LISF: An Invariant Local Shape Features Descriptor Robust to Occlusion

Leonardo Chang[1,2], Miguel Arias-Estrada[1], L. Enrique Sucar[1] and José Hernández-Palancar[2]
[1]*Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE),*
Luis Enrique Erro No.1, C.P. 72840, Tonantzintla, Puebla, Mexico
[2]*Advanced Technologies Application Center (CENATAV), 7thA No. 21406, Playa, C.P. 12200, Havana, Cuba*

Keywords: Shape Matching, Invariant Shape Features, Shape Occlusion.

Abstract: In this work an invariant shape features extraction, description and matching method (LISF) for binary images is proposed. In order to balance the discriminative power and the robustness to noise and occlusion in the contour, local features are extracted from contour to describe shape, which are later matched globally. The proposed extraction, description and matching methods are invariant to rotation, translation, and scale and present certain robustness to partial occlusion. Its invariability and robustness are validated by the performed experiments in shape retrieval and classification tasks. Experiments were carried out in the Shape99, Shape216, and MPEG-7 datasets, where different artifacts were artificially added to obtain partial occlusion as high as 60%. For the highest occlusion levels the proposed method outperformed other popular shape description methods, with about 20% higher bull's eye score and 25% higher accuracy in classification.

1 INTRODUCTION

Shape descriptors have proven to be useful in many image processing and computer vision applications (e.g., object detection (Toshev et al., 2011) (Wang et al., 2012), image retrieval (Shu and Wu, 2011) (Yang et al., 2013), object categorization (Trinh and Kimia, 2011) (Gonzalez-Aguirre et al., 2011), etc.). However, shape representation and description remains as one of the most challenging topics in computer vision. The shape representation problem has proven to be hard because shapes are usually more complex than appearance. Shape representation inherits some of the most important considerations in computer vision such as the robustness with respect to the image scale, rotation, translation, occlusion, noise and viewpoint. A good shape description and matching method should be able to tolerate geometric intra-class variations, but at the same time should be able to discriminate from objects of different classes. Some other important requirements for a promising shape descriptor include: computational efficiency, compactness, and generality of applications.

In this work, we describe object shape locally, but global information is used in the matching step to obtain a trade-off between discriminative power and robustness. The proposed approach has been named Invariant Local Shape Features (LISF), as it extracts, describes, and matches local shape features that are invariant to rotation, translation and scale. LISF, besides closed contours, extracts and matches features from open contours making it appropriate for matching occluded or incomplete shape contours. Conducted experiments showed that while increasing the occlusion level in shape contour, the difference in terms of bull's eye score, and accuracy of the classification gets larger in favor of LISF compared to other state of the art methods.

The rest of the paper is organized as follows. Section 2 discusses some shape description and matching approaches. Section 3.1 presents the local shape features extraction method. The features descriptor is presented in Section 3.2. Its robustness and invariability to translation, rotation, scale, and its locality property are discussed in Section 3.3. Section 4 describes the proposed features matching schema. The performed experiments and discussion are presented in Section 5. Finally, Section 6 concludes the paper with a summary of our proposed methods, main contributions, and future work.

Chang L., Arias-Estrada M., Sucar L. and Hernández-Palancar J..
LISF: An Invariant Local Shape Features Descriptor Robust to Occlusion.
DOI: 10.5220/0004825504290437
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 429-437
ISBN: 978-989-758-018-5

2 RELATED WORK

Some recent works where shape descriptors are extracted using all the pixel information within a shape region include Zernike moments (Kim and Kim, 2000), Legendre moments (Chong et al., 2004), and generic Fourier descriptor (Zhang and Lu, 2002). The main limitation of region-based approaches resides in that only global shape characteristics are captured, without taking into account important shape details. Hence, the discriminative power of these approaches is limited in applications with large intra-class variations or with databases of considerable size.

Curvature scale space (CSS) (Mokhtarian and Bober, 2003), multi-scale convexity concavity (MCC) (Adamek and O'Connor, 2004) and multi-scale Fourier-based descriptor (Direkoglu and Nixon, 2011) are shape descriptors defined in a multi-scale space. In CSS and MCC, by changing the sizes of Gaussian kernels in contour convolution, several shape approximations of the shape contour at different scales are obtained. CSS uses the number of zero-crossing points at these different scale levels. In MCC, a curvature measure based on the relative displacement of a contour point between every two consecutive scale levels is proposed. The multi-scale Fourier-based descriptor uses a low-pass Gaussian filter and a high-pass Gaussian filter, separately, at different scales. The main drawback of multi-scale space approaches is that determining the optimal parameter of each scale is a very difficult and application dependent task.

Geometric relationships between sampled contour points have been exploited effectively for shape description. Shape context (SC) (Belongie et al., 2002) finds the vectors of every sample point to all the other boundary points. The length and orientation of the vectors are quantized to create a histogram map which is used to represent each point. To make the histogram more sensitive to nearby points than to points farther away, these vectors are put into log-polar space. The triangle-area representation (TAR) (Alajlan et al., 2007) signature is computed from the area of the triangles formed by the points on the shape boundary. TAR measures the convexity or concavity of each sample contour point using the signed areas of triangles formed by contour points at different scales. In these approaches, the contour of each object is represented by a fixed number of sample points and when comparing two shapes, both contours must be represented by the same fixed number of points. Hence, how these approaches work under occluded or uncompleted contours is not well-defined. Also, most of these kind of approaches can only deal with closed contours and/or assume a one-to-one correspondence in the matching step.

In addition to shape representations, in order to improve the performance of shape matching, researchers have also proposed alternative matching methods designed to get the most out of their shape representations. In (McNeill and Vijayakumar, 2006), the authors proposed a hierarchical segment-based matching method that proceeds in a global to local direction. The locally constrained diffusion process proposed in (Yang et al., 2009) uses a diffusion process to propagate the beneficial influence that offer other shapes in the similarity measure of each pair of shapes. (Bai et al., 2010) replace the original distances between two shapes with distances induced by geodesic paths in the shape manifold.

Shape descriptors which only use global or local information will probably fail in presence of transformations and perturbations of shape contour. Local descriptors are accurate to represent local shape features, however, are very sensitive to noise. On the other hand, global descriptors are robust to local deformations, but can not capture the local details of the shape contour. In order to balance discriminative power and robustness, in this work we use local features (contour fragments) for shape representation; later, in the matching step, in a global manner, the structure and spatial relationships between the extracted local features are taken into account to compute shapes similarity. To improve matching performance, specific characteristics such as scale and orientation of the extracted features are used. The extraction, description and matching processes are invariant to rotation, translation and scale changes. In addition, there is not restriction about only dealing with closed contours or silhouettes, i.e. the method also extract features from open contours.

The shape representation method used to described our extracted contour fragments is similar to that of shape context (Belongie et al., 2002). Besides locality, the main difference between these descriptors is that in (Belongie et al., 2002) the authors obtain a histogram for each point in the contour, while we only use one histogram for each contour fragment, i.e. our representation is more compact. Unlike our proposed method, shape context assumes a one-to-one correspondence between points in the matching step, which makes it more sensitive to occlusion.

The main contribution of this paper is a local shape features extraction, description and matching schema that i) is invariant to rotation, translation and scaling, ii) provides a balance between distinctiveness and robustness thanks to the local character of the extracted features, which are later matched using global information, iii) deals with either closed or open contours, and iv) is simple and easy to compute.

3 PROPOSED LOCAL SHAPE FEATURES DESCRIPTOR

Psychological studies (Biederman and Ju, 1988) (De Winter and Wagemans, 2004) show that humans are able to recognize objects from fragments of contours and edges. Hence, if the appropriate contour fragments of an object are selected, they are representative of it.

Straight lines are not very discriminative since they are only defined by their length (which is useless when looking for scale invariance). However, curves provide a richer description of the object as these are defined, in addition to its length, by its curvature (a line can be seen as a specific case of a curve, i.e., a curve with null curvature). Furthermore, in the presence of variations such as changes in scale, rotation, translation, affine transformations, illumination and texture, the curves tend to remain present. In this paper we use contour fragments as repetitive and discriminant local features.

3.1 Features Extraction

The detection of high curvature contour fragments is based on the method proposed by Chetverikov (Chetverikov, 2003). Chetverikov's method inscribes triangles in a segment of contour points and evaluates the angle of the median vertex which must be smaller than α_{max} and bigger than α_{min}. The sides of the triangle that lie on the median vertex are required to be larger than d_{min} and smaller than d_{max}:

$$d_{min} \leq ||p - p^{+}|| \leq d_{max}, \quad (1)$$

$$d_{min} \leq ||p - p^{-}|| \leq d_{max}, \quad (2)$$

$$\alpha_{min} \leq \alpha \leq \alpha_{max}, \quad (3)$$

d_{min} and d_{max} define the scale limits, and are set empirically in order to avoid detecting contour fragments that are known to be too small or too large. α_{min} and α_{max} are the angle limits that determine the minimum and maximum sharpness accepted as high curvature. In our experiments we set $d_{min} = 10$ pixels, $d_{max} = 300$ pixels, $\alpha_{min} = 5°$, and $\alpha_{max} = 150°$.

Several triangles can be found over the same point or over adjacent points at the same curve, hence it is selected the point with the highest curvature. Each selected contour fragment i is defined by a triangle (p_i^-, p_i, p_i^+), where p_i is the median vertex and the points p_i^- and p_i^+ define the endpoints of the contour fragment. See Figure 1 (a).

The Chetverikov's corners detector has the disadvantage of not being very stable to noisy contours or highly branched contours, which may cause that false corners are selected. For example, see Figure 1(b). In order to deal with this problem, another restriction is added to the Chetverikov's method. Each candidate triangle (p_k^-, p_k, p_k^+) will grow while the points p_k^- and p_k^+ do not match any p_j point of another corner. Figure 1(c) shows how this restriction overcome the false detection in the example in Figure 1(b).

Then, each feature ς_i extracted from the contour is defined by $\langle P_i, T_i \rangle$, where $T_i = (p_i^-, p_i, p_i^+)$ is the triangle inscribed in the contour fragment and $P_i = \{p_1, ..., p_n\}, p_j \in \mathbb{R}^2$ is the set of n points which form the contour fragment ς_i, ordered so that the point p_j is adjacent to the point p_{j-1} and p_{j+1}. Points $p_1, p_n \in P_i$ match with points $p_i^-, p_i^+ \in T_i$, respectively.

3.2 Features Description

The definition of contour fragment given by the extraction process (specifically the triangle (p_i^-, p_i, p_i^+)) provides a compact description of the contour fragment as it gives evidence of amplitude, orientation and length; however, it has low distinctiveness due to the fact that different curves can share the same triangle.

In order to give more distinctiveness to the extracted features, we represent each contour fragment in a polar space of origin p_i, where the length r and the orientation θ of each point are discretized to form a two-dimensional histogram of $n_r \times n_\theta$ bins:

$$H_i(b) = |\{w \in P_i : (w - p_i) \in \text{bin}(b)\}| \quad . \quad (4)$$

Note that for a sufficiently large number of n_r and n_θ this is an exact representation of the contour fragment.

3.3 Robustness and Invariability Considerations

In order to have a robust and invariant description method, several properties must be met:

Locality: the locality property is met directly from the definitions of interest contour fragment and its descriptor given in Sections 3.1 and 3.2. A contour fragment and its descriptor only depend on a point and a set of points in a neighborhood much smaller than the image area, therefore, in both the extraction and description processes, a change or variation in a portion of the contour (produced, for example, by noise, partial occlusion or other deformation of the object), only affects the features extracted in that portion.

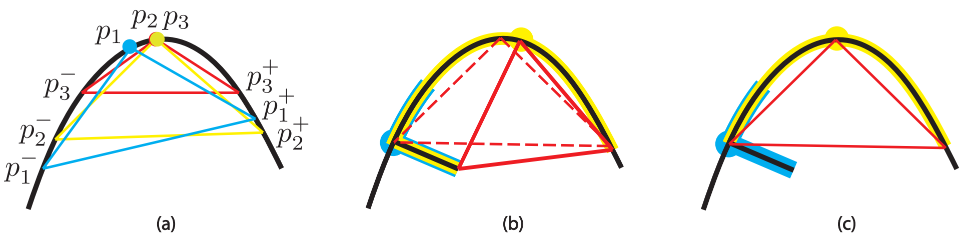

Figure 1: (*best seen in color*). Detection of contour fragments. Are candidates contour fragments those contour fragments where it is possible to inscribe a triangle with aperture between α_{min} and α_{max}, and adjacent sides with lengths between d_{min} and d_{max}. If several triangles are found on the same point or near points, the sharpest triangle in a neighborhood is selected. (b) Noise can introduce false contour fragments (the contour fragment in yellow). (c) To counteract the false contour phenomenon we add another restriction, candidate triangles will grow until another corner is reached.

Translation Invariance: by construction, both the features extraction and description processes are inherently invariant to translation since they are based on relative coordinates of the points of interest.

Rotation Invariance: the contour fragment extraction process is invariant to rotation by construction. An interest contour fragment is defined by a triangle inscribed in a contour segment, which only depends on the shape of the contour segment rather than its orientation. In the description process, it is possible to achieve rotation invariance by rotating each feature coordinate systems until alignment with the bisectrix of the vertex p_i.

Scale Invariance: this could be achieved in the extraction process by extracting contour fragments at different values of d_{min} and d_{max}. In the description process it is achieved by sampling contour fragments (i.e., P_i) to a fixed number M of points or by normalizing the histograms.

4 FEATURE MATCHING

In this section we describe the method for finding correspondences between LISF features extracted from two images. Let's consider the situation of finding correspondences between N_Q features $\{a_i\}$, with descriptors $\{H_i^a\}$, extracted from the query image and N_C features $\{b_i\}$, with descriptors $\{H_i^b\}$, extracted from the database image.

The simplest criterion to establish a match between two features is to establish a global threshold over the distance between the descriptors, i.e., each feature a_i will match with those features $\{b_j\}$ which are at distance $D(a_i, b_j)$ below a given threshold. Usually, matches are restricted to nearest neighbors in order to limit multiple false positives. Some intrinsic disadvantages of this approach limit its use; such as determining the number of nearest neighbors depends on the specific application and type of features and objects. The mentioned approach obviates the spatial relations between the parts (local features) of objects, which is a determining factor. Also, it fails in the case of objects with multiple occurrences of the structure of interest or objects with repetitive parts (e.g. buildings of several equal windows). In addition, the large variability of distances between the descriptors of different features makes the task of finding an appropriate threshold a very difficult task.

To overcome the previous limitations, we propose an alternative for feature matching that takes into account the structure and spatial organization of the features. The matches between the query features and database features are validated by rejecting casual or wrong matches.

Finding Candidate Matches. Let's first define the scale and orientation of a contour fragment.

Let the feature ς_i be defined by $\langle P_i, T_i \rangle$, its scale s_{ς_i} is defined as the magnitude of the vector $\mathbf{p_i^+} + \mathbf{p_i^-}$, where $\mathbf{p_i^+}$ and $\mathbf{p_i^-}$ are the vectors with initial point in p_i and terminal points in p_i^+ and p_i^-, respectively, i.e.,

$$s_{\varsigma_i} = |\mathbf{p_i^+} + \mathbf{p_i^-}|. \tag{5}$$

The orientation ϕ_{ς_i} of the feature ς_i is given by the direction of vector $\mathbf{p_i}$, which we will call orientation vector of feature ς_i, and is defined as the vector that is just in the middle of vector $\mathbf{p_i^+}$ and vector $\mathbf{p_i^-}$, i.e.,

$$\mathbf{p_i} = \hat{\mathbf{p}}_\mathbf{i}^+ + \hat{\mathbf{p}}_\mathbf{i}^-, \tag{6}$$

where $\hat{\mathbf{p}}_\mathbf{i}^+$ and $\hat{\mathbf{p}}_\mathbf{i}^-$ are the unit vectors with same direction and origin that $\mathbf{p_i^+}$ and $\mathbf{p_i^-}$, respectively.

We already defined the terms scale and orientation of a feature ς_i. In the process of finding candidate matches, for each feature a_i, its K nearest neighbors $\{b_j^K\}$ in the candidate image are found by comparing their descriptors (in this work we use χ^2 distance to compare histograms). Our method tries to find among the K nearest neighbors the best match (if any), so K can be seen as an accuracy parameter. To provide the method with rotation invariance the feature descriptors are normalized in terms of orientation. This normalization is performed by rotating the polar coordinate system of each feature by a value equal to $-\phi_{\varsigma_i}$ (i.e., all features are set to orientation zero) and calculated their descriptors. The scale and translation invariance in the descriptors is accomplished by construction (for details see Section 3.2).

Rejecting Casual Matches. For each pair $\langle a_i, b_j^k \rangle$, the query image features $\{a_i\}$ are aligned according to this correspondence:

$$a_i' = (a_i \cdot s + \mathbf{t}) \cdot R(\theta(a_i, b_j^k)), \tag{7}$$

where $s = s_{a_i}/s_{b_j^K}$ is the scale ratio between the features a_i and b_j^k, $\mathbf{t} = p_{a_i} - p_{b_j^k}$ is the translation vector from point p_{a_i} to point $p_k^{b_j}$, and $R(\theta(a_i, b_j^k))$ is the rotation matrix for a rotation, around point p_{a_i}, equal to the direction of the orientation vector of feature a_i with respect to the orientation of b_j^k, (i.e., $\phi_{a_i} - \phi_{b_j^k}$).

Once aligned both images (same scale, rotation and translation) according to correspondence $\langle a_i, b_j^k \rangle$, for each feature a_i' its nearest neighbor b_v in $\{b_j^k\}$ is found. Then, vector $\mathbf{m}$ defined by (l, φ) is calculated, where l is the distance from point p_{b_v} of feature b_v to a reference point $p_\bullet$ in the candidate object (e.g., the object centroid, the point p of some feature or any other point, but always the same point for every candidate image) and φ is the orientation of feature b_v with respect to the reference point $p_\bullet$, i.e., the angle between the orientation vector $\mathbf{p_{b_v}}$ of feature b_v and the vector $\mathbf{p_\bullet}$, the latter defined from point p_{b_v} to point $p_\bullet$,

$$l = ||p_{b_v} - p_\bullet||, \tag{8}$$

$$\varphi = \arccos\left(\frac{\mathbf{p_{b_v}} \cdot \mathbf{p_\bullet}}{||\mathbf{p_{b_v}}|| \; ||\mathbf{p_\bullet}||}\right). \tag{9}$$

Having obtained $\mathbf{m}$, the point $p_\circ$, given by the point at a distance l from point $p_{a_i'}$ of feature a_i' and orientation φ respect to its orientation vector $\mathbf{p_{a_i'}}$, is found,

$$p_\circ^x = p_{a_i'}^x + l \cdot \cos(\phi_{a_i'} + \varphi), \tag{10}$$

$$p_\circ^y = p_{a_i'}^y + l \cdot \sin(\phi_{a_i'} + \varphi). \tag{11}$$

Intuitively, if $\langle a_i, b_j^k \rangle$ is a correct match, most of the points $p_\circ$ should be concentrated around the point $p_\bullet$. This idea is what allows us to accept or reject a candidate match $\langle a_i, b_j^k \rangle$. With this aim, we defined a matching measure Ω between features a_i and b_j^k as a measure of dispersion of points $p_\circ$ around point $p_\bullet$,

$$\Omega = \sqrt{\frac{\sum_{i=1}^{N_Q} ||p_\circ^i - p_\bullet||^2}{N_Q}}. \tag{12}$$

Using this measure, Ω, we can determine the best match for each feature a_i of the query image in the candidate image, or reject any weak match having Ω above a given threshold λ_Ω. A higher threshold means supporting larger deformations of the shape, but also more false matches. In Figure 2, the matches between features extracted from silhouettes of two different instances of the same object class are shown, the robustness to changes in scale, rotation and translation can be appreciated.

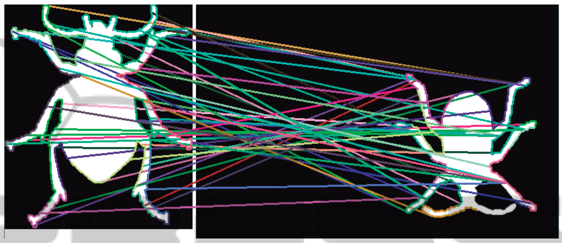

Figure 2: Matches between local shape descriptors in two images. It can be seen how these matches were found even in presence of rotation, scale and translation changes.

5 EXPERIMENTAL RESULTS

Performance of the proposed LISF method has been evaluated on three different well-known datasets. The fist dataset is the Kimia Shape99 dataset (Sebastian et al., 2004), which include nine categories and eleven shapes in each category with variations in form, occlusion, articulation and missing parts. The second dataset is the Kimia Shape216 dataset (Sebastian et al., 2004). The database consists of 18 categories with 12 shapes in each category. The third dataset is the MPEG-7 CE-Shape-1 dataset (Latecki et al., 2000). The database consists of 1400 images (70 object categories with 20 instances per category). In the three datasets, in each image there is only one object, defined by its silhouette, and at different scales and rotations. Example shapes are shown in Figure 3.

In order to show the robustness of the LISF method to partial occlusion in the shape, we generated another 15 datasets by artificially introducing occlusion of different magnitudes (10%, 20%, 30%, 45% and 60%) to the Shape99, Shape216 and MPEG-7 datasets. Occlusion was added by randomly choosing rectangles that occlude the desired portion of the shape contour. A sample image from the MPEG-7 dataset at different occlusion levels is shown in Figure 4.

As a measure to evaluate and compare the performance of the proposed shape matching schema in a shape retrieval scenario we use the so-called bull's

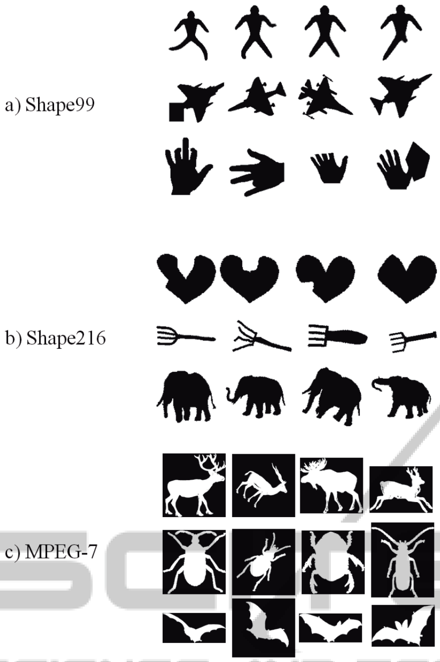

Figure 3: Example images and categories from a) the Shape99 dataset, b) the Shape216 dataset, and c) the MPEG-7 dataset.

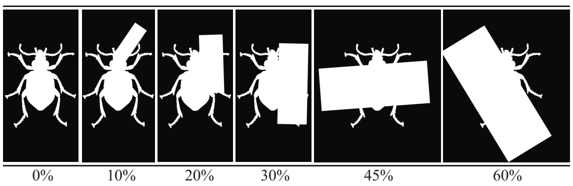

Figure 4: Example image from the MPEG-7 dataset with different levels of occlusion (0%, 10%, 20%, 30%, 45% and 60%) used in the experiments.

eye score. Each shape in the database is compared with every other shape, and the number of shapes of the same class that are among the 40 most similar is reported. The bull's eye score is the ratio between the total number of shapes of the same class and the largest possible value.

The results obtained by LISF ($n_r = 5$, $n_\theta = 10$, $\lambda_\Omega = 0.9$) were compared with those of the popular shape context descriptor (100 points, $n_r = 5$, $n_\theta = 12$) (Belongie et al., 2002), the Zernike moments (using 47 features) (Khotanzad and Hong, 1988) and the Legendre moments (using 66 features) (Chong et al., 2004). Rotation invariance can be achieve by shape context, but it has several drawbacks, as mentioned in (Belongie et al., 2002). In order to perform a fair comparison between LISF method (which is rotation invariant) and shape context, in our experiments the non-rotation invariant implementation of shape context is used, and images used by shape context were rotated so that the objects had the same rotation.

Motivated by efficiency issues, for the MPEG-7 CE-Shape-1 dataset we used only 10 of the 70 categories (selected randomly) with its 20 samples each. The bull's eye score implies all-against-all comparisons and experiments had to be done across the 18 datasets for the LISF, shape context, Zernike moments and Legendre moments methods. There is no loss of generality in using a subset of the MPEG-7 dataset since the aim of the experiment is to compare the behavior of the LISF method against other methods, across increasing levels of occlusion.

As a similarity measure of image a with image b, with local features $\{a_i\}$ and $\{b_j\}$ respectively, we use the ratio between the number of features in $\{a_i\}$ that found matches in $\{b_j\}$ and the total number of features extracted from a.

Figure 5 shows the behavior of the bull's eye score of each method while increasing partial occlusion in the Shape99, Shape216 and MPEG-7 datasets. Bull's eye score is computed for each of the 18 datasets independently.

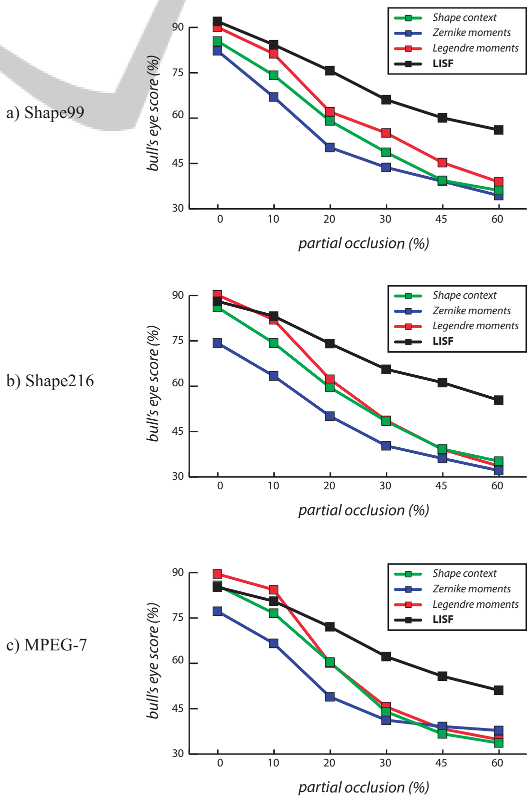

Figure 5: (*best seen in color*). Bull's eye score comparison between LISF, shape context, Zernike moments and Legendre moments in the a) Shape99, b) Shape216 and c) MPEG-7 datasets with different partial occlusions (0%, 10%, 20%, 30%, 45% and 60%).

As expected, the LISF method outperforms the shape context, Zernike moments and Legendre moments methods. Moreover, while increasing the occlusion level, the difference in terms of bull's eye score gets bigger, with about 15 - 20% higher bull's eye score across highly occluded images; which shows the advantages of the proposed method over the other three.

Figure 6 shows the top 5 retrieved images and its retrieval score for the *beetle-5* image with different occlusions. Top 5 retrieved images are shown for each database at different occlusion levels, respectively (MPEG-7 with 0% to 60% partial occlusion). The robustness to partial occlusion of the LISF method can be appreciated. Retrieval score of images that do not belong to the same class as the query image are depicted in red.

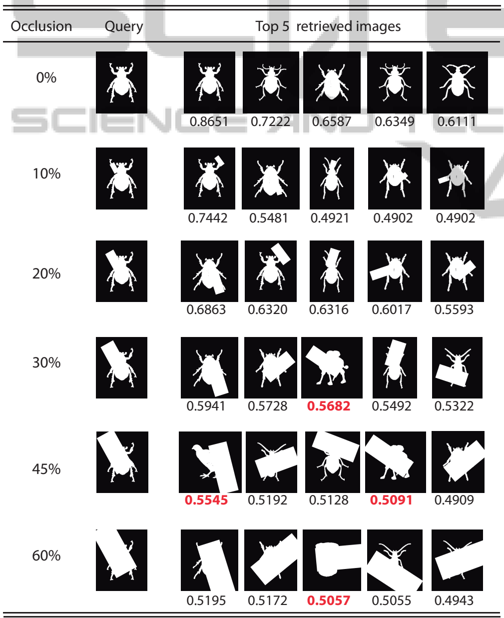

Figure 6: Top 5 retrieved images and similarity score. In each row retrieval results for the *beetle-5* image in the six MPEG-7 based databases. Red retrieval scores represent images that do not belong to the same class of the query image.

In a second set of experiments, the proposed method is tested and compared to shape context, Zernike moments and Legendre moments in a classification task also under varying occlusion conditions. A 1-NN classifier was used, i.e., we assigned to each instance the class of its nearest neighbor. The same data as in the first set of experiments is used. In order to measure the classification performance, accuracy measure was used. Accuracy is the percentage of data that are correctly classified. Figure 7 shows the results of classification under different occlusion magnitudes (0%, 10%, 20%, 30%, 45% and 60% occlusion).

In this set of experiments, a better performance of the LISF method compared to previous work can also be appreciated. As in the shape retrieval experiment, while increasing the occlusion level in the test images, the better is the performance of the proposed method with respect to shape context, Zernike moments and Legendre moments, with more than 25% higher results in accuracy.

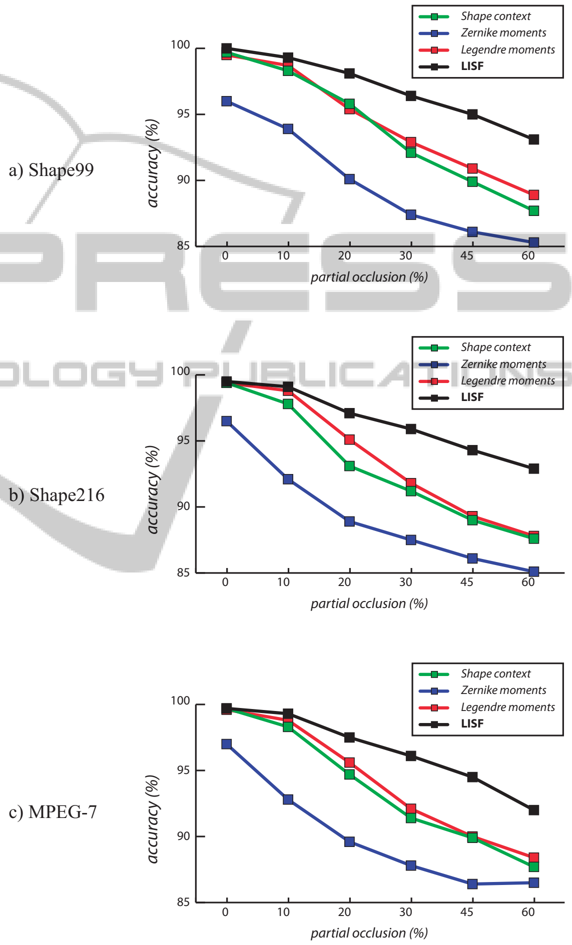

Figure 7: (*best seen in color*). Classification accuracy comparison between LISF, shape context, Zernike moments and Legendre moments in the a) Shape99, b) Shape 216, and c) MPEG-7 dataset, with different partial occlusions (0%, 10%, 20%, 30%, 45% and 60%).

The computation time of LISF has been also evaluated, and compared to other methods. Table 1 shows the comparison of LISF computation time against shape context, Legendre moments, and Zernike moments. The reported times correspond to the average time needed to describe and match two shapes of the MPEG-7 database over 500 runs. These results were obtained on a single thread of a 2.2 GHz processor and 8Gb RAM PC. As can be seen in Table 1, LISF

Table 1: Average feature extraction and matching time for two images of the MPEG7 database, in seconds.

Method	Computation time (s)
Shape context	2.66
Legendre moments	7.48
Zernike moments	26.47
LISF	**0.47**

is the least time-consuming method compared with shape context, Legendre moments, and Zernike moments.

6 CONCLUSIONS AND FUTURE WORK

As a result of this work, a method for shape features extraction, description and matching, invariant to rotation, translation and scale, have been developed. The proposed method allows us to overcome the intrinsic disadvantages of only using local or global features by capturing both local and global information. The conducted experiments supported the mentioned contributions, showing larger robustness to partial occlusion than other methods in the state of the art. It is also more efficient in terms of computational time than the other techniques.

Moreover, the feature extraction process does not depend on accurate and perfect object segmentation since the features are extracted from both the contour and the internal edges of the object. Therefore, the method has great potential for use in "real" images (RGB or grayscale images) and also, as a complement to certain limitations of appearance based methods (e.g., SIFT, SURF, etc.); particularly in object categorization, where shape features usually offer a more generic description of objects. Future work will focus on this subject.

ACKNOWLEDGEMENTS

This project was supported in part by CONACYT grant Ref. CB-2008/103878 and by Instituto Nacional de Astrofísica, Óptica y Electrónica. L. Chang was supported in part by CONACYT scholarship No. 240251.

REFERENCES

Adamek, T. and O'Connor, N. E. (2004). A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Trans. Circuits Syst. Video Techn.*, 14(5):742–753.

Alajlan, N., Rube, I. E., Kamel, M. S., and Freeman, G. (2007). Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*, 40(7):1911 – 1920.

Bai, X., Yang, X., Latecki, L. J., Liu, W., and Tu, Z. (2010). Learning context-sensitive shape similarity by graph transduction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(5):861–874.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522.

Biederman, I. and Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20(1):38–64.

Chetverikov, D. (2003). A Simple and Efficient Algorithm for Detection of High Curvature Points in Planar Curves. *Proceedings of the 23rd Workshop of the Austrian Pattern Recognition Group*, pages 746–753.

Chong, C.-W., Raveendran, P., and Mukundan, R. (2004). Translation and scale invariants of legendre moments. *Pattern Recognition*, 37(1):119–129.

De Winter, J. and Wagemans, J. (2004). Contour-based object identification and segmentation: stimuli, norms and data, and software tools. *Behavior research methods instruments computers. A journal of the Psychonomic Society Inc*, 36(4):604–624.

Direkoglu, C. and Nixon, M. (2011). Shape classification via image-based multiscale description. *Pattern Recognition*, 44(9):2134–2146.

Gonzalez-Aguirre, D. I., Hoch, J., Rhl, S., Asfour, T., Bayro-Corrochano, E., and Dillmann, R. (2011). Towards shape-based visual object categorization for humanoid robots. In *ICRA*, pages 5226–5232. IEEE.

Khotanzad, A. and Hong, Y. H. (1988). Rotation invariant pattern recognition using zernike moments. *Pattern Recognition, 1988., 9th International Conference on*, pages 326–328 vol.1.

Kim, W.-Y. and Kim, Y.-S. (2000). A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication*, 16(12):95 – 102.

Latecki, L. J., Lakmper, R., and Eckhardt, U. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, pages 1424–1429. IEEE Computer Society.

McNeill, G. and Vijayakumar, S. (2006). Hierarchical procrustes matching for shape retrieval. In *CVPR (1)*, pages 885–894. IEEE Computer Society.

Mokhtarian, F. and Bober, M. (2003). *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. Kluwer.

Sebastian, T. B., Klein, P. N., and Kimia, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571.

Shu, X. and Wu, X.-J. (2011). A novel contour descriptor for 2D shape matching and its application to image retrieval. *Image and Vision Computing*, 29(4):286–294.

Toshev, A., Taskar, B., and Daniilidis, K. (2011). Shape-based Object Detection via Boundary Structure Segmentation. *International Journal of Computer Vision*, 99(2):123–146.

Trinh, N. H. and Kimia, B. B. (2011). Skeleton Search: Category-Specific Object Recognition andSegmentation Using a Skeletal Shape Model. *International Journal of Computer Vision*, 94(2):215–240.

Wang, X., Bai, X., Ma, T., Liu, W., and Latecki, L. J. (2012). Fan shape model for object detection. In *CVPR*, pages 151–158. IEEE.

Yang, X., Bai, X., Kknar-Tezel, S., and Latecki, L. (2013). Densifying distance spaces for shape and image retrieval. *Journal of Mathematical Imaging and Vision*, 46(1):12–28.

Yang, X., Kknar-tezel, S., and Latecki, L. J. (2009). Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR*.

Zhang, D. and Lu, G. (2002). Shape based image retrieval using generic fourier descriptors. In *Signal Processing: Image Communication 17*, pages 825–848.