THE RGF PANDEMION: A LOW-LEVEL REPRESENTATIONAL MODEL FOR IMAGES

ROSA RODRIGUEZ-SANCHEZ,† J. A. GARCÍA*‡ and J. FDEZ-VALDIVIA‡
† Departamento de Informática, Escuela Politécnica Superior, Universidad de Jaen, 23071 Jaen, Spain
‡ Departamento de Ciencias de la Computación e I.A., E.T.S. de Ingeniería Informática,
Universidad de Granada, 18071 Granada, Spain

(Received 14 August 1997; in revised form 9 February 1998)

Abstract—In this work, we have reformulated the selection of a few responding units tuned to patterns of activity from the Fourier spectrum as a multilayer system composed of four stages of different kinds of feature detectors. The system output is a sparse representation for images that produces an organization according to object power discrimination. The resultant multilayer system provides a more efficient representation for later stages of processing because it possesses a higher degree of perceptual independence among its outputs.

© 1998 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Object power discrimination Responding units
Object detection Clutter

1. INTRODUCTION

The performance of the high-level mechanisms of a visual system (e.g. decision processes) depends on the signal-to-noise ratio (or the information content) of the outputs of the automatic low-level mechanisms of the visual system, and, to the extent that high-level mechanisms can be characterized as dependent on signal-to-noise ratio of the low-level outputs then the better the chances of developing a general computational model of visual processing consisting of a near optimal decision rule applied to the outputs of independent, automatic, perceptual mechanisms.

In this paper, we propose a multilayer system providing a more complete description of where perceptually independent features are present. To produce the proposed multilayer system, we have reformulated an automatic mechanism to select out the strongly responding sensors from a data-driven multichannel partition in the way of a pandemonium (i.e. a system such that its operation is dependent upon features but not upon templates). The system is composed of four stages of different kinds of feature detectors:

Stage 1: A spatial to 2-D spatial-frequency transformation capable of achieving the enhancement of the activity patterns in the Fourier spectrum. This will be a process of completion by means of a nonlinear enhancement and inhibition of stray and weak signals from the detectors in process.

Stage 2: A data-driven multichannel design, based on a function that returns a first-order statistics indicative of the character of each particular sector of the spectrum from level 1.

Stage 3: The selection of the most activated sensors of the second-tier detectors.

Stage 4: The grouping of the sensors from level 3 into a few complex sensors in order to maximize the signal to clutter ratio of sensor responses.

In the following, the technical details of the multilayer system are presented. Sections 2–5 describe each stage of the system. A comprehensive analysis of the model is presented using a set of experiments and discussion in Section 6. Finally, the main conclusions of this work are summarized in Section 7.

2. STAGE 1: A SPATIAL TO 2-D SPATIAL-FREQUENCY TRANSFORMATION

The objective of this first stage is twofold: first, to compute the Fourier spectrum of the input image; second, the patterns of activity spread across the Fourier spectrum, with values in the range of interest—computed as the dynamic range of a logarithm transformation on Fourier spectra—are to be contrasted with a transformation used for histogram stretching, so that the result is more suitable than the original for the subsequent stages.

The magnitude function |F(ϕ, θ)|, with F(ϕ, θ) being the Fourier transform in polar coordinates of the input image, is called the Fourier spectrum of the image. It has values in a large dynamic range, but the activity patterns have values in a specific range of interest, so when Fourier spectra is scaled linearly for processing to 8-bit, the highest values dominate the activity patterns, as expected for such a large dynamic
range. So enhancing this range of interest is desired. To this aim, two subsequent transformations on the spectrum are to be performed.

First, to extract an estimate of the range of interest, a logarithm transformation function performs a compression of the dynamic range of values, so \( |F(\rho, \theta)| \) is transformed to the function

\[
c \log(1 + |F(\rho, \theta)|)
\]

where \( c \) is a scaling constant selected to normalize the values between 0 and 255 for each scene. The range of interest is computed as the resultant range of the logarithm transformation function.

Second, to enhance the resultant range of interest, a histogram stretching\(^1\) on the result of the transformation in equation (1) is performed that produces the spread in the values of the compressed range so affecting its contrast. The histogram stretch here adopted occurs over a range defined by integrating in from each end of the histogram until 1 percent of the cells have been found. This prevents straggler channels from ruining the contrast. The histogram within the 1 percent points is given a linear stretch mapping onto the full contrast range. All this process is illustrated on a synthetic image in Fig. 2 (see Section 7 for further details). As a result of this level, the activity patterns in the Fourier spectrum are to be more easily isolated in specific sensors of a multichannel partition.

3. STAGE 2: A DATA-DRIVEN MULTICHANNEL DESIGN

The selection of an appropriate set of responding units, is a central issue in multi-channel approaches. Due to conjugate symmetry, the sensor design is only carried out on half of the 2-D frequency plane. In general, on the 2-D spatial-frequency plane, the multichannel approaches\(^2\) produce the desired organization by superposition of a fixed number of spatial-frequency channels empirically tuned within each orientation band, which following the biological evidence that demonstrates that the median orientation bandwidth of visual cortex cells is about 40° to 60°, are selected with fixed orientations (i.e., 0°, 45°, 90°, and 135°, with an orientation bandwidth of 45°). However, the problems with such a priori definitions of the decomposition rule are that: (1) they may not actually reflect the underlying structures for images under analysis, and (2) their lack of adaptability may well bias any posterior processing.

The second level of the RGF pandemonium is intended to perform a partitioning of the spectrum from level 1 into a set of data-driven band-limited regions of (radial) spatial frequency and orientation. The result is a set of units (hereafter named as simple sensors) where features triggering them are to be patterns of activity from the receptors in level 1 rather than the geometric features of the stimulus form.

3.1. The data-driven selection of bands of orientation

The Fourier spectrum from level 1, noted as \( R(\rho, \theta) \), is firstly partitioned into a number of bands of orientation (Fig. 3). This is accomplished by using a function that returns the deviation to mean ratio of an orientation band \( (0, \theta) \) in \( R(\rho, \theta) \). This function, noted as \( \eta(\theta) \), returns a first-order statistic providing a quantitative shape description of the orientation band between angles of 0 and \( \theta \) degrees:

\[
\eta(\theta) = \frac{\sigma_\theta}{\mu_\theta},
\]

where \( \sigma_\theta \) and \( \mu_\theta \) represent, respectively, the standard deviation and the mean of the spectrum distribution over the orientation band in \( R \) between angles of 0 and \( \theta \).

Let \( \eta''(\theta) \) be the second derivative of \( \eta \) computed as

\[
\eta''(\theta) = \eta(\theta) \ast \frac{d^2}{d\theta^2} G_s(\theta),
\]

where \( \eta \) is convolved with the second derivative of the Gaussian at scale \( s \) noted as \( (d^2/d\theta^2) G_s(\theta) \), to both smooth and differentiate the function. The scale \( s \) for the Gaussian is derived as described in (4).

An activity pattern in \( R \) produces a meaningful discontinuity in value of deviation to mean ratio of the orientation band \( (0, \theta) \) as \( \theta \) increases. This is illustrated in Fig. 4(a). On the other hand, the second derivative \( \eta'' \) has a zero crossing at the midpoint of a transition in value of \( \eta(\theta) \) as \( \theta \) increases (see Fig. 4(b)). So zero crossings provide a powerful approach for locating activity patterns in the Fourier spectrum which we assume that produce the meaningful discontinuities in \( \eta \).

Once a pattern of activity has been detected via a zero crossing, in order to enclose it, a new band of orientation is to be produced [see Fig. 4(c)]. This is performed taking into account that [as shown in Fig. 4(b)]: (i) the extremes of \( \eta'' \) correspond to the locations of change in the rate of variation in \( \eta(\theta) \); and (ii) the second derivative has one local extreme at location \( \theta_j \), for the start of the discontinuity signaled by the zero crossing, and another local extreme (of opposite sign) at \( \theta_j \) for the end of the discontinuity. Hence, the band of orientation isolating a pattern, noted as \( O(\theta_i, \theta_j) \), can be computed through the pair of local extremes of the second-derivative \( \eta'' \) enclosing the respective zero crossing.

3.2. The data-driven selection of radial frequency channels

To obtain the sensors combining the spatial frequency selectivity and the orientation selectivity, the resultant bands of orientation are independently partitioned into a number of channels of (radial) spatial frequency.

For each orientation band \( O(\theta_i, \theta_j) \) created earlier, a number of radial frequency channels can be created
using a function that returns the standard deviation to mean ratio of the radial frequency channel \((0, \rho)\), within the band \(O(\theta_l, \theta_r)\), as \(\rho\) increases [see Fig. 5(a)]. This function, noted as \(\eta_{i,j}(\rho)\), returns a first-order statistic providing a quantitative shape description of the sector in \(O(\theta_l, \theta_r)\) between radial frequencies of 0 and \(\rho\):

\[
\eta_{i,j}(\rho) = \frac{\sigma_{i,j}}{\mu_{i,j}},
\]

(4)

where \(\sigma_{i,j}\) and \(\mu_{i,j}\) represent, respectively, the standard deviation and the mean of the spectrum distribution over the band given by \((0, \rho)\) in radial frequency, and \((\theta_l, \theta_r)\) in orientation.

Let \(\eta_{i,j}'(\rho)\) be the second derivative of \(\eta_{i,j}\) computed as in equation (3). The process to detect and extract the radial frequency channels at each band of orientation is similar to that used for obtaining the orientation bands. So, in order to detect the locations \(\rho\) at which \(\eta_{i,j}\) undergoes a change in the rate of increment or decrement, a technique based on the extrema of the second derivative, \(\eta_{i,j}''\), will be used as well. The extremes of the second derivative, as shown in Fig. 5(b), correspond to positions marking a change in the rate of the increase or decrease in \(\eta_{i,j}\) and so, they are positions dividing the orientation band \(O(\theta_l, \theta_r)\) into

![Diagrams showing steps of the process](image)

**Fig. 1.** The three examples of images with 256 gray levels: (a) and (b) are synthetic scenes \((128 \times 128)\) containing white stripes with different orientations; (c) a biomedical scene \((256 \times 256)\).

**Fig. 2.** (a) the input image; (b) the Fourier spectrum of the input image; (c) result of a logarithm transformation on Fourier spectra; (d) result of a transformation for histogram stretching.
a number of sequences (the desired sensors) which isolate the activity patterns. The result is shown in Fig. 5(c).


Once the partition of the Fourier spectrum into a number of sensors, noted as \( S_n \), has been carried out, the significance of the sensor response is analyzed by classifying sensors into two classes: the active sensors and the non-active ones. The only sensors worth noting regarding the image representation, are those that exhibit a strong response from the significant structures in the image, namely the active sensors. In fact, with relatively few sensors we may still obtain excellent extraction of parameters of activity profiles in the images. Furthermore, if there is little time or processing capacity available, it is possible to settle for a rough extraction of structure parameters localizing the major structures by picking the most actively responding sensors.

Each sensor should be described by a sensor measure (feature) that can successfully characterize it. Here, we propose one feature derived from the mean of the spectra distribution over the sensor. Of course other measures intended to capture relevant characteristics of the sensor spectra are conceivable: measures such as location, size and orientation of peaks and entropy of the Fourier spectrum in sensors. To evaluate these frequency domain features according to their ability to discriminate active sensors, a method of successive selection and deletion based on Wilks criterion may be used.\(^{(5)}\) Finally, we have
Fig. 4. (a) For image in Fig. 2(a), deviation to mean ratio of the orientation band $(0, \theta)$ as $\theta$ increases; (b) the second derivative $\eta''$ has a zero crossing at the midpoint of a transition in value of $\eta(\theta)$ as $\theta$ increases; (c) a new band of orientation is to be produced taking into account that, as shown in (b), the second derivative has one local extreme at location $\theta$, for the start of the discontinuity signaled by the zero crossing, and another local extreme (of opposite sign) at $\theta_j$ for the end of the discontinuity.
Fig. 5. (a) for the orientation band $O(\theta_1, \theta_2)$ in Fig. 4(c), the function $\eta_{i,j}$ returns the standard deviation to mean ratio computed for the radial frequency channel $(0, \rho)$ as $\rho$ increases within the band $O(\theta_1, \theta_2)$; (b) the extremes of the second derivative; (c) they are positions dividing the orientation band $O(\theta_1, \theta_2)$ into a number of sequences (the desired sensors) which enclose the desired activity patterns.
found that the mean of the spectrum distribution over the sensor provides an effective feature for discriminating the set of sensors on training sets.

In the formulation proposed, cluster analysis is then used to group sensors together, since unsupervised learning may exploit the statistical regularities of the sensors by using the available sensor responses. We determine the two classes of sensors by the square-error clustering method which is implemented here as the k-means algorithm.

5. STAGE 4: THE GROUPING OF THE MOST ACTIVATED SENSORS INTO A NUMBER OF COMPLEX SENSORS

The four stage of the RGF pandemonium is intended to perform the grouping of the sensors from level 3 into a few complex sensors that maximize the signal to clutter ratio of the respective response images—computed by multiplying the DFT of the input image by each 2-D ideal complex-sensor filter, and computing the inverse DFT.

In this approach, the perceived contrast of a feature on a complex background is represented by a quantitative measure for signal-to-clutter ratio (SCR) that effectively combines (a) the contrast of the feature with (b) the number of significant edge points as a measure of image clutter—clutter can be viewed as localized effects in an image that tend to make accurate discrimination of objects from their background harder. For each sensor $S_i$, the signal to clutter ratio of the respective response image, noted as $I(S_i)$, is here defined as

$$SCR[I(S_i)] = \frac{\text{Contrast}[I(S_i)]}{\text{POE}[I(S_i)]},$$

where $\text{Contrast}[I(S_i)]$ measures the contrast of the signal computed as

$$\text{Contrast}[I(S_i)] = \frac{\text{magn}(\nabla I(S_i))}{\text{mean response by magn}(\nabla I(S_i))}$$

with $\text{magn}(\nabla I(S_i))$ being the magnitude of the gradient of $I(S_i)$; and where $\text{POE}[I(S_i)]$ measures the image clutter as the probability of edge. The POE metric assumes that the number of edge points drives the search or fixation process and is then calculated as follows:

$$\text{POE}[I(S_i)] = \frac{1}{N} \sum_{i=1}^{N} \text{POE}_i^2[I(S_i)],$$

where we divide $I(S_i)$ into $N$ blocks of $8 \times 8$ pixels in size, and an edge detector is applied to each block to compute the number of edges in a block, noted as $\text{POE}_i[I(S_i)]$. The motivation for this approach is that preattentive vision is known to be drawn to edges.

In order to group the activated sensors from level 3, noted as $\{S_1, \ldots, S_n\}$, into a few complex sensors maximizing the SCR of the respective response images, a clustering procedure is used, and each cluster
of the resultant partition will determine a complex sensor.

A similarity measure for a pair of sensors is defined as
\[ d(S_i, S_j) = \text{SCR}[I(S_i + S_j)], \]  
(6)

where \( I(S_i + S_j) \) is the response image of the complex sensor that the grouping of \( S_i \) and \( S_j \) produces.

Simple clustering is done based on this similarity measure. First, the maximum similarity between any two sensors is found as
\[ d(S_{max}, S_{max}) = \max_{i} d(S_i, S_i); \]

with \( d(S_i, S_j) \leq \max(S_{\text{SCR}}[I(S_i)], S_{\text{SCR}}[I(S_j)]) \).  
(7)

Then the pair of sensors \( S_i, S_j \) computed as given in equation (7) at one level are to be fused to one at the next hierarchical level.

The clustering procedure progresses until no further joining can be found, what happens if at a stage there is no pair of sensors \( S_i, S_j \) verifying the equation (7). The stopping rule suggested here, it is objective in the sense that it assumes no prior knowledge about the number of complex sensors to be found.

Finally, the stage 4 of the RGF pandemonium produces the system outputs in the shape of the response images of the complex sensors, so deriving a sparse representation of the input image as well as increasing the object power discrimination.

6. EXPERIMENTAL RESULTS

To examine the performance of the multilayer system, we used three examples of images with 256 gray levels: the first two images, as shown in Figs. 1(a, b), are synthetic scenes (128 \( \times \) 128) containing white stripes with different orientations; the third image is
a biomedical scene (256 × 256) as shown in Fig. 1(c). The software implementing the proposed model is available by anonymous ftp from decesai.ugr.es in the tar file pub/diata/software/rgf-punde.tar.gz. In the following, the results of the proposed model are described on the incoming images.

To see the performance of the first stage of the model, we present its result on a synthetic image.
Fig. 9. (a) The incoming image; (b) its most activated sensors from level 3; (c)–(m) the complex sensors from level 4.

Fig. 2 is used to illustrate its different steps. First, the Fourier spectrum of the input image is computed; second, the range of interest is estimated as the dynamic range of a logarithm transformation on Fourier spectra; third, the compressed spectra is contrasted by using a transformation for histogram stretching. A last point worth noting is that to highlight the activity patterns in the resultant spectra, a simple variation of this process may also be considered (as shown in Fig. 3): the modified spectra may
be multiplied by the result of thresholding this image. This additional transformation preserves the patterns of activity in the modified spectra while diminishing all others to a constant, low level.

The second stage of the RGF pandemonium is illustrated in Figs 4 and 5. In Fig. 5 is illustrated the process used for obtaining the sensors combining the spatial frequency selectivity and the orientation selectivity.

In Fig. 6, the output of the stages 2 and 3 are presented on two incoming images which are given in the left column. The middle column shows the respective multichannel partition as derived from level 2. The second-level detectors are analogous to a data-driven collection of oriented spatial-frequency channels. In the same figure, the right column shows the only activated sensors as derived from level 3. By simply viewing the data, it can be determined that the most active of the second-tier units were detected so bringing out the important perceptual features in the context of the image understanding.

The output of the four stage of the RGF pandemonium is illustrated in Figs 7(c–i) and 9(c–m) on the same incoming images as above. In this level, the Canny edge detector was used to compute both the gradient of the response images and the number of edges in each image block [see equation (5) for further details]. The incoming image and its most activated sensors from level 3 are shown, respectively, in Figs 7 and 9(a,b). In order to improve subsequent processing, the activated sensors $S_i$ with a value of $SCR[I(S_i)]$ less than 5 per cent of the mean $SCR$ over the set of activated sensors, can be ignored. The complex sensors from level 4 are presented in the other images. From these figures, it can be noted that, for each white stripe in the input image, there is only one distinct complex sensor isolating its respective high-, medium- and low-frequency components. The output system in the shape of the response images, is presented, respectively, in Figs 8 and 10. In each figure, the left column shows the different complex sensors (in these images, the center of the frequency spectrum is also plotted); and the middle column presents the respective response images. To compute them, the DFT of the input image is multiplied by an ideal filter such that all frequencies inside the complex sensor in addition to the center of the spectrum are passed with no attenuation, whereas all frequencies outside are completely attenuated, and then computed the inverse DFT. For each white stripe in the original image, there is only a response image that isolates such a spatial structure. In the same Figs 8 and 10, the right column shows the edge maps computed on the response images through the Canny edge detector. Again, for each white stripe in the original image, the respective illusory contour is also present in only an edge image. A point worth noting is that this study is by nature exploratory, and so we did not try to
optimize the performance of the system (there are several ways of computing the response images: the system might act as either a spatial-frequency analyzer, or between the spatial sampling and the Fourier transformation). Further studies are needed to examine whether the representation of images can be improved by deriving response images through 2-D Gabor filtering.8

Finally, the output of each level of the system on a biomedical image is illustrated in Figs 11 and 12. The original image is given in Fig. 11(a). The output from level 1, 2 and 3 is presented, respectively, in Fig. 11(b)–(d). A sample of the complex sensors from level 4 and their respective response images is given in Fig. 12. The same figure also shows the corresponding edge maps in the right column.

7. CONCLUSIONS

In absence of further information, the proposed system code achieves two goals. First, it produces a sparse representation by deriving the minimum number of complex sensors to represent images. So the multilayer system takes advantage of the redundancy of the environment. For a particular image, only a small subset of the population of sensors needs to be active to represent a scene. However, across all scenes,
Fig. 12. A sample of the complex sensors from level 4 as well as the output system (the respective response images). The corresponding edge maps are in the right column.
every sensor is as likely to be achieved as any other. Second, the RGF pandemonium provides us with an organization of the image content according to object power discrimination.

Acknowledgements—This research was sponsored by the Spanish Board for Science and Technology (CICYT) under grant TIC97-1150. The authors would like to thank an anonymous referee for helpful comments.

REFERENCES


About the Author — ROSA RODRIGUEZ-SANCHEZ was born in Granada, Spain. She received the M.S. degree in Computer Science from the University of Granada in 1996. Currently, she is with the Informatics Department at the University of Jaen and working for her Ph.D. degree at the Computer Science Department (DECSAI) at Granada University. Her current interests include image processing and visual perception.

About the Author — J. A. GARCIA was born in Almeria, Spain. He received the M.S. and Ph.D. degrees both in Mathematics from the University of Granada in 1987 and 1992, respectively. Since 1988 he has been with the Computer Science Department (DECSAI) at Granada University where he is now a Lecturer. His current interest is computer vision. Dr Garcia is a member of the IAPR Association and the IEEE Computer Society.

About the Author — J. FDEZ-VALDIVIA was born in Granada, Spain. He received the M.S. and Ph.D. degrees both in Mathematics from the University of Granada in 1986 and 1991, respectively. Since 1988 he has been with the Computer Science Department (DECSAI) at Granada University where he is now a Lecturer. His current interests include image processing, pattern recognition, multiresolution methods, image coding, and biomedical applications. Dr Fdez-Valdivia is a member of the IAPR Association, the IEEE Computer Society, ACM Computer Society and the OSA.