An Approach to Visual Scenes Matching with Curvilinear Regions

J.M. Pérez-Lorenzo[1], A. Bandera[2], P. Reche-López[1], R. Marfil[2], and R. Vázquez-Martín[2]

[1] Dept. Ing. Telecomunicación, Universidad de Jaén, Linares-23700, Spain
[2] Dept. Tecnología Electrónica, Universidad de Málaga, Málaga-29071, Spain

Abstract. This paper presents a biologically-inspired artificial vision system. The goal of the proposed vision system is to correctly match regions among several images to obtain scenes matching. Based on works that consider that humans perceive visual objects divided in its constituent parts, we assume that a particular type of regions, called curvilinear regions, can be easily detected in digital images. These features are more complex than the basic features that human vision uses in the very first steps in the visual process. We assume that the curvilinear regions can be compared in their complexity to those features analysed by the IT cortex for achieving objects recognition. The approach of our system is similar to other existing methods that also use intermediate complexity features for achieving visual matching. The novelty of our system is the curvilinear features that we use.

1 Introduction

Biological vision and artificial vision are two strongly interconnected fields. Biological vision theories can be helpful for new insights and development of artificial visual systems. At the same time, artificial visual systems allow proving models and theories based on biological experiments. Also, a better theoretical understanding of the computation when processing visual information can supply useful guidelines for empirical studies of the biological processes. Despite of these interrelations, even the best artificial system cannot rival the performance of a few years old child. For example, image matching is still a fundamental aspect of many problems in computer vision, including scene recognition, stereo correspondence and motion tracking. A good starting point to approach this problem can be found in visual theories, which include scientific fields such as biology, psychology and neuroscience.

In artificial vision, image matching is defined as the process of bringing two images into agreement so that corresponding pixels in the two images correspond to the same physical region of the scene. The similarity may be applied to global features derived from the original images. However, this is not the more efficient solution and we think it does not match with the biological principles of human vision. The approach we have used is to analyse first the scenes in order to extract some features and build a scheme to work with these features.

J. Mira and J.R. Álvarez (Eds.): IWINAC 2007, Part II, LNCS 4528, pp. 409–418, 2007.

2 Biological Inspiration

In recent years great advances have been reached in our understanding in the primate visual system. Some experiments show that primate vision works in a hierarchical way. The first stages use simple local features, and the image is subsequently represented in terms of larger and more complex features. In the earliest processing stages, which involve the retina, lateral geniculate nucleus (LGN) and primary visual cortex (V1), the image is represented by simple local features such as center-surround receptive fields and oriented lines and edges. After these steps, moderately complex features are represented in areas V4 and the adjacent region TEO, and finally, partial or complete object views are represented in anterior regions of inferior temporal (IT) cortex. Indeed, there are now great evidences showing that object recognition makes use of neurons in IT that respond to features of intermediate complexity [7][12]. So, a partial answer to objects recognition is to consider that every image can itself be broken down into components and that visual cortical neurons are specialized to detect only a subset of these components. In [12] it is shown by computational analysis and simulations that features of intermediate complexity and partial object views are optimal for visual object classification. Ullman suggests, based on equations of mutual information and entropies, that intermediate complexity features are more informative than very simple ones (detected in V1 receptive fields) or very complex ones. Indeed, it is shown that visual features of intermediate complexity emerge naturally from a coding principle of maximizing the delivered information with respect a class of objects.

In object recognition no much work has solved the question of specificity, that is, which properties of an object are exactly encoded in the neural representation used to recognize that object, and also the problem of feature selection is a fundamental question [12] [4]. Nevertheless, when using intermediate complexity features, these features seem to be very sensitive to particular combinations of local shape, colour, and texture properties [7]. Some researchers have suggested that objects are represented in terms of their constituent parts. Some of the most important ones are Biederman's "Recognition by Components" [2], where objects are formed with an alphabet of simple geometrical shapes called geons, and Marr's "Generalized Cylinders" [9], where these parts are cylinders along the main axes of the object. In both cases, the constituent parts of the object are "intuitive" in the sense that they correspond to the parts in terms of which normal observers commonly understand the objects (e.g. the legs of a table). An alternative approach can be found in [12], who suggested that the intermediate complexity fragments need not correspond to those intuitive parts for the purpose of classifying an object. The work that we present here is more inspired in Marr's theory. In fact, based on the supposition of the generalized cylinders we assume that digital images have got visual regions that can be easily detected thanks to their cylindrical shape, and we call them curvilinear regions. The purpose of our system is not to detect every component described by Marr's theory because it would be a quite complex computational task. However, our experiments have shown that with the detection of some of them, and using them

as features for matching and classification stages, the system is able to identify images and scenes with promising results.

Also the above ideas lead to the discussion among "image-based" models, in which objects are represented as collection of viewpoint-specific features and "structural-description" models, in which objects are represented as configurations of 3D volumes. Nowadays, there is still no common agreement among researchers. This issue is extensively discussed in the literature. In [6] a better performance is suggested for 3D shape representations compared with open contour and closed contour models. However, they find also an advantage for using 2D polygons to approximate the object surface shapes, which provides a new framework for the studies of 3D shape representation. In [11] a mixture of both models is suggested to obtain the most appealing aspects of both of them. In our work, although the curvilinear regions are based on the existence of generalized cylinders, we do not build a 3D framework. In fact, our work is more similar to the "image-based" models because we work with the projections of the features, although we transform them in a way that can be interpreted as an extended image-based approach according to [11].

In the computational literature there are other methods that uses intermediate complex features for classifying objects and scenes with good results [7][12][10]. Our work differs mainly from these in the features that we are using, the curvilinear regions.

3 Curvilinear Regions

3.1 Definition

We define our features of intermediate complexity as curvilinear regions represented by a parameter vector $\{a(l), w_l(l), w_r(l)\}_{l=0..L}$, being L the length of the region, $a(l)$ a vector defining the axis between the right and left borders ($b_r(l)$ and $b_l(l)$), and $w_r(l)$ and $w_l(l)$ the widths of the curvilinear region (see Fig. 1). We consider that a region is a curvilinear region if it satisfies the following conditions: i) there must be a geometric similarity around the region axis, ii) ratio between its average width and its total length must be less than a predefined threshold, iii) left and right borders must be locally parallel, iv) the colour along this axis should be homogeneous. The first three properties are geometrical properties. The last one is a restriction colour in order to make the detection easier, as well as it provides the existence of a simple colour descriptor for the region.

3.2 Overview of the Proposed Method

The algorithm for detecting the curvilinear regions can be divided in several steps. Firstly, the original image is segmented into homogeneous colour regions using a pyramid algorithm based on the Bounded Irregular Pyramid (BIP) [8]. The obtained regions comply with the homogeneous colour property, so the geometric properties of every region are checked looking for those regions with the

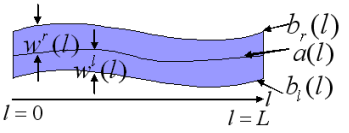

Fig. 1. Definition of curvilinear region. $b_l(l)$,$b_r(l)$: left and right borders, $a(l)$: medial axis, $w_l(l)$,$w_r(l)$: left and right widths.

curvilinear properties (this step requires the extraction of the medial axis for each region). As the result of this analysis step the method obtains the set of the detected curvilinear regions. Then, a normalisation step is applied. If the curvilinear region is enclosed inside an elliptical region whose centre is the centre of mass of the region, this step transforms the elliptical region into a circular reference region of fixed size. These normalised regions are employed as the input of the shape descriptor, which is a contour-based approach to object representation that uses a curvature function for the description of the boundary region. This contour descriptor is invariant to rotation and translation, and partially invariant to noise, scaling and skewing. The recognition stage matches the obtained individual features to a database of features from known scenes using a nearest-neighbour algorithm. This algorithm uses the curvature descriptor, colour descriptor and position of the region in the image. Experimental results show that this approach to scene recognition can match images taken from diffe-rent viewpoints if they present a similar layout, i.e. spatial distribution of curvilinear objects.

3.3 Image Segmentation Based on the Bounded Irregular Pyramid

In the present work we have used a pyramid segmentation algorithm. We have used a framework where the pyramid represents the image at multiple levels of abstraction. In each level the hierarchy presents a set of vertices V_l connected by a set of edges E_l. These edges connect vertices at the same level (intra-level edges) and define also relationships between pyramid levels (inter-level edges), where the vertices at level l connected with one vertex at level $l+1$ are defined as the sons of the vertex at level $l+1$, being this one the parent. The value of each parent is computed from its sons using a reduction function. Using this framework, the local evidence accumulation is achieved by the successive building of level $G_{l+1} = (V_{l+1}, E_{l+1})$ from level $G_l = (V_l, E_l)$. We have used the bounded irregular pyramid (BIP) proposed in [8]. In the irregular pyramids the main problem is that the size is not bounded, but in the BIP this is solved by combining the simplest regular and irregular structures. As we can see in Fig. 2, the segmented images are obtained correctly with this technique.

3.4 Medial Axis Extraction and Skeleton Classification

The skeleton of the region is used for the analysis of the geometric properties that define if a region is curvilinear or not. The skeleton is defined as a subset

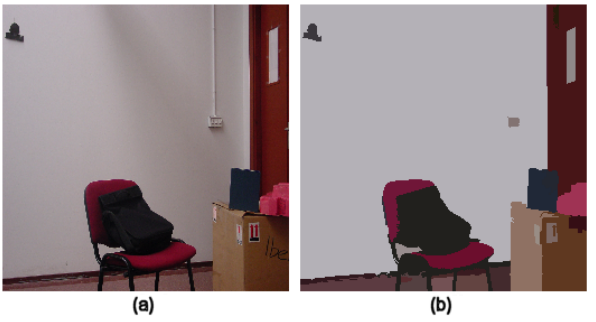

Fig. 2. Segmentation results obtained using the BIP structure: a) original image; b)segmentation results

of pixels that preserve the topological information of the region and it must approximate the medial axis. In this work a distance transformed approach is used for each colour segmented region, where a distance skeleton is a subset of grid points such that every point represents the centre of a maximal disc contained in the given component. For estimating the distance transform of a region we use the algorithm based on the $d8$-distance described in [5] which can approximate the distance transform inside the region in only two steps, so it has got a low computational cost. Those pixels which present a local maximum in the distance transform belong to the distance skeleton, and by choosing them we can obtain a skeleton for each region. These distance skeletons are generally not connected, so we post-process them with morphological operations to obtain connected and smooth skeletons that are used to estimate further geometric properties.

Our method decides which parts of the skeleton belong to a curvilinear region by estimating a set of geometric characteristics: symmetry around the skeleton, ratio between its average width and its total length, and borders parallelism.

Symmetry Around the Skeleton. The method looks for those pixels which comply with the requirement of symmetry around the axis as indicated in 3.1. To describe the algorithm we can define a skeleton as the set of connected pixels $p_s = (i_s, j_s), 0 \leq s \leq N-1$, being N the number of pixels being evaluated of the skeleton. In a first step, the normal vector is calculated for each pixel ps in the skeleton, and the cross-points between the normal and the left and right borders of the region are estimated. If we define p_s^l and p_s^r as these cross-points, then we obtain the triplets $(p_s, p_s^l, p_s^r), 0 \leq s \leq N-1$. The symmetry condition can be defined as:

$$\frac{1}{N}\sum_{s=0}^{N-1}(\Delta w_s - \overline{\Delta w})^2 \leq U_{\Delta w}(1 - e^{-\frac{N^2}{2\sigma^2_{\Delta w}}}) \tag{1}$$

with

$$\Delta w_s = |w_s^l - w_s^r| \tag{2}$$

$$\overline{\Delta w} = \frac{1}{N} \sum_{s=0}^{N-1} \Delta w_s \tag{3}$$

being w_s^l the Euclidean distance between pixels p_s and p_s^l and w_s^r the Euclidean distance between pixels p_s and p_s^r. The left side in Ec. (1) is a term that grows with the asymmetries of the region and the values $U_{\Delta w}$ and $\sigma_{\Delta w}$ in the right side are parameters of the method, we have used $U_{\Delta w} = 10$ and $\sigma_{\Delta w} = \sqrt{50}$ in our experiments.

Ratio. Given a position s in the skeleton, the width w_s of the region is estimated as the Euclidean distance between pixels p_s^l and p_s^r. In order to assure that the ratio between the width and the total length of the region is less than a predefined threshold, the following condition must be satisfied:

$$L_{max} \geq U_w \frac{1}{N} \sum_{s=0}^{N-1} w_s \tag{4}$$

being L_{max} the maximum length that the curvilinear region could have, which is estimated with all the connected pixels of the skeleton. U_w is also a parameter of the method set to 1.5.

Borders Parallelism. To check the borders parallelism requirement we estimate the tangential vectors on the borders at pixels p_s^l and p_s^r. Then, we calculate the angle between those vectors and the normal vector given a position s, obtaining angles α_s^l and α_s^r. We consider that the borders are parallel if the following condition is satisfied:

$$\frac{1}{N} \sum_{s=0}^{N-1} |\alpha_s^l - \alpha_s^r| \leq U_{\Delta\alpha} \tag{5}$$

$U_{\Delta\alpha}$ is a parameter of the method set to 30 degrees.

Classification Algorithm. When the skeleton has been extracted from the distance transform image associated to an object, the algorithm tries to join as many pixels as possible to form a curvilinear skeleton. The algorithm starts in an endpoint of the skeleton and it looks for adding the connected pixels checking if Ec. (1), Ec. (4) and Ec. (5) are true with the new added pixel. If they are true, the pixel is added and the next pixel will be studied. In case that any requirement is not fulfilled, the curvilinear skeleton is finished and a new curvilinear region will begin with the next positive evaluation. When all the pixels have been evaluated inside a region, the curvilinear skeletons with close endpoints are linked. Those parts of the objects with a skeleton evaluated as a curvilinear skeleton are considered curvilinear regions. In our experiments, we demand that these regions must have a minimum length of 10 pixels. Figs. 3.a and 3.b present an experiment and its results.

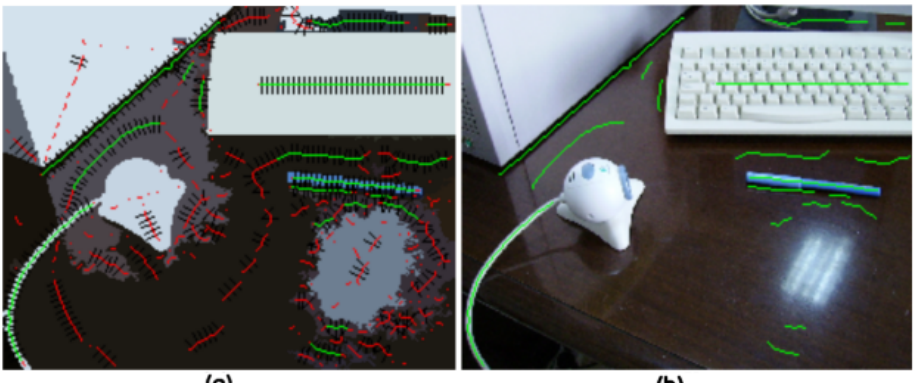

Fig. 3. a) Regions in a segmented image. The extracted skeletons have been drawn (in green colour the skeletons classified as curvilinear and in red colour as not curvilinear). Also some estimated normal vectors (black colour) to the skeletons have been drawn. b) Original image with the detected curvilinear skeletons superimposed. Several interesting objects as the ball pen, keyboard and webcam cable have been detected.

3.5 Normalisation Stage and Shape Description

In the normalisation stage we apply a 2D normalisation before obtaining the descriptor. For this we enclose the region inside an elliptical-shaped region and transform it into a circular region of fixed size (see Fig. 4.a)

The transformed curvilinear regions extracted are characterized using a shape descriptor. We have used the curvature function of the boundary of the region as the shape descriptor, which encodes the shape contour in terms of their local curvature or orientation. We have used an angle-based curvature estimator, where the curve orientation is estimated at each point with respect to a reference direction. In this method, the contour curvature $K(t)$ can be defined as the variation of the curve slope $\psi(t)$ with respect to t, or the inverse of the curvature radius $\rho(t)$:

$$K(t) = \frac{\delta\psi(t)}{\delta t} = \frac{1}{\rho(t)} \tag{6}$$

To extract $K(t)$ we have used an approach based on a k-slope algorithm which estimates the curvature using a k value which is adaptively changed [1]. Fig. 4.c shows an example of a curvature function associated to a shape contour.

4 Experimental Results

The curvilinear regions detected are characterised using a 260-dimensional space whose first two dimensions $(x, y)_i$ are the co-ordinates of the centre of mass of the region, the second two dimensions $(h, s)_i$ are the mean hue and saturation values of the region (HSV colour space), and the other 256 values $fc_{i_{i=1..256}}$

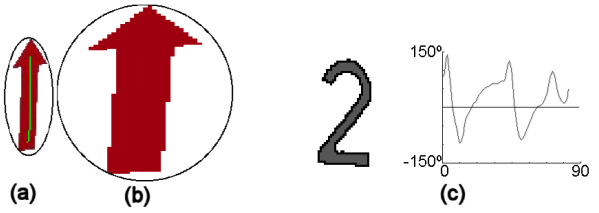

Fig. 4. a),b) Original and normalized curvilinear region; c)Curvilinear region shape and associated curvature function

are the curvature function of the object shape. So, each region is characterised by the shape of the contour, the position in the image and the homogeneous colour inside the region. We decide that two images will be similar if their sets of curvilinear regions are similar. For checking this similarity we define the distance between two curvilinear regions i and j as

$$D(i,j) = \alpha_1\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + \alpha_2\sqrt{s_i^2 + s_j^2 - 2s_i s_j cos\theta} + \\ +\alpha_3 max\{(fc_i * fc_j)_{k=1..256}\} \quad (7)$$

where θ is equal to $|h_i - h_j|$ if this value is less than π, or equal to $(2\pi - |h_i - h_j|)$ in any other case. Parameters α_i are chosen to define the importance of the position, colour and shape into the distance measure.

Given a query image Q and a dataset of images B_i with the curvilinear regions already characterised, the image matching process firstly extracts the set of N_Q curvilinear regions $\{c_{Q_i}\}_{i=1..N_Q}$ of the query image. The comparison between Q and each image B_i is achieved by comparing each curvilinear region in Q, c_{Q_i}, with all the N_{B_i} curvilinear regions present in B_i, $\{c_{Bi}\}_{i=1...N_{B_i}}$. The most similar region is then selected (if the distance value is less than a threshold) and both curvilinear regions are paired. When all the regions are processed then a similarity value is assigned to the image of the dataset. This similarity value is defined as

$$\lambda = (N_{B_i} - N'_Q + 1)\sum_{i=1}^{N'_Q} D(cQ_i, cBi_j) \quad (8)$$

where N'_Q is the number of paired curvilinear regions. Once this similarity value is computed for each image B_i of the dataset, they are sorted according to it. Those images B_i with a similarity value below a threshold are possible candidates to be matched with the query image Q.

To test this scheme, a database of 40 images obtained in an office-like environment has been created, which are divided into 10 different scenarios. In Fig. 5 we present two example retrievals for this database where the query image is the leftmost image in the rows, and subsequent images are nearest neighbours. In the images, the detected regions for the match among the scenes have been marked. To evaluate the matching performance, a normalised average rank has been used [3]:

$$\overline{R} = \frac{1}{NN_R}\left(\sum_{i=0}^{N_R-1} R_i - \frac{N_R(N_R-1)}{2}\right) \tag{9}$$

R_i is the rank at which the ith relevant image is retrieved, N_R is the number of relevant images for a given query, and N is the number of examples in the database. In our experiments the normalised average rank of relevant images present an average value of 0.025 and a standard deviation of 0.001 (if the average value is equal to zero it means a perfect performance)

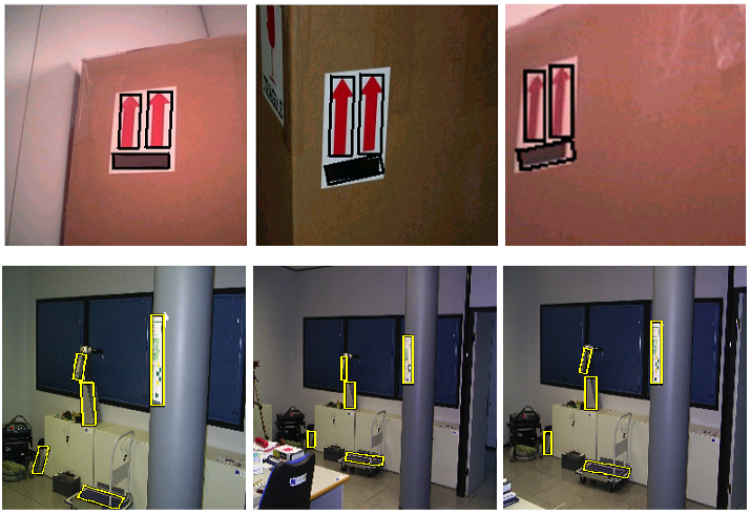

Fig. 5. Example retrievals for a database of office-like environment images

5 Conclusions and Future Work

We have presented a method which compares scenes based on the matching of detected curvilinear regions. The proposed method is inspired in biological theories which claim that objects can be divided in their constituent parts or regions in order to recognize them.We represent and compare these regions using their colour and position in the scene as well as a set of geometric properties of them. According to the literature, we think that there is an analogy between the curvilinear used features and the intermediate complexity features for the identification of objects in visual processes. These intermediate complexity features seem to be very sensitive to combinations of local shape, colour and texture properties and they are supposed to be used in the process of identifying visual objects. By adding the relative position in the scene to the characterization of the features, we have extended their use not only to object recognition but also to identify the visual scene.

Future work should be focused to further experiments in more scenarios in order to detect more regions in the images, and to study in more depth the possible analogies between the curvilinear regions used in this work and the features used in IT cortex for visual recognition, according to the most recent discoveries in visual neuroscience.

Acknowledgements

This work has been partially granted by FEDER and the Spanish Ministry of Education and Science under Project TEC2006-13883-C04-03 and project TIN2005-01359.

References

1. Bandera, A., Urdiales, C., Arrebola, F., Sandoval, F., "Corner detection by means of adaptively estimated curvature function", *Electronics Letters*, 36(2), 124-126, 2000.
2. Biederman, I., "Recognition-by-components: A theory of human image understanding", *Psychological Review*, 94, 115-147, 1987.
3. Grauman, K., Darrell, T., "Efficient image matching with distributions of local invariant features", *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, 627-634, 2005.
4. Haushofer, J., Baker, Cl., Kanwisher, N.,"Greater Sensitivity to Convexities than Concavities in Human Lateral Occipital Complex",*Society for Neuroscience Annual Meeting*, 2005.
5. Klette, G., "A comparative discussion of distance transformation and simple deformations in digital image processing", *Machine Graphics & Vision*, 12(2), 235-256, 2003.
6. Leek, E.C., Reppa, I., Arguin, M. "The structure of three-dimensional object representations in human vision: evidence from whole-part matching", *Journal of Experimental Psychology: Human Perception and Performance*, 31(4), 668-684, 2005.
7. Lowe, D., "Towards a computational model for object recognition in IT Cortex". *First IEEE International Workshop on Biologically Motivated Computer Vision*, Seoul, Korea, 20-31, 2000.
8. Marfil, R., Rodriguez, J.A., Bandera, A., Sandoval, F., "Bounded irregular pyramid: a new structure for color image segmentation", *Pattern Recognition*, 37(3), 623-626, 2004.
9. Marr D., Nishihara, H.K., "Representation and recognition of the spatial organization of three-dimensional shapes", *Proceedings of the Royal Society of London B: Biological Sciences*, 200, 269-294, 1978.
10. Matas, J., Chum, O., Urban, M., Pajdla, T.,"Robust wide baseline stereo from maximally stable extremal regions", *Proceedings of the British Machine Vision Conference*, 1, 384-393, 2002.
11. Tarr, M., Bulthof, H., "Image-based object recognition in man, monkey and machine", *Cognition*, 67, 1-20, 1998.
12. Ullman, S.,Vidal-Naquet, M.,Sali E., "Visual features of intermediate complexity and their use in classification", *Nature Neuroscience*, 5, 682-687, 2002.