Image Applications of Agglomerative Clustering using Mixtures of non-Gaussian Distributions

Addisson Salazar[1], Jorge Igual[1], Gonzalo Safont[1], Luis Vergara[1], Antonio Vidal[2]
Universitat Politècnica de València
[1]Instituto de Telecomunicaciones y Aplicaciones Multimedia
[2]Departamento de Sistemas Informáticos y Computación
asalazar@dcom.upv.es

***Abstract*— We present two applications in image processing of an agglomerative clustering method based on mixtures of non-Gaussian distributions. The method joins pair-wise the mixture models estimated for every cluster building a pyramidal or hierarchical structure by using the Kullback-Leibler divergence. This process can be related with the feedforward process of abstraction carried out by the brain. The applications consist of grouping images based on their content similarities and segmentation of regions of an image in similar areas. The capability of the method to distinguish between natural and artificial images is also demonstrated.**

Keywords*— *image processing, non-Gaussian distributions, clustering, pattern recognition, ICA, pyramidal structures

I. INTRODUCTION

Image pattern recognition tasks involve process such as noise filtering and border detection of the incorporated objects in an image. This is carried out by humans using visual capabilities of the brain. The visual stimulus of the image cause depolarization of millions of neurons that activate brain circuits to interpret it, which is commonly related with the cognitive function of learning and memory [1]. The recognition process should be independent of the object size, spatial location, and other variables belonging to the image context where it is embedded. Thus, the final result of the task is the division of an image in several parts that potentially can be connected with some cognitive perception, e.g., "a car on a road trip in a sunny day". Actually, depending on the human health condition and age, the perception and understanding of the image could become quite different (e.g., [2]).

The study of methods for automatic pattern recognition tasks in image processing is a booming field of research which include distinguishing objects with similar shapes and separating different parts with similar characteristics in an image. The challenges are enormous considering the objective of emulating human capabilities as explained above. The importance of such automatic capabilities can be particularly acknowledged in applications of big data where not only the quality of the recognition is important, but also the capabilities for rapid processing of volumes of data. Examples of these applications are the following: imputation of missing functional MRI (Magnetic Resonance Imaging) cardiac images; processing geographic images to determine the current terrain type from onboard mobile sensors; and QBSE (query by semantic example) systems for large-scale histopathological image analysis (see for instance [3][4] and the references within).

Currently, image pattern recognition is applied using statistical operators such as oriented statistical edge detector filters (for improved 3D surface detection) [5]; local polynomial regression (for cell-average multiresolution) [6]; and independent component analysis (ICA) for automatic detection of Parkinsonism in nuclear medicine imaging [7]. In this line of work, the method applied in this paper models the probability density function of the data as a mixture of non-Gaussian distributions in the framework of latent variable modelling (random variable generators that are hidden behind the observable variables). The mixture model consists of groups of multivariate densities, where each one group is modelled by using ICA [8]. ICA is an extension of principal component analysis (PCA), but imposing the condition of statistical independence in the estimated latent variables that, usually, are called sources. The methods for ICA implementation can be roughly classified considering how they approximate the parameters of the model by minimizing a cost function so-called contrast using: non Gaussianity; mutual information; higher order statistics (cumulants); and time structures [8]. Applications of ICA comprise such diverse disciplines such as telecommunications; non-destructive testing; and biomedicine (e.g., [9]).

The degrees of freedom provided by the statistical method applied allow local independence and global dependence relationships of the variables to de modelled, besides of obtaining an automatic partition of the processed image. An important part of our methodology is hierarchical and so related to pyramidal structures in image processing. The ICA decompositions are joined using hierarchical clustering of agglomerative type. It starts from a set of ICA mixture parameters that are extracted from the data using a learning process as explained in [10]. Each cluster at the first level of the hierarchy is characterized with the parameters of a single ICA model. The proximities between clusters are estimated pair-wise using the Kullback-Leibler (KL) divergence [11].

During the combination of the clusters, the entropy and cross-entropy of the sources have to be estimated [12]. This

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.118

cannot be obtained analytically, and thus an iterative suboptimal approach is applied using a numerical approximation from the training data. The structure of several ICA subspaces at the bottom level of the hierarchy allows non-Gaussian mixtures to be modelled. The independence relations between the hidden variables at this lowest level are relaxed at higher levels of the hierarchy allowing more flexible modelling. The subspaces constructed by the method at intermediate levels of the hierarchy represent different degrees of dependence of the variables. In addition, considering recent works, these subspaces might be analyzed as different levels of information fusion (see for instance [13] and the references within).

The next sections are organized as follows: Section II explains the hierarchical classification clustering; Section III includes the image pattern recognition applications (object recognition and image segmentation); and finally Section IV contains the conclusions of this work.

II. HIERARCHICAL ICA MIXTURES

This section explains the formulation of the hierarchical agglomerative algorithm. We define the ICA mixture model ICAMM as follows:

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{s}_k + \mathbf{b}_k, \quad k = 1, \ldots, K \tag{1}$$

For K clusters, where each cluster is defined using an ICA, i.e., with a mixing matrix $\mathbf{A}_k$, an independent vector source $\mathbf{s}_k$ and a bias vector $\mathbf{b}_k$. Essentially, the geometry of the data is explained by the parameters $\mathbf{A}_k\mathbf{s}_k$ and the bias $\mathbf{b}_k$ is used to locate the data in the feature hyperspace.

We suppose that the ICAMM parameters have been estimated as in [10], and summarize the hierarchical procedure that is introduced in [14]. The clusters are joined attending to the symmetric KL divergence between the clusters C_u, C_v as follows:

$$D_{KL}\left(C_u // C_v\right) = \int p\left(\mathbf{x}/C_u\right) \log \frac{p\left(\mathbf{x}/C_u\right)}{p\left(\mathbf{x}/C_v\right)} d\mathbf{x} + \int p\left(\mathbf{x}/C_v\right) \log \frac{p\left(\mathbf{x}/C_v\right)}{p\left(\mathbf{x}/C_u\right)} d\mathbf{x} \tag{2}$$

where each term in (2) is calculated using $p_{x_u}(\mathbf{x}) = p(\mathbf{x}/C_u)$:

$$D_{KL}\left(p_{\mathbf{x}_u}(\mathbf{x}) // p_{\mathbf{x}_v}(\mathbf{x})\right) = -H(\mathbf{x}_u) - H(\mathbf{x}_v) - \int p_{\mathbf{x}_u}(\mathbf{x}) \log p_{\mathbf{x}_v}(\mathbf{x}) d\mathbf{x} - \int p_{\mathbf{x}_v}(\mathbf{x}) \log p_{\mathbf{x}_u}(\mathbf{x}) d\mathbf{x} \tag{3}$$

where $H(\mathbf{x})$ is the entropy, which is estimated as the negative of the logarithm of the probability distribution, i.e., $H(\mathbf{x}) = -E\left[\log p_x(\mathbf{x})\right]$; and the other terms are the cross-entropies $E_{\mathbf{x}_v}\left[\log p_{\mathbf{x}_u}(\mathbf{x})\right]$, $E_{\mathbf{x}_u}\left[\log p_{\mathbf{x}_v}(\mathbf{x})\right]$. Using the independence of the sources in the ICA model, (3) simplifies, but we still have to approximate the densities of the sources since they are unknown. We use non-parametric kernel-based densities:

$$p_{s_u}(s_u) = \sum_{n=1}^{N} a e^{-\frac{1}{2}\left(\frac{s_u - s_u(n)}{\Delta}\right)^2} \tag{4}$$

where $s_u(n)$ is the source s_u at sample n. The final algorithm is (see [14]):

$$D_{KL}\left(p_{\mathbf{x}_u}(\mathbf{x}) // p_{\mathbf{x}_v}(\mathbf{x})\right) = -\sum_{i=1}^{M} \hat{H}\left(s_{u_i}\right) - \sum_{j=1}^{M} \hat{H}\left(s_{v_j}\right) - \sum_{i=1}^{M} \hat{H}\left(\mathbf{s}_v, s_{u_i}\right) - \sum_{j=1}^{M} \hat{H}\left(\mathbf{s}_u, s_{v_j}\right) \tag{5}$$

where M is the number of sources in every ICAMM (the same for every cluster) and the estimates for Q observations of every source:

$$\begin{aligned}
&\hat{H}\left(s_{u_i}\right) = -\frac{1}{Q} \sum_{n=1}^{Q} \log p_{s_{u_i}}\left(s_{u_i}(n)\right), \\
&p_{s_{u_i}}\left(s_{u_i}(n)\right) = \sum_{l=1}^{N} a e^{-\frac{1}{2}\left(\frac{s_{u_i}(n) - s_{u_i}(l)}{\Delta}\right)^2} \\
&\hat{H}\left(\mathbf{s}_v, s_{u_i}\right) = \frac{1}{Q^M} \sum_{s_{v_1}=1}^{Q} \cdots \sum_{s_{v_M}=1}^{Q} \log \sum_{n=1}^{N} a e^{-\frac{1}{2}\left(\frac{\left[\mathbf{A}_u^{-1}\left(\mathbf{A}_v \mathbf{s}_v + \mathbf{b}_v - \mathbf{b}_u\right)\right]_i - s_{u_i}(n)}{\Delta}\right)^2} \\
&\hat{H}\left(\mathbf{s}_u, s_{v_j}\right) = \frac{1}{Q^M} \sum_{s_{u_1}=1}^{Q} \cdots \sum_{s_{u_M}=1}^{Q} \log \sum_{n=1}^{N} a e^{-\frac{1}{2}\left(\frac{\left[\mathbf{A}_v^{-1}\left(\mathbf{A}_u \mathbf{s}_u + \mathbf{b}_u - \mathbf{b}_v\right)\right]_j - s_{v_j}(n)}{\Delta}\right)^2}
\end{aligned} \tag{6}$$

where $\hat{H}\left(s_{v_j}\right)$ in (5) is calculated in the same way as $\hat{H}\left(s_{u_i}\right)$ in (6).

The objective is to build a hierarchy from an initial number of K clusters to only one cluster by combining the closest clusters in successive iterations. The number of hierarchy levels will be K, for instance, given $K = 4$, we will obtain 4 hierarchy levels l, $l = 1, \ldots, 4$ with $4, 3, 2, 1$ number of clusters in each level, respectively. For each hierarchy level, a pairwise calculation of the proximities between clusters is done using KL divergence. The proximity between clusters depends on all the ICA mixture parameters ($\mathbf{A}_k$, $\mathbf{s}_k$, and $\mathbf{b}_k$). Thus, the clusters to be combined at hierarchy level l are selected by using the proximities estimated at the previous hierarchy level $l-1$.

The proximity at level l between a cluster C_z^l to a combined cluster C_w^l obtained from clusters C_u^{l-1} , C_v^{l-1} at level $l-1$ is calculated as follows:

$$D_l\left(p_l\left(\mathbf{x}/C_w^l\right)//p_l\left(\mathbf{x}/C_z^l\right)\right)=\left[\frac{1}{p_{l-1}\left(C_u^{l-1}\right)+p_{l-1}\left(C_v^{l-1}\right)}\right]\cdot$$
$$\left[\begin{array}{l} p_{l-1}\left(C_u^{l-1}\right)\cdot D_{l-1}\left(p_{l-1}\left(\mathbf{x}/C_u^{l-1}\right)//p_{l-1}\left(\mathbf{x}/C_z^{l-1}\right)\right)+ \\ p_{l-1}\left(C_v^{l-1}\right)\cdot D_{l-1}\left(p_{l-1}\left(\mathbf{x}/C_v^{l-1}\right)//p_{l-1}\left(\mathbf{x}/C_z^{l-1}\right)\right) \end{array}\right] \quad (7)$$

The iterative application of (7), combining two clusters at every step, produces a dendrogram of proximities that measures the similarity between clusters at every step.

III. IMAGE PROCESSING APPLICATIONS

A. Object Recognition

We used the public database "Columbia object image library-100 (COIL-100)" [15]. This database contains a large number of colour images (pictures) of objects over a black background (i.e., views of the objects taken from different angles). The size of the image is 128x128 pixels. Some of the object images are very similar to other images depending on the angle from which the picture of the object was made. In all the experiments, each image was first converted to greyscale and linearly normalized (pixels had zero mean and unit variance).

The first test was to compare the base functions of different objects in COIL-100 database. A total of 20 images for each one of 8 selected objects were randomly taken from the database. A total of 2000 patches of 8x8 pixels were randomly taken for each object, subtracting the local mean. These data were used for base function calculation after whitening using PCA. Thus, dimensionality was reduced from 64 to 40. This procedure is explained in detail in [8][16][17].

The base functions were calculated with the ICAMM algorithm that was performed one time per each object data set in order to estimate the parameters. Supervised training and the Laplacian prior was used in order to estimate the source probability density functions. The estimated base functions were converted to the original feature space using the dewhitening matrix previously estimated by PCA. Fig. 1 shows the estimated base functions of 8x8 for two of the objects: a box with an inscribed label (Fig. 1a) and a vegetable (Fig. 1b). The similarities and differences between the bases of each object (borders) can be observed in the figure.

The obtained base functions were used as input to apply the algorithm explained in Section II. Experiments to create a hierarchical classification of objects were also performed using the data set of the previous example. The ICA mixture parameters were estimated using the algorithm in [10] for $K=8$ (a class per object). These ICA parameters build the lowest level of the hierarchy. Fig. 2 shows the obtained hierarchical classification of the eight objects correctly grouped into three main kinds of objects: cars and boxes; bottles and cans; and vegetables. The proposed clustering algorithm found meaningful groupings that describe the objects at higher hierarchy levels from the bases extracted in ICA mixtures at the bottom of the hierarchy.

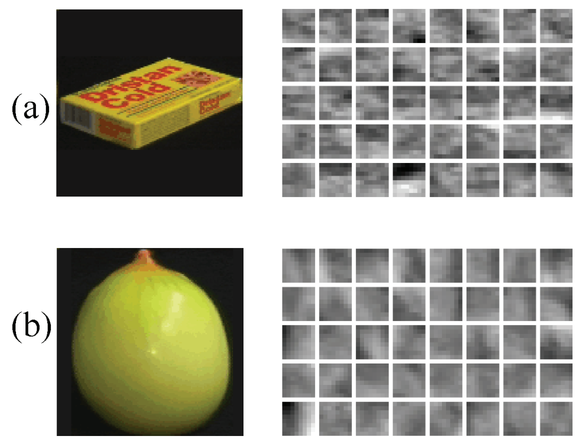

Fig. 1. Estimated base functions for two objects from the COIL-100 database.

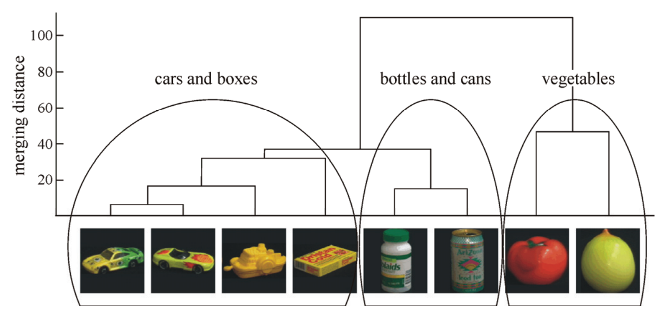

Fig. 2. Hierarchical clustering of the objects.

B. Image Segmentation

The goal of this application was to obtain logical perceptual partitions of an image such as those obtained by a human. Fig. 3 shows an image with 9 numbered zones. The total size of the image is 449x512 pixels. In all the experiments of image segmentation the following was done:

(i) a set of 1000 image patches (windows) of 8x8 pixels were taken at random location from each zone,

(ii) the normalization, whitening, and dewhitening procedure explained above was applied,

(iii) the number of classes of the ICA mixture algorithm was configured to be the number of zones of the image,

(iv) supervised training was used to estimate the ICA parameters for the lowest level of the hierarchy, and

(v) a hierarchical representation using the proposed clustering algorithm was obtained.

Fig. 3 includes a dendrogram that describes how the zones are combined starting from the base functions. Five segments have been found: sky and persons; roof skeleton; cone; stairs; stairs and persons. These segments are grouped into two broad segments (stairs and sky) distinguishing the differences in the base functions. The values of the proximities explain the similarities of the different partitions of the image.

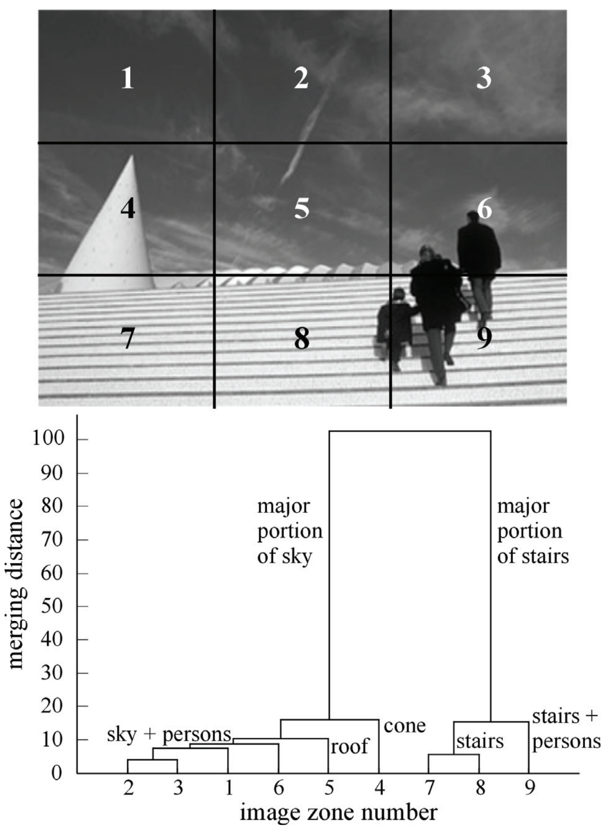

Fig. 3. Image segmentation; (a) image partition; (b) hierarchical clustering (two broad groups of zones are found).

Fig. 4 and Fig. 5 correspond to images of 1344x800 and 1140x786 pixels that are divided into 16 and 40 zones, respectively. The order of zone numbering is left to right at columns and top to down at rows; for instance, in Fig. 4 (left) the lowest zone number (1) is located at the top left corner and the highest zone number (16) is at bottom right corner of the image. Fig. 4 (left) shows a mixed image that includes two subimages: a natural image (a frog) and a text image. There are clear differences in the borders of each subimage, which are indicated in the proximities at which these subimages are combined at the penultimate level of the hierarchy (Fig. 4 (right)). Thus, the segmentation of the image into the two different subimages has been found.

Fig. 5 shows an image segmentation example where all the zones are natural. The algorithm built a suitable hierarchy by grouping zones with similar textures and borders. The proximities of the objects in the photo also determine their borders and thus the way that zones are clustered. The result in Fig. 5 indicates that there are three broad segments in the image: the sky, the horizon line area, and the rice field. Therefore, the hierarchical structures obtained from the zones of the natural image allow for an intuitive interpretation of the scene from different degrees of generalization. This is significant since it can be related with a complex abstraction process.

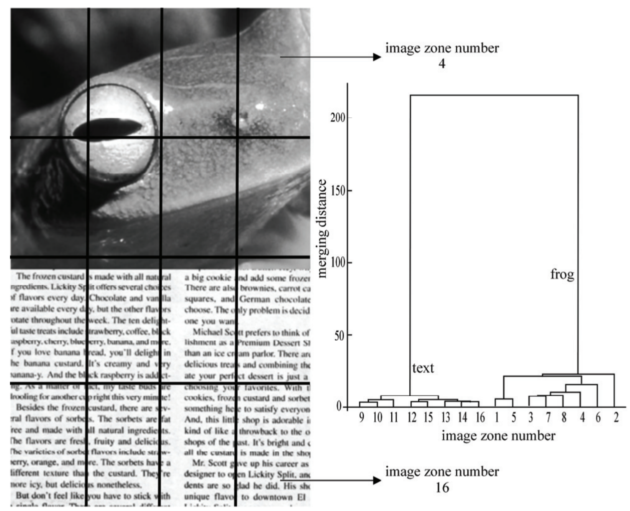

Fig. 4. Natural and text subimages.

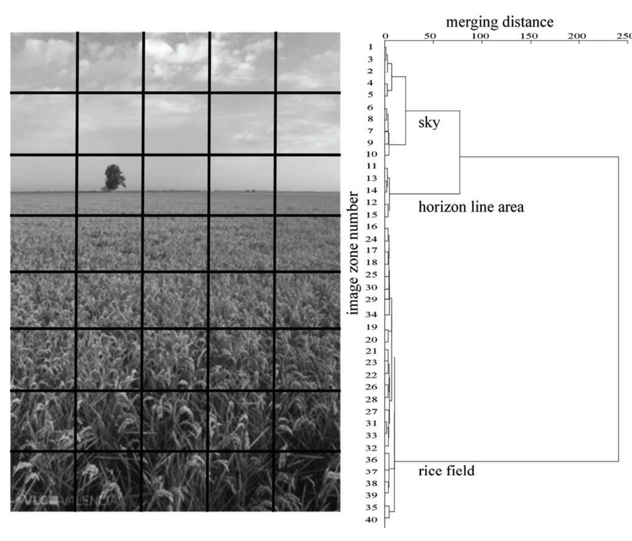

Fig. 5. Natural image.

IV. CONCLUSIONS

Two applications of image pattern recognition using mixtures of non-Gaussian distributions have been presented. Logical partitions and classifications of object images and natural images have been obtained using an agglomerative clustering algorithm formulated in the framework of independent component analysis. The statistical differences captured in the base functions of the images, that were combined step by step using the Kullback-Leibler divergence, allow hierarchical pattern structures like those produced in human perception were obtained.

The importance of image pattern recognition is currently recognized, especially in big data applications. The degrees of freedom provided by the statistical method applied allow local independence and global dependence relations of the variables were modelled for an automatic partition of the processed images. Thus, from the presented results, applications such as content-based image retrieval in semantic spaces might be attempted.

ACKNOWLEDGMENT

This work was supported by *Generalitat Valenciana* under grants PROMETEO II/2014/032, ISIC 2012/006, and GV/2014/034.

REFERENCES

[1] J.M. Kizilirmaka, F. Röslerb, P.H. Khader, “Trial-to-trial dynamics of selective long-term-memory retrieval with continuously changing retrieval targets”, Brain and Cognition, vol. 90, pp. 8-18, 2014.

[2] J.Z. Chadick, T.P. Zanto, and A. Gazzaley, “Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing”, Nature Communications, vol. 5, 5:4223, doi: 10.1038/ncomms5223, 2014.

[3] G.B. Giannakis, F. Bach, R. Cendrillon, M. Mahoney, J. Neville (Guest Editors), “Big Data: Theoretical and algorithmic foundations,” IEEE Signal Processing Magazine, vol. 31, no 5, Sept. 2014.

[4] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, “Towards large-scale histopathological image analysis: Hashing-based image retrieval”, IEEE Transactions on Medical Imaging, vol. 34 no. 2, pp. 496-506, 2015.

[5] S. Smith and I. Williams, “A statistical method for improved 3D surface detection”, IEEE Signal Processing Letters, vol. 22 no. 8, pp. 1045-1049, 2015.

[6] F. Arandiga and D.F. Yañez, “Cell-average multiresolution based on local polynomial regression. Application to image processing”, Applied Mathematics and Computation, vol. 245, pp. 1-16, 2014.

[7] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, A. Ortiz, “Automatic detection of Parkinsonism using significance measures and component analysis in DaTSCAN imaging”, Neurocomputing, vol. 126, pp. 58-70, 2014.

[8] P. Comon, C. Jutten eds, Handbook of Blind Source Separation, Independent Component Analysis and Applications, Academic Press, 2010.

[9] R. Llinares, J. Igual, A. Salazar, A. Camacho, "Semi-blind source extraction of atrial activity by combining statistical and spectral features", Digital Signal Processing: A Review Journal, vol. 21, no. 2, pp. 391-403, 2011.

[10] A. Salazar, L. Vergara, "ICA mixtures applied to ultrasonic nondestructive classification of archaeological ceramics", Eurasip Journal on Advances in Signal Processing, vol. 2010, Article number 125201, doi:10.1155/2010/125201, pp. 1-11, 2010.

[11] D.J. Mackay, Information theory, inference, and learning algorithms. Cambridge University Press, 2004.

[12] C.E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948.

[13] A. Soriano, L. Vergara, A. Bouziane, A. Salazar, "Fusion of scores in a detection context based on Alpha integration", Neural Computation, vol. 27, no. 9, pp. 1983-2010, 2015.

[14] A. Salazar, J. Igual, L. Vergara, and A. Serrano, “Learning hierarchies from ICA mixtures”, Proceedings of IEEE International Conference on Neural Networks, IJCNN 2007, Article number 4371312, pp. 2271-2276, Orlando, USA, 2007.

[15] S.A. Nene, S.K. Nayar, and H. Murase, “Columbia Object Image Library (COIL-100),” Technical Report CUCS-006-96, February 1996.

[16] A.J. Bell and T.J. Sejnowski, “The ‘Independent Components’ of natural scenes are edge filters,” Vision Research, vol. 37, no. 23, pp. 3327-3338, 1997.

[17] J.H. Van Hateren and A. van der Shaaf, “Independent component filters of natural images compared with simple cells in primary visual cortex,” Proc. of Royal Society of London: B, vol. 265, pp. 359-366, 1998.