9

Image Matching based on Curvilinear Regions

J. Pérez-Lorenzo, R. Vázquez-Martín, R. Marfil, A. Bandera & F. Sandoval
Grupo de Ing. de Sist. Integrados, Dept. Tecnología Electrónica, Universidad de Málaga
Spain

1. Introduction

Image matching, or comparing images in order to obtain a measure of their similarity, is a fundamental aspect of many problems in computer vision, including object and scene recognition, content-based image retrieval, stereo correspondence, motion tracking, texture classification and video data mining. It is a complex problem, that remains challenging due to partial occlusions, image deformations, and viewpoint or lighting changes that may occur across different images (Grauman & Darrell, 2005).

Image matching can be defined as "the process of bringing two images geometrically into agreement so that corresponding pixels in the two images correspond to the same physical region of the scene being imaged" (Dai & Lu, 1999). Therefore, according to this definition, image matching problem is accomplished by transforming (e.g., translating, rotating, scaling) one of the images in such a way that the similarity with the other image is maximised in some sense. The 3D nature of real-world scenarios makes this solution complex to achieve, specially because images can be taken from arbitrary viewpoints and in different illumination conditions. Instead, the similarity may be applied to global features derived from the original images. However, this is not the more efficient solution. Besides, these global statistics cannot usually deal with real-world scenarios because they do not often give adequate descriptions of the local structures or discriminating features which are present on the image (Grauman & Darrell, 2005).

Other solution to the image matching problem is to describe the image using a set of *distinguished regions* (Matas et al., 2002). These regions must own some invariant and stable property in order to be detected with high repeatability in images taken from arbitrary viewpoint. Then, the matching between two images is posed as a search in the correspondence space established between the associated sets of distinguished regions. If each region is described by a vector of image pixels, then cross-correlation can be used to obtain a similarity value between two regions (Mikolajczyk & Schmid, 2005). However, due to the high dimensionality of such vector, the generation of the correlation space typically presents a high computational cost. In order to reduce the computational complexity, the number of tentative correspondences can be limited by computing local invariant descriptors for distinguished regions (Matas et al., 2002; Grauman & Darrell, 2005). These descriptors can be also employed to estimate the similarity value between two regions.

In this paper, we have adopted an approach which describes the image using a set of distinguished regions and exploits local invariant descriptors to estimate the similarity value between two distinguished regions belonging to different images. Thus, there are four

Source: Vision Systems: Segmentation and Pattern Recognition, ISBN 987-3-902613-05-9,
Edited by: Goro Obinata and Ashish Dutta, pp.546, I-Tech, Vienna, Austria, June 2007

main procedures involved in the image matching process: i) detection of distinguished regions, ii) local invariant description of these regions, iii) definition of the correspondence space, and iv) searching of a globally consistent subset of correspondences. This subset of correspondences will permit to associate a similarity score to the images being matched. The main contribution of this work is the introduction of a new set of distinguished regions, the so called *curvilinear regions*.

The choice of the location and shape of the distinguished regions can be considered as a crucial issue in these image matching approaches (Matas et al., 2002). In a typical case, when images are taken from different viewpoints, local image deformations cannot be realistically approximated by translations and rotations, and it is required a full affine model. Then, correspondence cannot be established by comparing regions of a fixed shape like rectangles or circles since their shape is not preserved under affine transformation. Region shape must depend on the image data (Dai & Lu, 1999; Matas et al., 2002). In our case, the proposed method exploits a particular image structure. It is based on the presence, in a typical image, of numerous objects which can be built using cylinders or generalized cylinders (Biederman, 1987). The main disadvantage of the method is to use shapes which must be explicitly present in the image, so it depends on the presence of these specific structures in the scene. On the contrary, curvilinear regions automatically deform with changing viewpoint as to keep on covering identical physical parts of a scene.

This chapter is organised as follows: Section 2 describes related work. The curvilinear region detector is presented in Section 3. Section 4 describes the contour-based descriptor computed for each extracted region. This descriptor is compared to other similar approaches in Section 5.1. The correspondence algorithm is presented in Section 5.2. This Section also describes some experimental results and finally, Section 6 discusses extracted conclusions and future work.

2. Related work

The development of algorithms which use a set of local distinguished items for image matching can be traced back to the works of Moravec (1981) and Harris and Stephens (1988). Although the initial applications of both approaches are for stereo and short-range motion tracking, it can be considered that a similar strategy has been later extended to deal with more difficult problems. Thus, Zhang et al. (1995) propose to match Harris points over a large image range by using a correlation window around each point. The Harris point detector selects any image location that has large gradients in all directions at a predetermined scale. Outliers are then removed by solving for a fundamental matrix describing the geometric constraints between the two views of a rigid scene and removing matches that did not agree with the majority solution.

Local invariant feature matching is extended to general image recognition problems in which a feature is matched against a large set of images by Schmid and Mohr (1997). This approach also employs Harris points as distinguished items, but rather than matching with a correlation window, they use a rotationally invariant descriptor of the local image region. The 2D translation and 2D rotation invariant features are extracted from the intensity pattern in fixed circular regions around Harris points. Invariance under scaling is handled by including circular regions of several sizes. This allows features to be matched under arbitrary orientation change between the two images. Besides, they demonstrate that multiple feature matches could accomplish general recognition under occlusion and clutter

by identifying consistent clusters of matched features. This method has been modified to deal with very large scale changes (Dufournaud et al., 2000) or with colour images (Montesinos et al., 2000).

The Harris point detector is very sensitive to scale changes, so it does not provide a good basis for matching images of different sizes. In any case, representations that are stable under scale change have been proposed. Crowley and Parker (1984) developed a detector that identifies peaks and ridges in scale-space and links these into a tree structure. The tree structure can then be matched between images with arbitrary scale change. The Harris point local feature approach has been modified by Lowe (1999) to achieve scale invariance. Circular regions that maximise the output of a difference-of-Gaussian (doG) filters in scale-space are employed. More recent work on graph-based matching by Shokoufandeh et al. (1999) provides more distinctive feature descriptors using wavelet coefficients. Harris-Laplace regions (Mikolajczyk & Schmid, 2001) are also invariant to rotation and scale changes. These points are detected by the scale-adapted Harris function and selected in scale-space by the Laplacian-of-Gaussian operator. Hessian-Laplace regions (Lowe, 2004) are localised in space at the local maxima of the Hessian determinant and in scale at the local maxima of the Laplacian-of-Gaussian. This detector obtains higher localisation accuracy than the doG approach and the scale detection accuracy is also higher than in the case of the Harris-Laplace detector (Mikolajczyk & Schmid, 2005). The problem of identifying an appropriate and consistent scale for feature detection has been studied in depth by Lindeberg (1993, 1994).

As it is commented above, when images are taken from different viewpoints, image regions are subject to affine transformations. The affine transformation includes rotation, scaling, skewing and translation (Bala & Cetin, 2004). It preserves parallel lines and equispaced points along a line. Therefore, it has been used to approximate the perspective transformation in some cases. Local features have been extended to be invariant to full affine transformations. Harris-affine regions (Mikolajczyk & Schmid, 2004) and Hessian-affine regions (Mikolajczyk et al., 2005) are invariant to affine image transformations. However, they start with initial feature scales and locations selected in a non-affine-invariant manner. Then, the affine neighbourhood is determined by the affine adaptation process based on the second moment matrix. Baumberg (2000) has proposed an invariant descriptor which cannot deal with scale changes. Thus, these regions are invariant under rotation, stretch and skew, but scale changes are dealt with by applying a scale-space approach. The error on the scale also influences the other components of the transformation. Tuytelaars and Van Gool (2004) propose two types of affine-invariant regions, one based on a combination of Harris points and edges and other one based on image intensities. Matas et al. (2002) describe the Maximally Stable Extremal Regions (MSER). They are extracted with a watershed like segmentation algorithm. An important issue that affine invariant approaches must take into account is the sensitivity to noise. Thus, affine features are sensitive to noise, so in practice they have typically lower repeatability than the scale-invariant features (Mikolajczyk, 2002). To deal with this problem, the local descriptor must allow relative feature positions to shift significantly with only small changes in the descriptor. This not only allows the descriptors to be reliably matched across a considerable range of affine distortion, but it also makes the features more robust against changes in 3D viewpoint for non-planar surfaces (Lowe, 2004).

Many other features have been proposed. Some of them make use of region boundaries, which should make them less likely to be disrupted by cluttered backgrounds near object

boundaries. Thus, Matas et al. (2002) have shown that their MSERs can produce large numbers of matching features with good stability. Mikolajczyk et al. (2003) uses local edges while ignoring unrelated nearby edges, providing the ability to find stable features even near the boundaries of narrow shapes superimposed on background clutter. Nelson and Selinger (1998) employ local features based on groupings of image boundaries. Finally, Pope and Lowe (2000) use features based on the hierarchical grouping of image boundaries. A curvilinear-based region detector has been proposed by Deng et al. (2006). It starts by detecting curvilinear structures followed by watershed segmentation to define regions. On the other hand, phase-based local features have been described by Carneiro and Jepson (2002). These features represent the phase rather than the magnitude of local spatial frequencies, which is likely to provide improved invariance to illumination. Schiele and Crowley (2000) have proposed the use of multidimensional histograms. These histograms represent the distribution of measurements within image regions and they may be particularly useful for matching textured regions with deformable shapes. Other useful properties to incorporate include colour, motion, figure-ground discrimination, region shape descriptors, and stereo depth cues.

3. Curvilinear regions

3.1 Definition

Basically, in a digital image, a curvilinear region is a set of pixels delimited by left and right boundaries, $r_l(l)$ and $r_r(l)$. This region can be defined by the parameter vector, $\{a_i,w_i\}_{i=0...L}$, where L is the length of the region, a_i a vector defining the axis between the boundaries and w_i the width of the curvilinear region (see Fig. 1). In a curvilinear region, the ratio between its average width and its total length should be less than a predefined threshold. Besides, left and right borders should be locally parallel, it should exist a geometric similarity around the region axis and the colour along this axis should be homogeneous. These items will be extended in next epigraphs.

I. Symmetry around the axis

If we define $\Delta w(l)$ as the difference of width at both sides of the medial axis:

$$\Delta w(l) = \mid w_l(l) - w_r(l) \mid \tag{1}$$

Then, we can evaluate the error on the symmetry around the axis as:

$$E_{\Delta w}(L) = \frac{1}{L}\int_0^L (\Delta w(l) - \overline{\Delta w})^2 \, dl \tag{2}$$

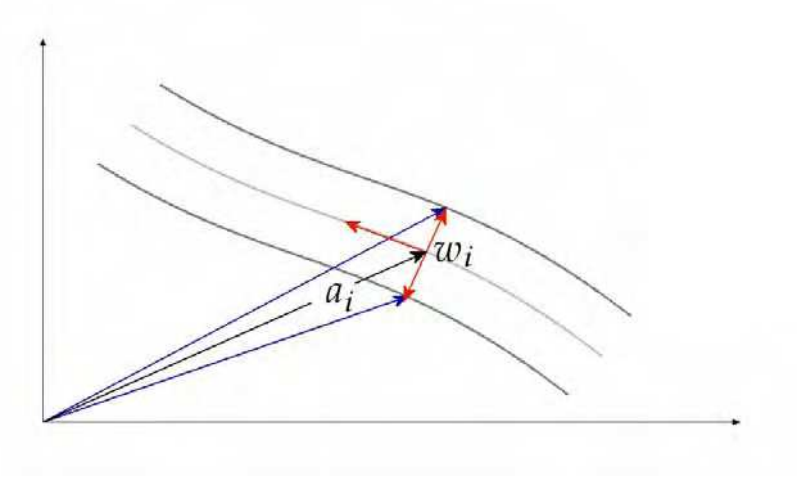

Fig. 1. Curvilinear region definition

In a curvilinear region, this error must be limited by a threshold. In our case, this threshold depends on two parameters, $U_{\Delta w}$ and $\sigma_{\Delta w}$. A curvilinear region complies with:

$$E_{\Delta w}(L) \le U_{\Delta w}(1 - e^{\frac{-L^2}{2\sigma_{\Delta w}^2}}) \tag{3}$$

II. Ratio between average width and length

If we define *w(l)* as

$$w(l) = w_l(l) + w_r(l) \tag{4}$$

Then, a curvilinear region complies with:

$$L_{\max} \ge U_w \cdot \overline{w} \tag{5}$$

where U_w is a parameter of the method and $L_{\max}$ is the maximum length of the curvilinear region. This length is obtained from all connected pixels inside the region.

III. Left and right borders locally parallel

The mean value of the difference of the tangents at both sides of the region, $\overline{\Delta\alpha}$, must be also bounded. If

$$\Delta\alpha(l) = |\alpha_l(l) - \alpha_r(l)|, \tag{6}$$

then

$$\overline{\Delta\alpha} \le U_{\Delta\alpha} \tag{7}$$

where $U_{\Delta\alpha}$ is a parameter of the method.

Section 3.5 will present an extended description of these three curvilinear region restrictions.

3.2 Overview of the proposed method

The algorithm for detecting the curvilinear regions works in a simple way. Firstly, the input image is segmented into a set of homogeneous colour regions, so the obtained regions comply with the requirement that colour must be homogeneous through the region. In order to achieve it in a fast way, a pyramid algorithm is employed: the Bounded Irregular Pyramid (BIP) (Marfil et al., 2004). The BIP divides the original image into a set of connected regions which present an homogeneous colour. Then, every image region is checked in order to look for curvilinear regions by analysing its medial axis and borders. Several curvilinear regions can be detected in the same object. Once the curvilinear regions have been extracted from the input image, an extra normalisation step is applied to compensate for part of the deformations (Tuytelaars & Van Gool, 2004). If the curvilinear region is enclosed inside an elliptical region whose centre is obtained as the centre of mass of the region, the normalisation step transforms this elliptical region to a circular reference region of fixed size. Then, normalised curvilinear regions are employed as the input of a shape descriptor. The used shape descriptor is described in Section 4. Basically, it is a contour-based approach to object representation which characterises the region boundary using a curvature function. The obtained contour descriptor is invariant to rotation and translation, and partially invariant to noise, scaling and skewing.

Finally, the approach uses these high-level features for scene recognition. The recognition proceeds with matching individual features to a database of features from known scenes using a nearest-neighbour algorithm based on a curvature matching criterion. The relative pose of recognised features is employed to identify the image layout. Experimental results show that this approach to scene recognition can match images taken from different viewpoints if they present a similar layout, i.e. spatial distribution of curvilinear objects. The image matching process is described in Section 5.2.

3.3 Image segmentation based on the Bounded Irregular Pyramid

In our approach, image segmentation is employed to obtain a global set of image regions. Subsequent stages will perform the region characterisation and they will obtain the final set of curvilinear regions. Particularly, we have used a pyramid segmentation algorithm because these approaches exhibit interesting properties with respect to segmentation algorithms based on a single representation. Thus, local operations can adapt the pyramidal hierarchy to the topology of the image, allowing the detection of global features of interest and representing them at low resolution levels. This general principle was briefly described by Jolion and Montanvert (1992): "*a global interpretation is obtained by a local evidence accumulation*."

In order to accumulate the local evidence, a pyramid represents the contents of an image at multiple levels of abstraction. Each level of this hierarchy is at least defined by a set of vertices V_l connected by a set of edges E_l. These edges define the horizontal relationships of the pyramid and represent the neighbourhood of each vertex at the same level (*intra-level edges*). Another set of edges define the vertical relationships by connecting vertices between adjacent pyramid levels (*inter-level edges*). These inter-level edges establish a dependency relationship between each vertex of level l+1 and a set of vertices at level l (*reduction window*). The vertices belonging to one reduction window are the sons of the vertex which defines it. The value of each parent is computed from the set of values of its sons using a

reduction function. The ratio between the number of vertices at level l and the number of vertices at level l+1 is the *reduction factor*.

Using this general framework, the local evidence accumulation is achieved by the successive building of level $G_{l+1}=(V_{l+1},E_{l+1})$ from level $G_l=(V_l, E_l)$. This procedure consists of three steps:

1. Selection of the vertices of G_{l+1} among V_l: This selection step is a *decimation procedure* and selected vertices V_{l+1} are called the surviving vertices.
2. Inter-level edges definition: Each vertex of G_l is linked to its parent vertex in G_{l+1}. This step defines a partition of V_l.
3. Intra-level edges definition: The set of edges E_{l+1} is obtained by defining the adjacency relationships between the vertices V_{l+1}.

The parent–son relationship defined by the reduction window may be extended by transitivity down to the base level. The set of sons of one vertex in the base level is named its *receptive field*. The receptive field defines the embedding of this vertex in the original image. In a general view of the pyramid hierarchy, the vertices of the bottom pyramidal level (level 0, also called base level) can be anything from an original image pixel via some general numeric property to symbolic information, e.g. a vertex can represent an image pixel grey level or an image edge. Corresponding to the generalization of the vertex contents, the intra-level and inter-level relations of the vertices are also generalized.

After building the pyramidal structure, the segmentation of the input image can be achieved either by selecting a set of vertices from the whole hierarchy as region roots, or by choosing as roots all the vertices which constitute a level of this hierarchy. In any case, this selection process depends on the final application and it must be performed by a higher level task. The efficiency of a pyramid to solve segmentation tasks is strongly influenced by two related features that define the intra-level and inter-level relationships. These features are the data structure used within the pyramid and the decimation scheme used to build one graph from the graph below (Brun & Kropatsch, 2003). The choice of a data structure determines the information that may be encoded at each level of the pyramid and it defines the way in which edges E_{l+1} are obtained. Thus, it roughly corresponds to setting the horizontal properties of the pyramid. On the other hand, the reduction scheme used to build the pyramid determines the dynamics of the pyramid (height, preservation of details, etc.). It defines the surviving vertices of a level and the inter-level edges between levels which correspond to the vertical properties of the pyramid. Taking into account these features, pyramids have been roughly classified as regular and irregular pyramids. A *regular pyramid* has a rigid structure where the intra-level relationships are fixed and the reduction factor is constant. In these pyramids, the inter-level edges are the only relationships that can be changed to adapt the pyramid to the image layout. The inflexibility of these structures has the advantage that the size and the layout of the structure are always fixed and well-known. However, regular pyramids can suffer several problems (Bister et al., 1990): non-connectivity of the obtained receptive fields, shift variance, or incapability to segment elongated objects. In order to avoid these problems, *irregular pyramids* were introduced. In the irregular pyramid framework, the spatial relationships and the reduction factor are not constant. Original irregular pyramids presented a serious drawback with respect to computational efficiency because they gave up the well-defined neighbourhood structure of regular pyramids. Thus, the pyramid size cannot be bounded and hence neither can the time to execute local operations at each level (Willersinn & Kropatsch, 1994). This problem has

been resolved by recently proposed strategies (Brun & Kropatsch, 2003; Haxhimusa et al., 2003; Marfil et al., 2004).

The bounded irregular pyramid (BIP) (Marfil et al., 2004) is a hierarchical structure that merges characteristics from regular and irregular pyramids. Its data structure combines the simplest regular and irregular structures: the 2×2/4 regular one and the simple graph irregular representation. The algorithm firstly tries to work in a regular way by generating, from level l, a 2×2/4 new level l+1. However, only the 2×2 homogeneous arrays of V_l generate a new vertex of V_{l+1}. Therefore, this step creates an incomplete regular level l + 1 which only presents vertices associated to homogeneous regions at the level below. Vertices of level l which generate a new vertex in V_{l+1} are linked to this vertex (son–parent edges). Then, all vertices without parent (orphan vertices) of level l search for a neighbour vertex with a parent in level l + 1 whose colour will be similar to the orphan vertex's colour (*parent search step*). If there are several candidate parents, the orphan vertex is linked to the most similar parent. Finally, the irregular part of the BIP is built. In this step, orphan vertices, of level l, search for all neighbour orphan vertices at the same level. Among the set of candidates, they are linked with the most similar. When two orphan vertices are twined, a new parent is generated at level l + 1 (*intra-level twining step*). This parent is a node of the irregular part of the BIP. The algorithm performs these two steps simultaneously. Thus, if an orphan vertex does not find a parent in the parent search stage, it will search for an orphan neighbour to link to it (*intra-level twining*). In the parent search stage an orphan vertex can be linked with the irregular parent of a neighbour. Once this is completed, intra-level edges are generated at level l + 1. The decimation process stops when it is no longer possible to generate new vertices in the regular part of the BIP. When all the levels are generated, homogeneous vertices without parent are regarded as roots and their corresponding receptive fields constitute the segmented image.

Fig. 2 shows some segmentation results obtained using the proposed algorithm. It can be noted as the different homogeneous regions present in the image have been correctly segmented.

3.4 Medial axis extraction

The geometric properties used to check if a region is curvilinear or not are based on the extraction of the skeleton of the region. The skeleton is defined as a subset of pixels that preserve the topological information of the region and it must approximate the medial axis. There are a lot of methods to estimate the skeleton of an object and they are either based on distance transforms defined by different metrics or algorithms based on simple shape deformations (Klette, 2003). The choice of the method often depends on the task, as there is no "best method". One category is based on distance transforms, where a *distance skeleton* is a subset of grid points such that every point of this subset represents the centre of a maximal disc contained in the given component. A second category is based on iterative thinning methods, where the term *linear skeleton* can be used for the result of a continuous deformation of the frontier of a connected subset without changing the connectivity of the original set, until only a set of lines and points remains. In this work a distance transformed approach is used for each colour segmented region, therefore obtaining a skeleton for each region. This skeleton will be used to estimate further geometric properties.

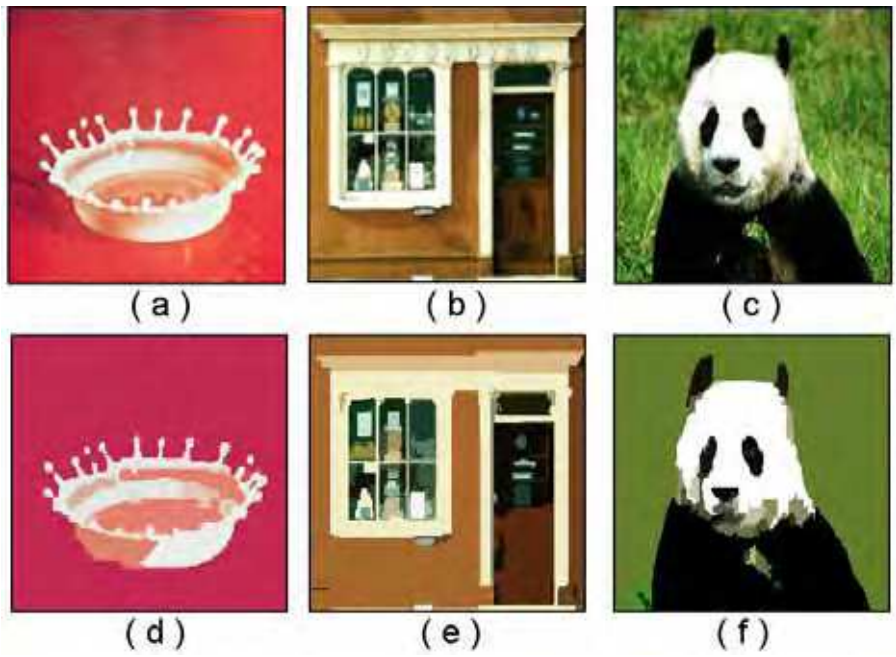

Fig. 2. Segmentation results obtained using the BIP structure (Marfil et al., 2004): a-c) original images; and d-f) segmentation results.

The distance from one point to another is the smallest positive integer n such that there exists a sequence of distinct points $p_0, p_1, p_2 \ldots p_n$ with p_i being an a-neighbour of p_{i-1}, $1 \le i \le n$. For $a = 8$, the distance $d(p,q)$ is called the d_8-distance. If (i_p, j_p) and (i_q, j_q) are the coordinates of p and q respectively, then

$$d_8(p,q) = \max\left\{\left|i_p - i_q\right|, \left|j_p - j_q\right|\right\} \tag{8}$$

For estimating the distance transform of a region we use the algorithm described in (Klette, 2003) which can approximate the distance transform inside the region in only two steps, so it has got a low computational cost. We define the original region as an image: $I(i,j) = 0$, if the pixel (i, j) belongs to the border of the region, and $I(i,j) = 255$ otherwise. In the first step the function f_1 is defined as

$$f_1(i,j,I(i,j)) \begin{cases} 0 \;\; if \;\; I(i,j) = 0 \\ \min\left\{I^*(i-1,j)+1, I^*(i,j-1)+1, I^*(i-1,j-1)+1, I^*(i-1,j+1)+1\right\} \\ \qquad\qquad if \;\; I(i,j) = 255 \;\; and \;\; i \neq 1 \;\; or \;\; j \neq 1 \\ i + j \;\; otherwise \end{cases} \tag{9}$$

The function f_1 is applied to the image I from top to bottom and from left to right, producing $I^*(i, j) = f_1(i, j, I(i, j))$. In the second step the function f_2 is defined as

$$f_2(i,j,I^*(i,j)) = \min\left\{I^*(i,j), T(i+1,j)+1, T(i,j+1)+1, T(i+1,j-1)+1, T(i+1,j+1)+1\right\} \tag{10}$$

and the resulting image T is calculated as $T(i, j) = f_2(i, j, I^*(i,j))$, applying f_2 from bottom to top and from right to left, and being T the distance transform image of I. If we choose those pixels (i_s, j_s) in the image T such as none of the points in the vicinity $A_8((i_s, j_s))$ has a value in T equal to $T(i_s, j_s)+1$ then those pixels (i_s, j_s) belong to the distance skeleton and they are supposed to be local maxima in the distance transform.
The resulting distance skeletons are generally not connected, so we post-process them with morphological operations (interpolation, dilatation, erosion and elimination of not useful pixels) to obtain a connected and smooth skeleton. By this way we obtain an approximation to the medial axis of the object.

3.5 Skeleton classification

Once the skeletons are calculated for each segmented region our method decides which parts of the skeleton belong to a curvilinear region and which not. In order to achieve this goal, several geometric characteristics are estimated: symmetry around the skeleton, ratio between average width and length, and borders parallelism (see Section 3.1).

3.5.1 Symmetry around the skeleton

The method checks those pixels which comply with the requirement of (3). To describe the algorithm we can define a skeleton as the set of connected pixels $p_s=(i_s, j_s)$, $0 \le s \le N$-1, and N the number of pixels being evaluated of the skeleton. In a first step, the normal vector is calculated for each pixel p_s 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), $0 \le s \le N$-1. We can implement (3) as

$$\frac{1}{N}\sum_{s=0}^{N-1}\left(\Delta w_s - \overline{\Delta w}\right)^2 \le U_{\Delta w}\left(1 - e^{-\frac{N^2}{2\sigma_{\Delta w}^2}}\right) \tag{11}$$

with

$$\Delta w_s = \left| w_s^l - w_s^r \right| \tag{12}$$

$$\overline{\Delta w} = \frac{1}{N}\sum_{s=0}^{N-1}\Delta w_s \tag{13}$$

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 (11) 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. For our experiments, we have used $U_{\Delta w} = 10$ and $\sigma_{\Delta w} = \sqrt{50}$. The number of pixels N also appears on the right side of (11), in a way that longer regions are allowed to have a higher value of asymmetry.

3.5.2 Ratio $L/\overline{w}$

In a similar way that Section 3.5.1, we define w_s as the width of the region estimated as the Euclidean distance between pixels p_s^l and p_s^r given a position s in the skeleton. Then, (5) is implemented as

$$L_{\max} \geq U_w \cdot \frac{1}{N} \sum_{s=0}^{N-1} w_s \tag{14}$$

$L_{\max}$ is the maximum length that the curvilinear skeleton could have and is calculated with all the connected pixels of the skeleton of the object. U_w is also a parameter of the method. In our experiments, it has been set to 1.5.

3.5.3 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. Equation (7) is implemented as

$$\frac{1}{N} \sum_{s=0}^{N-1} \left| \alpha_s^l - \alpha_s^r \right| \leq U_{\Delta\alpha} \tag{15}$$

$U_{\Delta\alpha}$ is a parameter of the method. For our experiments, it has been set to 30 degrees.

3.5.4 Classification algorithm

The algorithm to classify the skeletons into the curvilinear group or the not curvilinear one works in an easy way. Once 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. So the algorithm begins in an endpoint of the skeleton and it looks for adding the connected pixels checking if (11), (14) and (15) are true with each new added pixel. If these equations are true for a pixel, then the new pixel is added and the algorithm will check the next connected pixel in the extracted skeleton. If the new pixel does not comply with all the requirements, then the curvilinear skeleton is finished and a new curvilinear region will begin with the next positive evaluation.

Given an object and its skeleton, when all the pixels have been evaluated, the curvilinear skeletons whose endpoints are near are linked to form a longer curvilinear skeleton. At the end of the process, the parts of the objects whose skeleton has been evaluated as a curvilinear skeleton are considered as curvilinear regions. The algorithm allows to demand a minimum length L_{min} to the regions. In our experiments, the minimum length has been set to 10 pixels.

Figs. 3 and 4 present an experiment with a real scene obtained using our typical set of parameters. In Fig. 3, the results of the detection of objects and classification of the extracted skeletons are presented. Fig. 4 presents the original scene with the curvilinear skeletons superimposed.

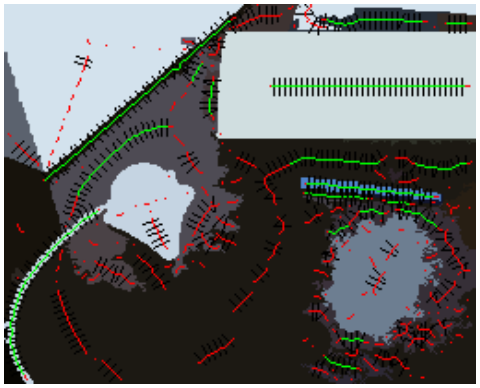

Fig. 3. Detected segmented regions in a segmentation image. The extracted skeletons have been drawn (in green colour the skeletons classified as curvilinear and in red colour as not curvilinear). Also some estimated normal vectors (black colour) to the skeletons have been drawn.

Fig. 4. Original image with the detected curvilinear skeletons (see Fig. 3). Several interesting objects as the ball pen, keyboard and webcam cable have been detected. Parameters used are: segmentation threshold = 95.0, $U_{\Delta w} = 10$, $\sigma_{\Delta w} = \sqrt{50}$, $U_{w} = 1.5$, $U_{\Delta \alpha} = 30^{o}$, $L_{min} = 10$ pixels.

3.6. Normalisation stage

As it is pointed out by Tuytelaars and Van Gool (2004), it is better to compensate for part of the geometric deformations through a normalisation stage, before obtaining the descriptor associated to the region. In our case, the geometric normalisation stage will be achieved by enclosing the curvilinear region inside an elliptically-shaped region and by transforming this region to a circular reference region of fixed size (see Fig. 5). This process leaves one degree of freedom to be determined which corresponds to a free rotation of the circular region around its centre. In our case, it is not a problem because the shape will be represented using a contour descriptor which is invariant to rotation distortions.

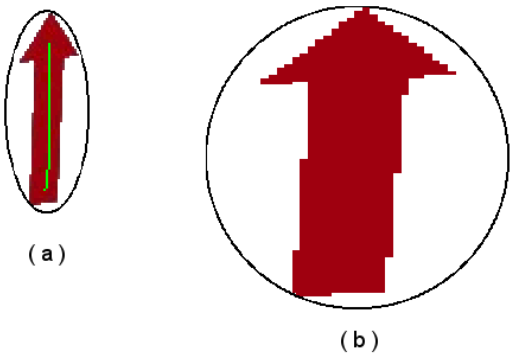

Fig. 5. a) Original curvilinear region; and b) normalised region.

4. Shape description

Once the curvilinear regions have been extracted from the input image, they are characterised using a shape descriptor. Shape representation constitutes one of the most powerful tools to represent a planar object. Therefore, many approaches have been proposed to describe shapes from a small set of features. These descriptors can be divided into those which work on a shape as a whole (*global descriptors*) and those which work on the contours of the shape (*boundary-based descriptors*). Boundary-based descriptors are less computationally intense than global ones. However, since they are based on the shape contour, they cannot take into account the internal structure of the object. Therefore, boundary-based methods are not suited to deal with certain kinds of applications. On the other hand, most of the boundary-based descriptors do not need to normalise the 2D representation of the object to achieve common geometrical invariance. Thus, a boundary-based method, the popular *curvature scale space* (Mokhtarian & Mackworth, 1986), has been used in the MPEG-7 standard.

In this work, we employ a boundary-based descriptor. Particularly, this descriptor is based on the estimation of the curvature associated to the shape contour. By definition, the curvature function encodes the shape contour in terms of their local curvature or orientation. If $c(t)=(x(t), y(t))$ is a parametric plane curve, then its curvature function $\kappa(t)$ can be calculated as (Mokhtarian & Mackworth, 1986)

$$\kappa(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}} \tag{16}$$

This equation implies that estimating the curvature involves the first and second order directional derivatives of the plane curve co-ordinates. This is a problem in the case of computational analysis where the plane curve is represented in a digital form. In order to solve this problem, two different approaches are often encountered: those that approximate the plane curve co-ordinates (*interpolation-based curvature estimators*), and those that estimate the curve orientation at each contour point with respect to a reference direction (*angle-based curvature estimators*). In addition, both type of methods can be subdivided in single scale methods and multiscale ones. Single scale methods are based upon an analysis of the

contour using a fixed set of parameters. Multiscale methods represent the evolution (or deformation) of the original contour when a certain parameter value is varied.
The described shape descriptor is grouped into the angle-based curvature estimators. These approaches propose an alternative curvature measure based on angles between vectors which are defined as a function of the curve co-ordinates. Thus, the contour curvature $\kappa(t)$ can be defined as the variation of the curve slope $\psi(t)$ with respect to t, that is, the inverse of the curvature radius $\rho(t)$:

$$\kappa(t) = \frac{\partial \psi(t)}{\partial t} = \frac{1}{\rho(t)} \tag{17}$$

In order to extract $\kappa(t)$ from a digital contour, several methods have been proposed. The majority of these approaches consist of comparing segments of k-points at both sides of a given point to estimate its curvature. Therefore, the value of k determines the cut frequency of the curve filtering. So, these algorithms are single scale methods in which only features unaffected by the filtering process may be detected. On the contrary, Beus and Tiu (1987) propose a multiscale angle-based approach which modifies the Freeman's approach (Freeman, 1978) by averaging the results obtained for several values of k. However, this approach is slow and, in any case, it must choose the cut frequencies for each iteration (Bandera et al., 2000).
Another solution is to adapt the cut frequency of the filter at each curve point as a function of the local properties of the shape around it. A k-slope algorithm which estimates the curvature using a k value which is adaptively changed according to the local information of the boundary is proposed by Bandera et al. (2000). In this work, we will employ this curvature estimator. Thus, Fig. 6 shows several examples of curvature functions associated to different shape contours.

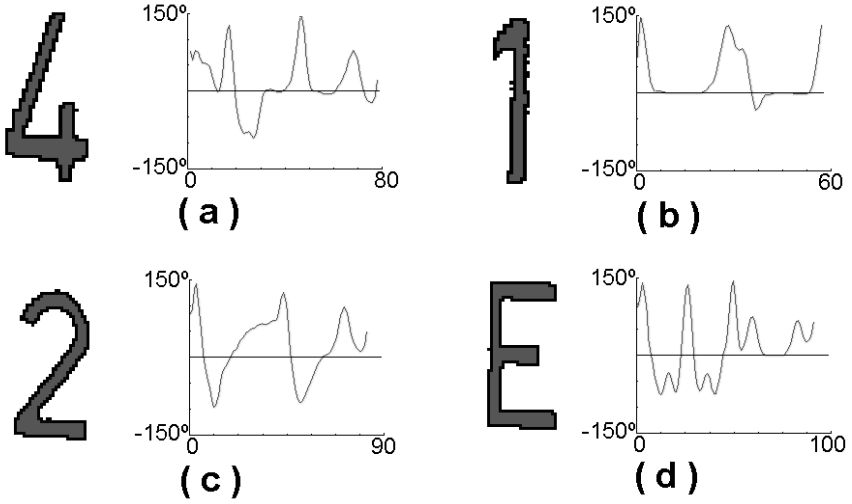

Fig. 6. a-d) Curvilinear region shapes and associated curvature functions

5. Experimental results

5.1. Shape description: a comparative study

The proposed shape descriptor has been compared to other methods to test its performance. Particularly, we chose for the purpose of comparison the methods proposed by Bernier and Landry (2003) and Zhang and Lu (2005). The first method employs a contour-based descriptor, whereas the second one is rather region-based. In order to compare the performance of the different methods, a publicly available data set (Sebastian et al., 2001) was employed[1]. This data set consists of nine classes with eleven shapes in each cluster (see Fig. 7).

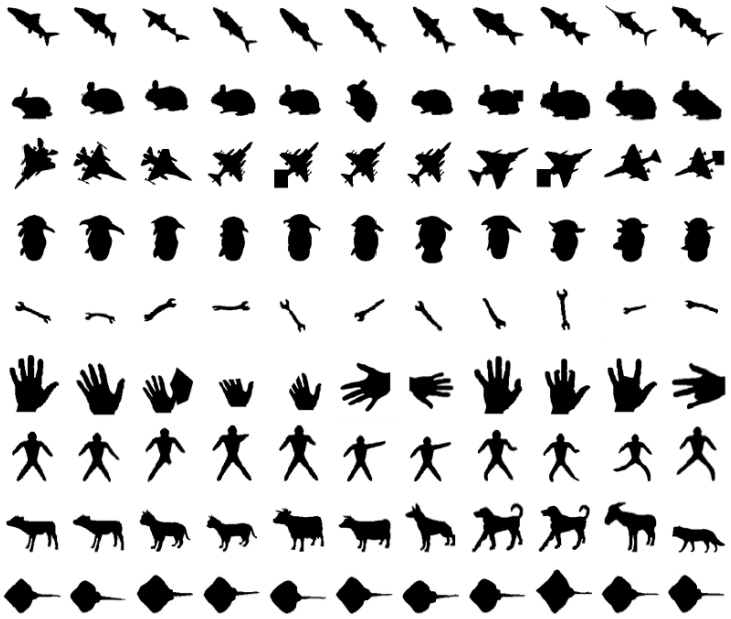

Fig. 7. A data set of 99 shapes (Sebastian et al., 2001)

The experiments were performed on a Pentium IV 2.6 GHz PC. Each shape was matched against all the other shapes of the data set and the number of times the test image was correctly classified was counted in the nth nearest neighbours (n ranging from 1 to 8) (Tabbone et al., 2006). Fig. 8 shows the nth nearest match rates for each approach. Although the results of the first nearest matches were quite similar among all methods, the results for the matches from 5 to 8 were better with our approach. Finally, it must be mentioned that these results are quite similar to the ones reported by Tabbone et al. (2006) which use a more computationally expensive shape descriptor defined on the Radon transform.

[1] http://www.lems.brown.edu/vision/researchAreas/SIID/

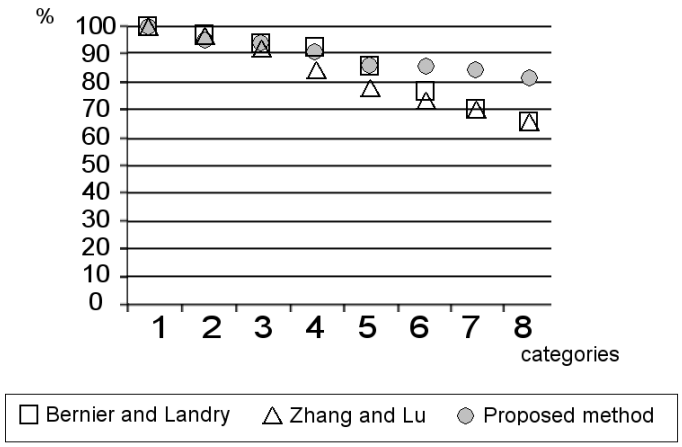

Fig. 8. Comparison of the employed shape descriptor with other approaches (see text)

5.2. Scene recognition experiments

Once the curvilinear regions have been detected, they 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 image co-ordinates are ranged from 0 to 256), 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 $\{fc_i\}_{i=1...256}$ are the curvature function of the object shape. Each image is then described by the properties of the associated set of curvilinear regions.

In this image matching scheme, two images will be similar if their associated sets of curvilinear regions are similar. The distance between two curvilinear regions i and j can be defined as

$$D(i,j) = \alpha_1 \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + \alpha_2 \sqrt{s_i^2 + s_j^2 - 2 s_i s_j \cos\theta} + \alpha_3 \max\{(fc_i * fc_j)_{k=1...256}\} \tag{18}$$

where θ is equal to $|h_i\text{-}h_j|$ if this value is less than π, or equal to $(2\pi - |h_i\text{-}h_j|)$ in any other case. The parameters a_i define the importance of the position, colour and shape into the distance measure and they have been experimentally adjusted. The * operator denotes the convolution and it is applied ranging from 1 to 256, providing rotation invariance. Then, given a query image Q and a dataset of images B_i, whose associated sets of curvilinear regions have been detected and characterised off-line, the image matching process firstly extracts the set of N_Q curvilinear regions $\{cQ\}_{i=1...NQ}$ present in the query image. They are sorted as a function of their lengths. Then, the comparison between Q and each image B_i is achieved by comparing each curvilinear region in Q, cQ_i, with all the N_{Bi} curvilinear regions present in B_i, $\{cBi\}_{i=1...NBi}$, using (18). The most similar region is selected and, if the similarity value, $D(cQ_i,cBi_j)$, is less than a given threshold U, both curvilinear regions are paired. This implies that the selected curvilinear region of B_i cannot be paired with other curvilinear region of Q. Finally, a similarity value is assigned to the comparison between images Q and Bi. This value is defined as

$$\lambda = \left(N_{Bi} - N'_Q + 1\right) \sum_{i=1}^{N'_Q} D(cQ_i, cBi_j) \tag{19}$$

where N'_Q is the number of paired curvilinear regions.

The images B_i are then sorted according to the obtained similarity values. To test the method, a database of 40 images obtained in an office-like environment has been created. This database can be divided into 10 different scenarios (4 different images for each scenario). Fig. 9 presents two example retrievals for this database. Query is the leftmost image in each row, and subsequent images are nearest neighbours. Detected curvilinear regions employed to match both images have been marked.

To evaluate the matching performance, we have employed the normalised average rank $\overline{R}$ (Grauman & Darrell, 2005)

$$\overline{R} = \frac{1}{NN_R}\left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R - 1)}{2}\right) \tag{20}$$

where 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. A normalised average rank equal to 0 implies a perfect performance, that is all relevant images in the database have been retrieved as nearest neighbours of the query image. For the reported experiment, the normalised average rank of relevant images present an average value of 0.025 and a standard deviation of 0.001.

Fig. 9. Example retrievals for a database of office-like environment images (see text for details)

6. Conclusions and future work

This chapter presents a method for image matching which is based on the detection and characterisation of curvilinear regions. In a curvilinear region, the ratio between its average width and its total length should be less than a predefined threshold. Besides, left and right borders should be locally parallel, it should exist a geometric similarity around the region axis and the colour along this axis should be homogeneous. That is, they constitute particular image structures and, therefore, the method is restricted to scenes where these particular items are presented. On the contrary, curvilinear regions automatically deform with changing viewpoint as to keep on covering identical physical parts of a scene. For this reason, they can be used as distinguished regions. The shape contour of these curvilinear regions is characterised using the adaptive curvature function. Experimental results show that this shape descriptor is invariant to rotation and translation, and partially invariant to noise and skewing. Scaling invariance is achieved by employing an extra normalisation stage. Thus, this descriptor, plus the region colour and position, can be used to match curvilinear regions detected on the input image with those previously stored in a database. This is the basis of the correspondence algorithm described in this paper: the similarity index between two images is determined by the presence of the same set of curvilinear regions localised in similar positions.

There are many directions for further research. One of this is the integration of several types of distinguished regions. As we commented above, the obligatory presence of these regions in the images is the main disadvantage of the proposed system. Besides, further work must be accomplished in the correspondence algorithm to employ the most similar region correspondences as ground control points. These points could be used to generate a fundamental matrix describing the geometric constraints between the two images. Thus, matches that did not agree with the majority solution could be removed.

7. Acknowledgments

This work has been partially granted by the Spanish Ministerio de Ciencia y Educación (MEC), under project nº TIN2005-01359.

8. References

Bala, E. & Cetin, A.E. (2004). Computationally efficient wavelet affine invariant functions for shape recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 8, 1095-1098

Bandera, A.; Urdiales, C.; Arrebola, F. & Sandoval, F. (2000). Corner detection by means of adaptively estimated curvature function. *Electronics Letters*, Vol. 36, No. 2, 124-126

Baumberg, A. (2000). Reliable feature matching across widely separated views. *Proc. of the Conference on Computer Vision and Pattern Recognition*, 774-781

Bernier, T. & Laundry, J.A. (2003). A new method for representing and matching shapes of natural objects. *Pattern Recognition*, Vol. 36, 1711-1723

Beus, L. & Tiu, S. (1987). An improved corner detection algorithm based on chain-coded plane curves. *Pattern Recognition*, Vol. 20, No. 3, 291-296

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, Vol. 94, No. 2, 115-147

Bister, M.; Cornelis, J. & Rosenfeld, A. (1990). A critical view of pyramid segmentation algorithms. *Pattern Recognition Lett.*, Vol. 11, 605–617

Brun, L. & Kropatsch, W.G. (2003). Construction of combinatorial pyramids, in: *Graph Based Representations in Pattern Recognition,* E. Hancock & M. Vento (Eds.), Vol. 2726, 1-12

Carneiro, G. & Jepson, A.D. (2002). Phase-based local features. *Proc. of the European Conference on Computer Vision (ECCV)*, 282-296

Crowley, J. L. & Parker, A.C. (1984). A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 2, 156-170

Dai, X.L. & Lu, J. (1999). An object-based approach to automated image matching. *Proc. of the IEEE Int. Conf. on Geoscience and Remote Sensing Symposium,* Vol. 2, 1189-1191

Deng, H.; Zhang, W.; Dietterich, T. & Mortensen, E. (2006). *A comparative evaluation of a new curvilinear region detector for object recognition*. Technical Report, Electrical Engineering and Computer Science, Oregon State University

Dufournaud, Y.; Schmid, C. & Horaud, R. (2000). Matching image with different resolutions. *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 612-618

Freeman, H. (1978). Shape description via the use of critical points. *Pattern Recognition*, Vol. 10, 159-166

Grauman, K. & Darrell, T. (2005). Efficient image matching with distributions of local invariant features. *Proc. of IEEE Conf. Computer Vision and Pattern Recogn.,* 627-634

Harris, C. & Stephens, M. (1988). A combined corner and edge detector. *Proc. of the Fourth Alvey Vision Conference*, 147-151

Haxhimusa, Y.; Glantz, R. & Kropatsch, W.G. (2003) Constructing stochastic pyramids by MIDES—maximal independent directed edge set, in: *Fourth IAPR-RC15 Workshop on GbR in Pattern Recognition,* E. Hancock & M. Vento (Eds.), Vol. 2726, 35-46

Jolion, J.M. & Montanvert, A. (1992). The adaptive pyramid, a framework for 2D image analysis. *CVGIP: Image Understanding*, Vol. 55, 339–348

Klette, G. (2003). A comparative discussion of distance transformations and simple deformations in digital image processing. *Machine Graphics & Vision*, Vol. 12, No. 2, 235-256

Lindeberg, T. (1993). Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, Vol. 11, No. 3, 283-318

Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, Vol. 21, No. 2, 224-270

Lowe, D.G. (1999). Object recognition from local scale-invariant features. *Proc. of the International Conference on Computer Vision*, 1150-1157

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, Vol. 60, No. 2, 91-110

Marfil, R.; Rodríguez, J.A.; Bandera, A. & Sandoval, F. (2004). Bounded irregular pyramid: a new structure for color image segmentation. *Pattern Recognition*, Vol. 37, No. 3, 623-626

Matas, J.; Chum, O.; Urban, M. & Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. *Proc. of the British Machine Vision Conf.*, 384-393

Mikolajczyk, K. & Schmid, C. (2001). Indexing based on scale invariant interest points. *Proc. of the 8th Int. Conf. on Computer Vision*, 525-531

Mikolajczyk, K. (2002). *Detection of local features invariant to affine transformations*, Ph.D. thesis, Institut National Polytechnique de Grenoble, France.

Mikolajczyk, K.; Zisserman, A. & Schmid, C. (2003). Shape recognition with edge-based features. *Proc. of the British Machine Vision Conference*, 799-788

Mikolajczyk, K. & Schmid, C. (2004). Scale and affine invariant interest point detectors. *Int. Journal on Computer Vision*, Vol. 1, No. 60, 63-86

Mikolajczyk, K. & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, 1615-1630

Mikolajczyk, K.; Tuytelaars, T.; Schmid, C.; Zisserman, A.; Matas, J.; Schaffalitzky, F.; Kadir, T. & Van Gool, L. (2005). A comparison of affine region detectors. *Int. Journal on Computer Vision*, Vol. 65, No. 1/2, 43-72

Mokhtarian, F. & Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Analysis and Machine Intell.*, Vol. 8, No. 1, 34-43

Montesinos, P.; Gouet, V. & Pele, D. (2000). Matching color uncalibrated images using differential invariants. *Image and Vision Computing*, Vol. 18, No. 9, 659-671

Moravec, H. (1981). Rover visual obstacle avoidance. *Proc. of the Int. Joint Conference on Artificial Intelligence*, 785-790

Nelson, R.C. & Selinger, A. (1998). Large-scale tests of a keyed, appearance-based 3-D object recognition system. *Vision Research*, Vol. 38, No. 15, 2469-2488

Pope, A.R. & Lowe, D.G. (2000). Probabilistic models of appearance for 3-D object recognition. *International Journal of Computer Vision*, Vol. 40, No. 2, 149-167

Schiele, B. & Crowley, J.L. (2000). Recognition without correspondence using multi-dimensional receptive field histograms. *International Journal of Computer Vision*, Vol. 36, No. 1, 31-50

Schmid, C. & Mohr, R. (1997). Local gray value invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vo. 19, No. 5, 530-534

Sebastian, T.; Klein, P. & Kimia, B. (2001). Recognition of shapes by editing shock graphs. *Proc. of the ICCV'2001*, 755-762

Shokoufandeh, A.; Marsic, I. & Dickinson, S.J. (1999). View-based object recognition using saliency maps. *Image and Vision Computing*, Vol. 17, 445-460

Tabbone, S.; Wendling, L. & Salmon, J.P. (2006). A new shape descriptor defined on the Radon transform. *Computer Vis. Image Understanding*, Vol. 102, 42-51

Tuytelaars, T. & Van Gool, L. (2004). Matching widely separated views based on affine invariant regions. *Int. Journal of Computer Vision*, Vol. 59, No. 1, 61-85

Willersinn, D. & Kropatsch, W.G. (1994). Dual graph contraction for irregular pyramids. *Proc. of the 12th IAPR International Conference on Pattern Recognition*, Vol. 3, 251–256

Zhang, Z.; Deriche, R.; Faugeras, O. & Luong, Q.T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, Vol. 78, 87-119

Zhang, D. & Lu, G. (2005). Study and evaluation of different Fourier methods for image retrieval. *Image and Vision Computing*, Vol. 23, 33-49

Vision Systems: Segmentation and Pattern Recognition

Edited by Goro Obinata and Ashish Dutta

ISBN 978-3-902613-05-9
Hard cover, 536 pages
Publisher I-Tech Education and Publishing
Published online 01, June, 2007
Published in print edition June, 2007

Research in computer vision has exponentially increased in the last two decades due to the availability of cheap cameras and fast processors. This increase has also been accompanied by a blurring of the boundaries between the different applications of vision, making it truly interdisciplinary. In this book we have attempted to put together state-of-the-art research and developments in segmentation and pattern recognition. The first nine chapters on segmentation deal with advanced algorithms and models, and various applications of segmentation in robot path planning, human face tracking, etc. The later chapters are devoted to pattern recognition and covers diverse topics ranging from biological image analysis, remote sensing, text recognition, advanced filter design for data analysis, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

J. Perez Lorenzo, R. Vazquez Martin, R. Marfil, A. Bandera and F. Sandoval (2007). Image Matching based on Curvilinear Regions, Vision Systems: Segmentation and Pattern Recognition, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-05-9, InTech, Available from: http://www.intechopen.com/books/vision_systems_segmentation_and_pattern_recognition/image_matching_based_on_curvilinear_regions

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821