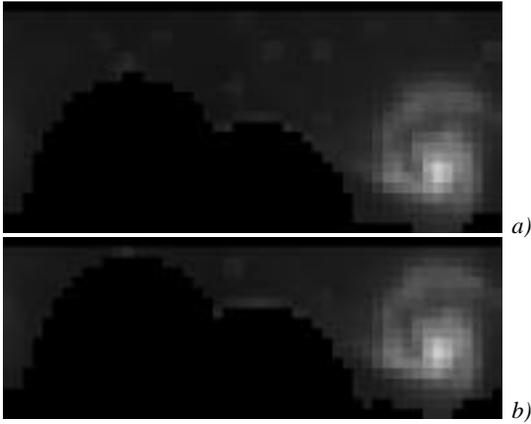


Figure 4. Encoded image by average sampling for a foveal radius of a) $R_0 = 15$ and b) $R_0 = 30$ with B as focusing point.

Figure 3 and 4 show the average values of integration sets for two different foveal diameters and two different focusing points. The formula giving the average value $v(d)$ on a disk d follows. Let $v(M)$ be the value at a point M and $card(d)$ be the number of points belonging to d :

$$v(d) = \frac{1}{card(d)} \sum_{M \in d} v(M)$$

Image Encoding by Median Filtering Integration

In order to determine the value computed by median filtering, let us consider the value set on each integration set. Sorting these values by increasing or decreasing order enables the extraction of the median value $v(d)$, associated with the disk d and given by the following formula. Let M_i be the sequence of points belonging to a disk d such as:

$$M_1, \dots, M_i, \dots, M_{card(d)} \in d$$

$$v(M_1) \leq \dots \leq v(M_i) \leq \dots \leq v(M_{card(d)})$$

$$v(d) = v\left(\frac{M_{card(d)}}{2}\right)$$

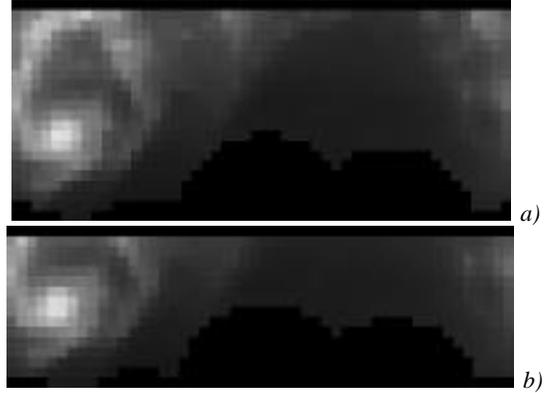


Figure 5. Encoded image by median filtering sampling for a foveal radius of a) $R_0=15$ and b) $R_0=30$ with A as focusing point.

Figure 5 and 6 show the encoded images from the original one by median filtering integration. The results are quite similar to the previous ones obtained by average computation. In other words, a loss of information more or less important can hardly be evaluated when comparing the two methods. Let us see how the reconstruction process works on these images.

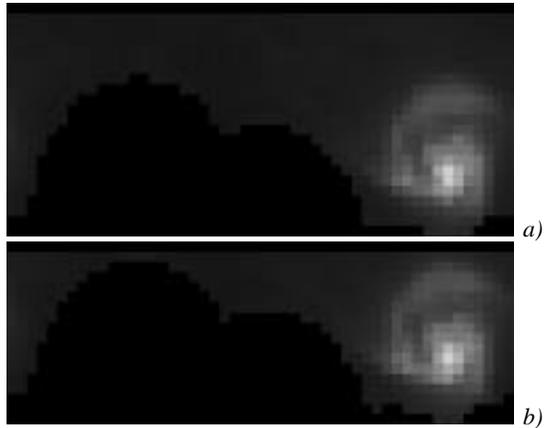


Figure 6. Encoded image by median filtering sampling for a foveal radius of a) $R_0 = 15$ and b) $R_0 = 30$ with B as focusing point.

Image Reconstruction from an Encoded Image

Previously, we have computed an encoded image corresponding to a radial-sampled image of the original one. We are now going to define the reverse process. In other words, the main problem to solve is to determine a way to obtain a square-sampled image from the radial-sampled image.

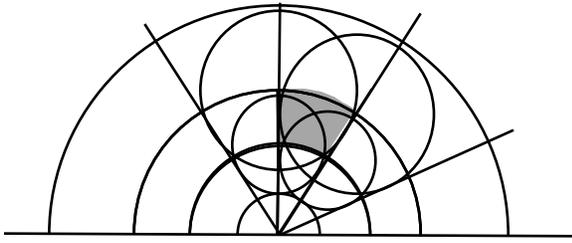


Figure 7. Radial neighborhood of a point M belonging to a given part of the plane (in gray). This part is defined by an angle (angular sector of sampling) and a ring. A point M of the gray region can belong to 1 to 4 disks.

According to the previous sampling, a point of the square grid can belong to 1, 2, 3 or 4 circular cells (see Figure 7). These few particular cells can be found comparing the point coordinates with the respective covered areas. Using complex coordinates is a good way to determine which part of angular sector and which ring the point belongs to, because the modulus of the point gives immediately the ring and its argument the angular sector (see Figure 8).

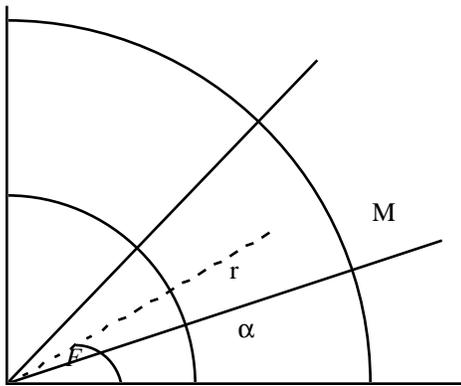


Figure 8. Position of a point M according to its complex coordinate z . The part of the angular sector to which a point M belongs is given by its modulus r (giving the ring) and its argument α with $z = e^{i\alpha}$.

It is then easy to find out which of the four disks determining the radial neighborhood may contain the point under study. As a point of the radial-sampled image corresponds to a full disk of the square-sampled image and vice versa, it is easy to find the gray-level values necessary for the reconstruction. They make out a square neighborhood on the encoded image, as shown in Figure 9.

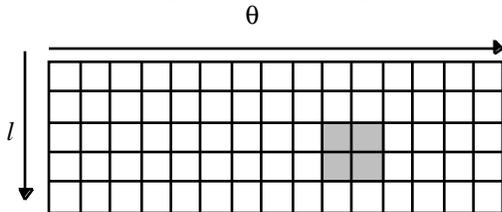


Figure 9. Location of the four considered disks in Figure 7 on the encoded image.

Then, the value of the point on the grid will be computed from the values associated with the previously selected cells. The way of computing this value depends on the sampling method used for the encoding step.

Reconstruction from an Encoded Image by Average Sampling

The reverse process for the average sampling is an average reconstruction. In other words, the value v corresponding to the point M on the grid (reconstructed image) is computed as the average value of those on the selected cells. Let D be the set of disks containing M and $v(d)$ the value on disk d .

$$v = \frac{1}{\text{card}(D)} \sum_{d \in D} v(d)$$

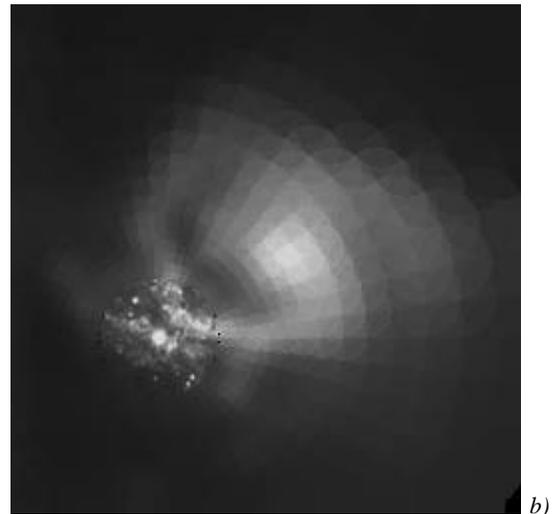
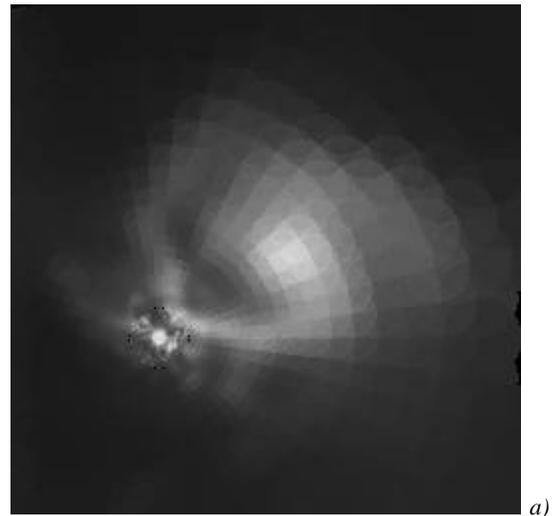


Figure 10. Reconstructed images from encoded images of Fig. 3.

Figure 10 and 11 show the reconstructed images with a foveal radius of a) 15 and b) 30. We can notice that some artifacts due to the sampling disks clearly appear on these images. But global information is represented because the

spiral can be well located though it is less luminous as it was in the original image.

When considering point A as the focusing point, the spiral can be easily located as it is close enough to the foveal area. But information is weakened enough not to prevent the star at the focusing point from being brighter and more accurately represented in the reconstructed image.

In the following images (Figure 11) with B as focusing point, the spiral can hardly be distinguished. That enables to point out local information all around the star under study, giving less importance to points too far from the focusing point.

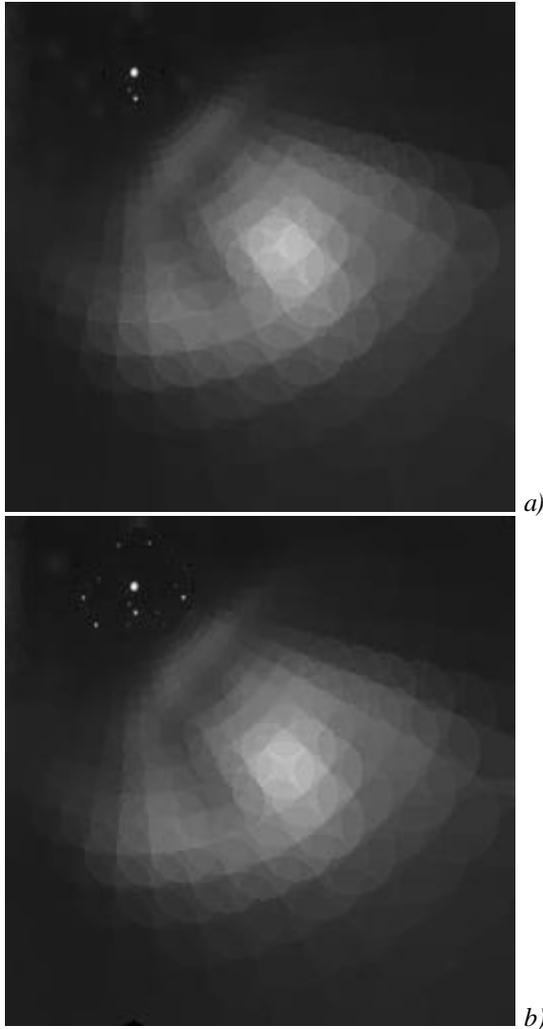


Figure 11. Reconstructed images from encoded images of Fig. 4.

Reconstruction From an Encoded Image by Median Filtering Sampling

The reverse process for the median filtering sampling is a median filtering reconstruction. In other words, the value v corresponding to the point M on the reconstructed images is computed as the median value out of those of the selected disks.

Let D be the set of disks d_i containing M and $v(d_i)$ the value on disk d_i . The disk values $v(d_i)$ are sorted in an increasing sequence and:

$$d_1, \dots, d_i, \dots, d_{card(D)} \in D$$

$$v(d_1) \leq \dots \leq v(d_i) \leq \dots \leq v(d_{card(D)})$$

$$v = v\left(d \frac{card(D)}{2}\right)$$

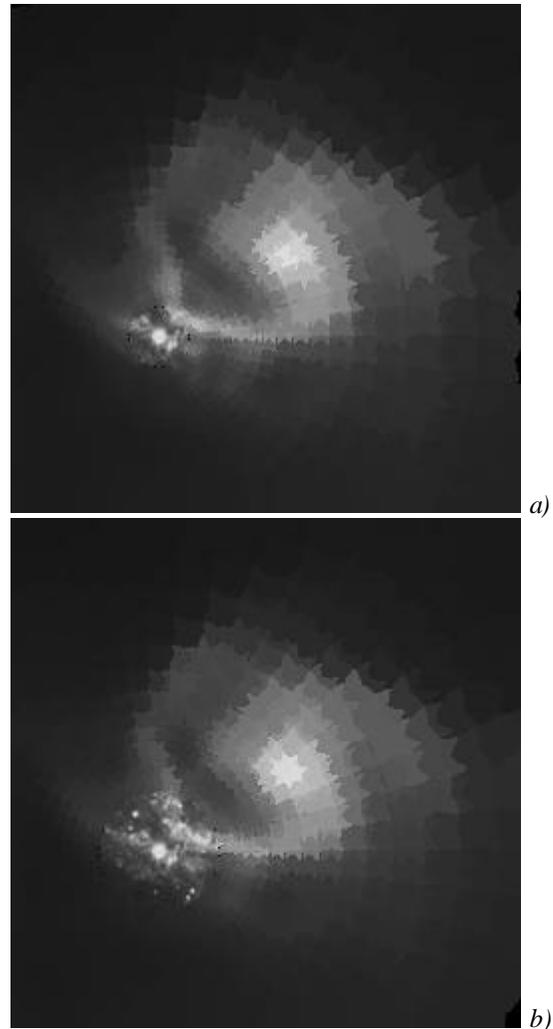


Figure 12. Reconstructed images from encoded images of Fig. 5.

Figure 12 and 13 show the reconstructed images with a foveal radius of a) 15 and b) 30.

The conclusion concerning this new set of images obtained from the median filtering process is quite the same as for the average process, though these last images appear noisier than the previous ones.

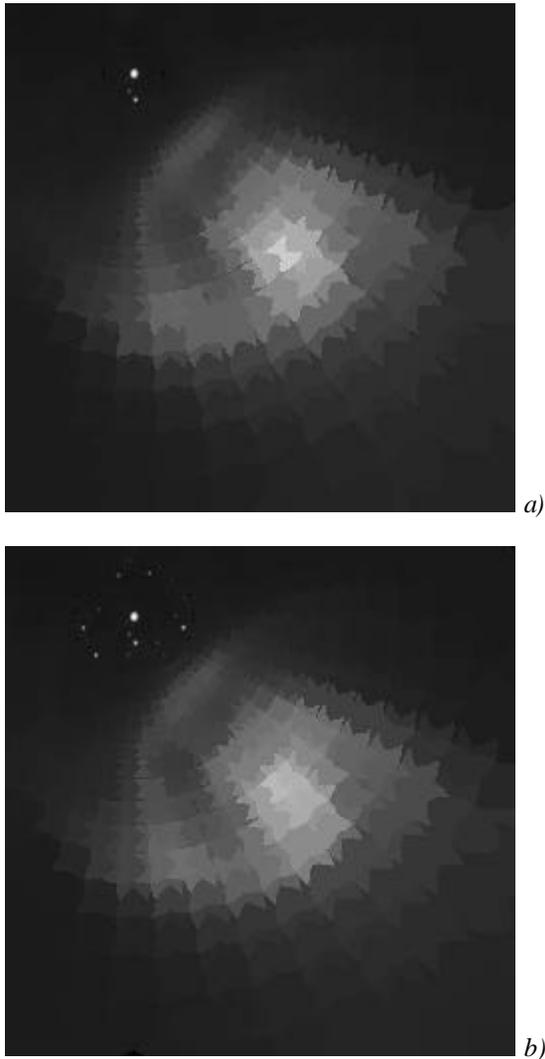


Figure 13. Reconstructed images from encoded images of Fig. 6.

Conclusion

Comparing the previous methods for a given radius of the fovea shows that on the resulting images, some circular shapes are visible for the average method, though it is noisier for the median filtering process. They reproduce the circular cells used for radial sampling. This space-variant filtering method is a good way to gather information all over an image, apart from in the foveal area, according to the radial distance to the focusing point.

While visual attention is a major research topic in psychology, neurobiology,⁴ and the computational aspects of vision, most of the work in this area relates to covert attention. This term refers to a situation where the gaze is fixed on a single image and the focus of attention moves covertly within that image. This way, all the interesting spots can be studied separately without any interference.

References

- 1.EL. Schwartz, *Vision Research*, **20**, (1980), pp 645-69.
 - 2.BA. Wandell, *Foundations of vision*, Sunderland, Sinauer Associates, (1995).
 - 3.H. Yamamoto, Y. Yeshurun, MD. Levine, *Computer Vision and Image Understanding*, **63**, (1996), pp 50-65., CL. Colby, *Child Neurol.*, **6**, (1991), pp S90-S118.
- § IS&T Member