

# TEXTUAL DESCRIPTION OF IMAGES

Larabi, S.

Computer Science Department, Faculty of Electronic and Computer Science,  
University of Science and Technology Houari Boumedienne, BP 32, El Alia, Algiers, Algeria  
E-Mail: [slarabi@usthb.dz](mailto:slarabi@usthb.dz)

## Abstract

We present in this paper a new method for image description based on the description of encompassed regions. Image is firstly segmented in meaningful regions and in the next step these regions are represented semantically in a tree structure. Regions are described by means the geometry of their outline contours and the description of encompassed internal regions. An XML language XLWDI (XML Language for Writing Descriptors of Images) is proposed to write these descriptions. This method may be used for image coding in multimedia applications or for image description in computer vision applications for well contrasted images. Experiments conducted over real images are presented and explained.

(Extended paper from the International Symposium CompIMAGE – Coimbra, Portugal, 20-21 October 2006.)

**Key Words:** Image, Region, Silhouette, Descriptor, XML Language

## 1. INTRODUCTION

Few methods have been proposed for the description of images, the most known is the method used in the MPEG-7 standard where image considered as initial region is decomposed into individual regions for which are associated the feature type, color histogram and textual annotation. The set of these regions are organized in tree structure associated to the image. Other known method is the SVG format that describes synthetic images with text using an XML format. Image for this format is a set of specific 2D primitives (circle, Bezier curve, rectangle...).

In case where image contains only one object, various methods have been proposed to describe the image in an abstract and efficient way while still preserving important shape feature. The most interesting methods may be classified as follow:

- Part-based methods [1, 2, 3, 4, 5]
- Aspect-graph methods [6, 7]
- Methods that use the medial axis of silhouettes [8, 9, 10]
- Methods based on the shock graph [11, 12]
- Methods using graph for shape representation [13, 14]
- Approximation of outline shape by 2D features [15, 16, 17, 18]
- Methods based on the reference points of outline shape [19, 20, 21, 22]
- Methods based on the attributes of outline shape [23, 24, 25, 26]
- Method based on the shape context [27]
- Appearance-based methods [28]

A review of shape representation methods can be found in [29, 30].

In order to homogeneous all databases of image descriptions and to offer accessibility to all users, a common and compact format of representation is recommended. In addition, this format must be easy to index, easy to compare and efficient for computation and storage. Our aim in this paper is to propose a new image descriptor verifying the above recommendations.

Our approach for image description is based on the part-based methods for silhouette description, shape description [31, 32] and on the basic principle of image representation

proposed by Z. Tu et al [33]. Image is firstly segmented into a set of regions corresponding either to shape of object or to area in the viewed world. Each region is described by means the geometry of its outline contours and the description of encompassed internal regions. The tree structure corresponding to the image is translated to a textual description written with an XML language XLWDI (XML Language for Writing Descriptors of Images).

The proposed description may be applied to images that may be segmented correctly as well contrasted images, industrial logos images, and synthetic images and provide the following properties:

- Coding of images with text following XML language
- Reduced size of computed description
- Easiness for index extraction
- Preservation of perceptual structure
- Complete storage of all object information
- Coding and visualization of images in multimedia applications.

This paper is organized as follows: Section 2 describes the basic principle of our method for image description. In section 3, we present the method for outline shape description. In section 4 we discuss results of experiments conducted over images demonstrating the usefulness of this description.

## **2. BASIC PRINCIPLE OF IMAGE DESCRIPTION**

We assume that image is segmented into regions using the one of known methods [34, 35]. This task is not discussed in this paper. Any image is characterized by the list of all regions that contains the inclusion relation between regions, the position of each region in the image and the description of each region (the geometry of its outline contour and its colour).

Having these attributes, we can represent image with a tree structure where the node root corresponds to the image, nodes correspond to regions and children nodes of any node correspond to the set of regions that encompasses (see Fig. 1).

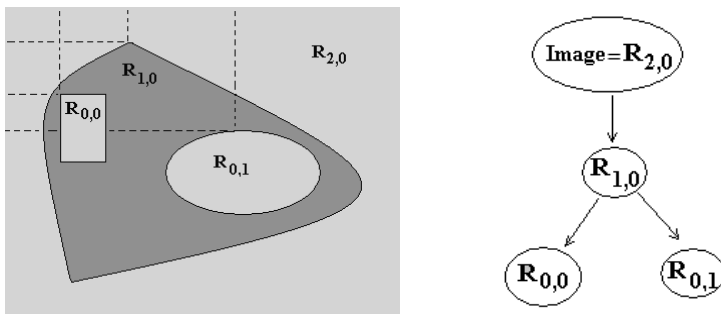


Figure. 1. Different regions of image and the associated tree structure

We associate for each region:

- the level  $i$  corresponding to its position in the tree beginning from the bottom to the up. This value begins from zero.
- the order number  $j$  of the region having the same level. This value beginning from 0 is given to regions located sweeping top-bottom and left-right the image.

Thus, the level zero (0) is given to region that not contains internal regions. The level one (1) is given to region that contains regions of level 0. In general case, the level ( $n$ ) is associated to regions encompassing regions of level less or equal to ( $n-1$ ) (see Fig. 1).

The tree structure can be automatically achieved starting from a segmented image, by the location on the image all elementary regions that are regions not encompassing other internal

regions. For each of these regions we associate the level 0. The next step is the location of regions of level 1 that are regions encompassing regions of level 0. We repeat this treatment in order to locate regions of level  $i$  corresponding to regions that encompass regions of level less or equal to  $(i-1)$ . An order number is given for regions of the same level starting from zero by the sweep of the image from the top to bottom and from the left to the right. The image of figure 1 contains for example two regions of level 0 ( $R_{0,0}$ ,  $R_{0,1}$ ) and one region of level 1 ( $R_{1,0}$ ) encompassing  $R_{0,0}$ ,  $R_{0,1}$  and one region  $R_{2,0}$  of level 2 containing  $R_{1,0}$ .

In case where regions of level less or equal to  $(n-1)$  are neighbouring one to other, a new region of level  $(n)$  is created and added in the tree as the union of these regions. Fig. 2 illustrates an example where the new region  $R_{1,0}$  is the region obtained as the union the neighbouring regions  $R_{0,0}$ ,  $R_{0,1}$  and  $R_{0,2}$ . Fig. 3 illustrates the different steps for the tree structure construction.

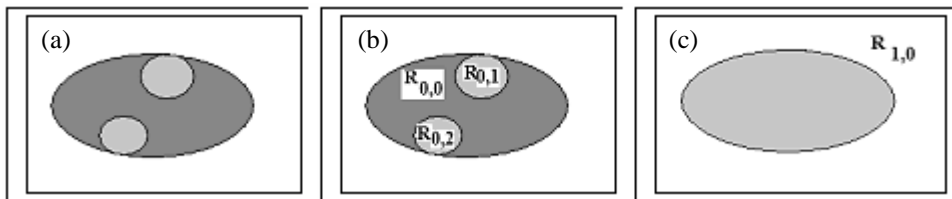


Figure 2: Example of region of level 1 that is the union of neighboring region of level 0 – (a) initial image, (b) regions of level 0, (c) regions of level 1.

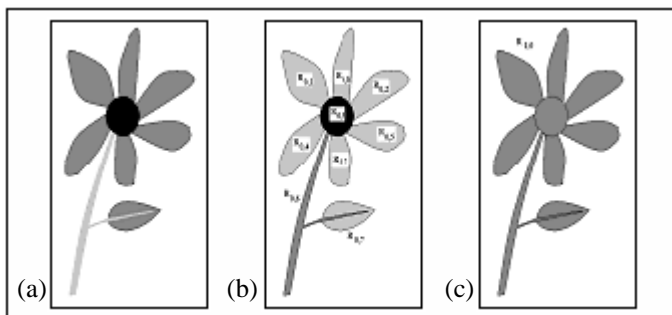


Figure 3: (a) Different regions of the image, (b) regions of level 0, (c) regions of level 1.

To achieve the image representation, we must characterize each region by the geometrical description of its outline and its position expressed by the  $x$  and  $y$ -coordinates of its left highest point.

### **3. DESCRIPTION OF REGIONS**

#### **3.1 Description of outline regions**

To describe outline regions we use the part-based method of Larabi et al [31] whose basic principle is presented in follow:

When we sweep up silhouette from the top to bottom, it may be locate concave points for which the direction of the outer contour changes following top-bottom-top or bottom-top-bottom (see Fig. 4). Silhouette is decomposed at these points onto parts, junction and disjunction lines: either, two parts or more are joined with a third part through a junction line, or a part is joined with two parts or more through a disjunction line. This process applied to silhouette of Fig. 4 produces five parts, one junction and one disjunction line.

Description of outline region is the grouping of descriptions of its elements.

Part is defined by its two boundaries (left and right) which begin at the highest left point and terminate at their lowest points. Using the inflection points that may be located using one of the known algorithms [36], these boundaries are segmented into set of primitives (line, convex and concave contours) and described by the parameters: type (line, convex or concave curve), degree of concavity or convexity, angle of inclination and length (see Fig. 5).

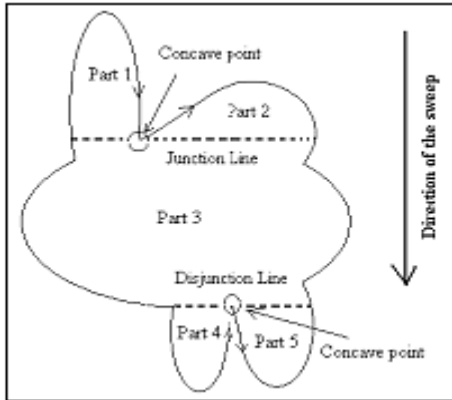


Figure 4: Partition of region into parts, junction and disjunction lines.

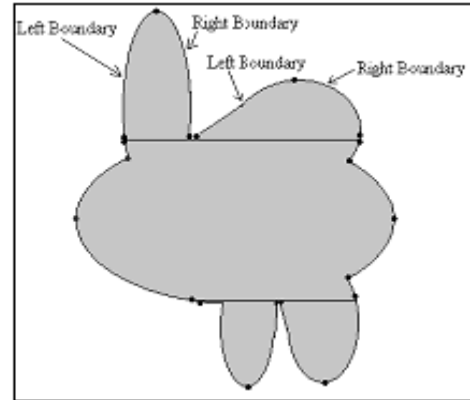


Figure 5: The beginning points, end points and the inflexion points of the boundaries (left and right) of each part.

Junction and disjunction lines are decomposed onto segments. Each segment is described with three parameters: type, the reference numbers of parts where it appertains and its length. Types of segment are: **Junction** if it is common for two parts, **Free-High** if it belongs only to the high part or **Free-Low** if it belongs to the low part (see Fig. 6).

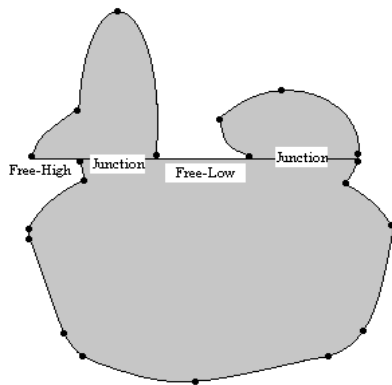


Figure 6: Description of junction and disjunction lines.

From this description of the outline region, it is easy to draw the correspondent region without ambiguity. The use of absolute lengths produces an absolute description of regions. Such description is recommended for coding images in multimedia applications. For computer vision applications as recognition of objects, relative lengths may be used to guaranty the scalability property.

### 3.2 The XML Language for writing descriptors of outline of regions

The proposed XML language for writing descriptors of outline regions is presented here. Firstly we give the set of rules allowing the writing of description of parts, junction and disjunction lines.

Part → <P num=x>Left boundary </L> <R> Right boundary </R ></P num=x>  
 LeftBoundary → <L>Contour 1... Contour n </L>  
 RightBoundary → <R>Contour 1... Contour n </R>  
 Contour i → <CV degree=x Inclination=x Length=x/> /  
                   <CC degree=x Inclination=x Length=x/> /  
                   <LN Inclination=x angle=x Length=x/>

where CV, CC, LN indicate respectively convex, concave, and line contour.

Junction line → <J> Segment 1 .... Segment n </J>  
 Disjunction line → <D> Segment 1 ...Segment n </D>  
 Segment i → <S Numpart1=x Numpart2=x Length=x /> /  
                   <W Numpart1=x Length=x /> /  
                   <H Numpart1=x Length=x />

where S, W and H denote respectively the attributes: **Shared**, **FreeLow** and **FreeHigh**.

The composition of different elements of the silhouette produces the **composed part** defined as the set of two (or more) parts joined to another part using junction or disjunction line, and written as follow:

Composed Part → <CP> P<sub>1</sub> P<sub>2</sub> .... P<sub>n-1</sub> <J> Junction line </J> P<sub>n</sub> </CP>  
 Composed Part → <CP> P<sub>1</sub> <D> Disjunction line </D> P<sub>2</sub> P<sub>3</sub> ..... P<sub>n</sub> </CP>

Recursively, a composed part is considered as a part and can constitutes with other elements other composed part.

The write the descriptor of silhouettes we use:

*Silhouette* → <DLWDOS><Name=Object name/> Composed Part </DLWDOS>

#### 4. LANGUAGE FOR WRITING DESCRIPTORS OF IMAGES

The tree structure proposed to represent images may be translated to XML file. For this we propose in this section the grammar of the language XLWDI (XML Language for Writing Descriptors of Images) that allow writing the description of images following XML syntax.

Description of Image → <XDLWDI>

    <Name=Name-of-image/>  
     ¿Description of initial region!  
     </XDLWDI>

¿Description of initial region! → ¿Description of Region! ¿Description of Initial Region! /  
   ¿Description of Region!

¿Description of Region! → <Region>¿Position of Outline Region!  
   ¿Description of Outline Region!  
   ¿Description of Internal Regions!  
   </Region>

¿Description of Internal Regions! → <Internal Region>  
   ¿Position of Outline Internal Region!  
   ¿Description of Outline Internal Region!  
   ¿Description of Internal Regions!  
   </Internal Region>

¿Position of Outline Region! → ¿X-Coordinate of Begin Point! ¿Y-Coordinate of Begin Point!

¿Description of Outline Internal Region! → ¿Description of Silhouette!

¿Description of Outline Region! → ¿Description of Silhouette!

¿X-Coordinate of Begin Point! → INTEGER VALUE

¿Y-Coordinate of Begin Point! → INTEGER VALUE

For example, the image of Fig. 7, contains two regions of level 0: R<sub>0,0</sub> and R<sub>0,1</sub> and one region of level 1: R<sub>1,0</sub>.

The descriptor of this image is written as follow (in bold character are written the non terminal):

```
<DXLWDI><Name=figure 7/>
<Region>111 127 Description of R1,0
<Internal Region>123 174 Description of R0,0 </Internal Region>
<Internal Region>124 143 Description of R0,1 </Internal Region></ Region>
</DXLWDI>
```

where (111,127), (123,174), (124,143) are respectively the x-y coordinates of  $R_{0,0}$ ,  $R_{0,1}$  and  $R_{1,0}$ .

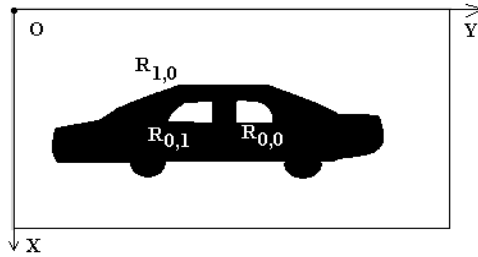


Figure 7: Image of a car.

## 5. EXPERIMENTS

Images used for the validation of our method are images of industrial logos and real images. Each image is segmented using JSEG method [34]. The processing of the segmented image produces a set of regions that are represented by a tree structure of regions.

Fig. 8 illustrates an example of real image, the segmented image and the tree structure associated where  $S_{3,0}$  corresponding to the entire image is composed by the region  $S_{0,0}$  and the  $S_{2,0}$ .  $S_{2,0}$  is contains the region  $S_{1,0}$  composed by three elementary regions:  $S_{0,1}$ ,  $S_{0,2}$  and  $S_{0,3}$ .

A textual descriptor is computed for each image as described in section 4. The descriptor of each region (elementary or composed) of the image is computed firstly using the Chetverikov's algorithm [36] in order to locate the curvature points. Two parameters are necessary: the angle and distance. Fig. 9 illustrates the result of the  $S_{0,1}$  partitioning into parts, junction and disjunction lines.

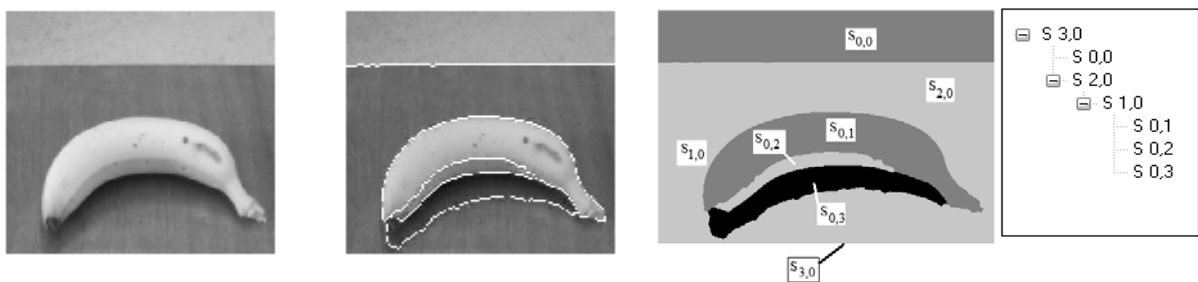


Figure 8: Segmentation and representation by a tree structure of banana image.

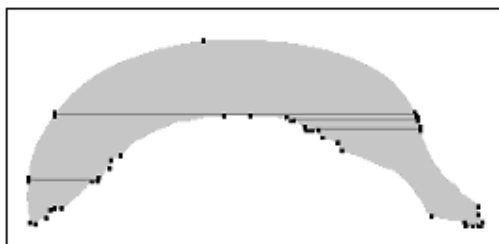


Figure 9: Example of computation of silhouette descriptor.

The computed descriptor for image includes all descriptors of internal regions and their (x,y) positions (see Fig. 10).

Images used for the validation of our method are images of a industrial logos and real images. Fig. 11 illustrates the developed software “Multimedia-XLWDI” that allows to convert an image giving the result of its segmentation into regions (two images: the map image and the segmented image) into a textual description. Each image is segmented using JSEG method [34]. The processing of the segmented image produce its representation by a tree structure of regions where to each region is associated a unique color.

A textual descriptor is computed for each image as described in section 4. The computed descriptor includes all descriptors of internal regions. Fig. 12 illustrates for example the result of the XLWDOS description of the region  $S_{1,1}$ .

```

<DXLWDOS>
<Name>S01</Name>
<CP>
<P1>
<L>CV 9 23 46</L>
<R>CV 12 164 46</R>
</P1>
<D1>W P2 1 5 P1 P2 123 H P: 18 5 P1 P3 122 W P3 1</D1>
<CP>
<P2>
<L>CV 7 65 41</L>
<R>CC 11 19 27 CC 13 23 4 R 79 6 R 41 7</R>
</P2>
<D4>S P2 P9 47 H P2 3 S P2 P10 1 H P2 2</D4>
<P9>
<L>CV 1 92 27 CV 24 166 1</L>
<R>CV 5 38 18 R 0 6 R 18 2 R 59 6 CC 25 27 5</R>
</P9>
<P10>
<L>R 90 1</L>
<R>R 90 1</R>
</P10>
</CP>
<CP>
<P3>
<L>CC 5 177 2</L>
<R>R 90 2</R>
</P3>
<D2>H P3 9 S P3 P4 1 H P1 1 S P3 P5 90 W P5 1</D2>
<P4>
<L>R 90 1</L>
<R>R 90 1</R>
</P4>
<P5>
<L>CC 12 150 5</L>
<R>R 117 5</R>
</P5>
</CP>
</DXLWDOS>
    
```

Figure 10: The computed descriptor for the region  $S_{0,1}$ .

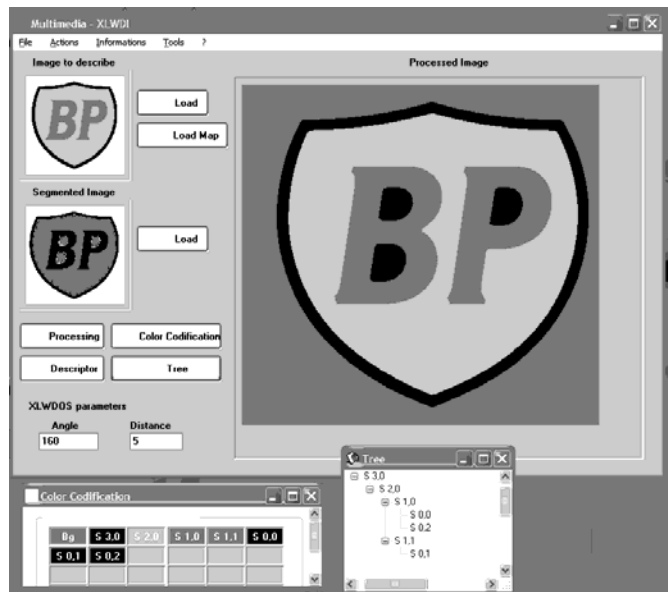


Figure 11: The developed software Multimedia-XLWDI and the tree structure associated for BP logo.

Same results are obtained for Esso logo, Peugeot logo and banane images. For each region that corresponds to the set of neighboring regions, a new region is created and the black color it will be affected.

For each image we give in table 1 the size of the computed XLWDI descriptor and the size of the different formats. We can see that in addition of the semantic contained in the XLWDI description, the correspondent size is also interesting.

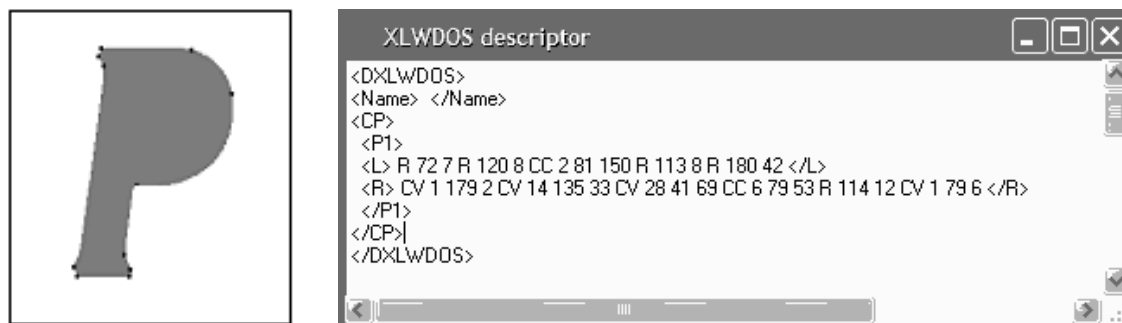


Figure 12: The computation of  $S_{1,1}$  description.

Table I: Comparison of the size of different formats.

	<i>XLWDI description</i>	<i>Format BMP</i>	<i>Format GIF</i>	<i>Format JPG</i>	<i>Scalar Format</i>
Logo of Esso	3 106 bytes	781 254 bytes	10 862 bytes	25 725 bytes	21 995 bytes
Logo of Peugeot	9 285 bytes	660 054 bytes	41 931 bytes	25 326 bytes	15 490 bytes
Logo of BP	1 992 bytes	554 454 bytes	12 837 bytes	57 232 bytes	21 667 bytes
Banane	4 002 bytes	330 054 bytes	10 052 bytes	9 464 bytes	---

## **6. CONCLUSION**

In this paper we have presented a new method for image representation and description. We presented also the XML language noted XLWDI for the writing with XML format the proposed description.

The proposed method can be also used to describe image for computer vision applications. In this case, regions must be described relatively to the minimum rectangle encompassing it in order to guaranty the invariance to rotation of shape of this description.

Another interesting application of our method is the coding of images using the XLWDI language. Current works are focused for the development of standard libraries to convert to and from segmented image the XML file format.

## **REFERENCES**

- [1] Kim, D. H.; Yun, I. D.; Lee, S. U. (2005). A new shape decomposition scheme for graph-based representation, *Pattern Recognition*, Vol. 38, 673-689
- [2] Pitas, I.; Venetsanopoulos, A. N. (1990). Morphological shape decomposition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 38-45
- [3] Rosin, P. L. (1999). Shape partitioning by convexity, British Machine Vision Conference
- [4] Siddiqi, K.; Kimia, B. B. (1995). Parts of visual form computational aspects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 3, 239-251
- [5] Yu, L.; Wang, R. (2004). Shape representation based on mathematical morphology, *Pattern Recognition Letters*
- [6] Cyr, C. M.; Kimia, B. B. (2004). A similarity-based aspect-graph approach to 3D object recognition, *International Journal of Computer Vision*, Vol. 57, No. 1, 5-22
- [7] Koenderink, J. J.; Doorn, V. (1976). The internal representation of solid shape with respect to vision, *Biological Cybernetics*, Vol. 32, No. 4, 211-216
- [8] Geiger, D.; Liu, T.; Kohn, R. V. (2003). Representation and Self-Similarity of shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 1
- [9] Ruberto, C. (2004), Recognition of shapes by attributed skeletal graphs, *Pattern Recognition*, Vol. 37, 21-31
- [10] Zhu, S. C.; Yuille, A. L. (1995). FORMS: A flexible object recognition and modelling system, *Proc. of the Fifth International Conference on Computer Vision*, M.I.T. Cambridge
- [11] Siddiqi, K.; Kimia, B. B. (1996). A shock grammar for recognition, *Conference of Computer Vision and Pattern Recognition*
- [12] Sebastian, T. B.; Klein, P. N.; Kimia, B. B. (2004). Recognition of Shapes by Editing Their Shock Graphs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5
- [13] Badawy, O. E.; Kamel, M. (2002). Shape representation using concavity graphs, *Proceedings of the 16<sup>th</sup> International Conference on Pattern Recognition*
- [14] Lourens, T.; Wurtz, R. P. (1998). Object recognition by matching symbolic edge graphs, *Proceedings of ACCV*, 193-200

- [15] Cronin, T. M. (2003). Visualizing concave and convex partitioning of 2D contours, *Pattern Recognition Letters*, Vol. 24, 429-443
- [16] Grosky, W. I.; Mehrota, R. (1990). Index-based recognition in pictorial data management, *Computer Vision Graphics and Image Processing*, Vol. 52, 416-436
- [17] Nelson, R. C.; Selinger, A. (1998). Large-scale tests of a keyed, appearance-based 3-D object recognition system, *Vision Research*, Vol. 38, No. 15-16
- [18] Petrakis, E. G. M.; Diplaros, A.; Milios, E. (2002). Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 11
- [19] Berretti, S.; Bimbo, A. (2000). Retrieval by shape similarity with perceptual distance and effective indexing, *Transactions on Multimedia*, Vol. 2, No. 4
- [20] Mokhtarian, F.; Mackworth, A. K. (1992). A Theory of Multiscale, Curvature-Based Shape Representation for Planar curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 8, 789-805
- [21] Mokhtarian, F. (1995). Silhouette-Based isolated object recognition through curvature scale space, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 5, 539-544
- [22] Shao, Z.; Kittler, J. (1999). Shape representation and recognition based on invariant unary and binary relations, *Image and Vision Computing*, Vol. 17, 429-444
- [23] Arica, N.; Vural, F. T. Y. (2003). BAS: a perceptual shape descriptor based on the beam angle statistics, *Pattern Recognition Letters*, Vol. 24, No. 9-10, 1627-1639
- [24] Bernier, T.; Landry, J. A. (2003). A new method for representing and matching shapes of natural objects, *Pattern Recognition*, Vol. 36, 1711-1723
- [25] Orrite, C.; Herreo, J. E. (2004). Shape matching of partially occluded curves invariant under projective transformation, *Computer Vision and Image Understanding*, Vol. 93, No. 1
- [26] Sethi, A.; Renaudie, D.; Kriegman, D; Ponce, J. (2004). Curve and surface duals and the recognition of curved 3D objects from their silhouette, *International Journal of Computer Vision*, Vol. 58, No. 1, 73-86
- [27] Belongie, S., Malik, J.; Puzicha, J. (2002). Shape Matching and Object Recognition Using Shape Contexts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24
- [28] Murase, H.; Nayer, S. K. (1995). Visual learning and recognition of 3-D objects from appearance, *International Journal of Computer Vision*, Vol. 14, No. 1
- [29] Campbell, R. J.; Flynn, P. J. (2001). A Survey of Free-Form Object Representation and Recognition Techniques, *Computer Vision and Image Understanding*, Vol. 81, 166-210
- [30] Zhang, D.; Lu, G. (2004). Review of shape representation and description techniques, *Pattern Recognition*, Vol. 37, 1-19
- [31] Larabi, S.; Bouagar, S.; Trespaderne, F. M.; Fuente, E. L. (2003). LWDOS: Language for Writing Descriptors of Outline Shapes, *Lecture Notes in Computer Science, Proceedings of Scandinavian Conference on Image Analysis*, Vol. 2749, 1014 – 1021
- [32] Larabi, S. (2006). Textual description of images, *Proceedings of the Conference Computational Modeling of objects represented in Images, Fundamentals Methods and Applications*, Coïmbra, Portugal
- [33] Tu, Z.; Chen, X.; Yuille, A. L.; Zhu, S. (2005). Image parsing: Unifying Segmentation, Detection, and Recognition, *International Journal of Computer Vision*, Vol. 63, No. 2, 113-140
- [34] Deng, Y.; Majunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video, *IEEE Transactions on Pattern Analysis Machine intelligence*, Vol. 23, No. 8, 800-810
- [35] Felzenszwalb, P. F.; Huttenlocher, D. P. (2005). Pictorial structures for object recognition, *International Journal of Computer Vision*, Vol. 61, No. 1, 55-79
- [36] Chetverikov, D. (2003). A Simple and Efficient Algorithm for Detection of High Curvature Points in Planar Curves, *10<sup>th</sup> International Conference CAIP*