

# Elastic Hand Contour Matching in NIR Images with a Novel Shape Descriptor Parametrization

A. Inogés, G. López-Nicolás, C. Sagüés

Instituto de Investigación en

Ingeniería de Aragón (I3A)

Universidad de Zaragoza, Zaragoza, Spain

ainoges@unizar.es, gonlopez@unizar.es, csagues@unizar.es

S. Llorente

Research and Development Department,

Induction Technology, Product Division Cookers

BSH Home Appliances Group, Zaragoza, Spain

sergio.llorente@bshg.com

**Abstract**—Detection and tracking of hands from visual information is a key issue for gesture recognition. Based on Near Infrared (NIR) imagery we present a new procedure for hand detection. The sensor we propose is a single NIR camera with an attached NIR light source and a visible light filter, which makes our system invariant under changing environments and uncontrolled lighting conditions. The main contributions are twofold. On the one hand we propose a new shape descriptor composed of inner-angles and inner-distances. On the other hand, our novel parametrization of the extracted descriptor allows for elastic hand contour matching. The proposal is validated experimentally and tested with the standard data set MPEG-7.

## I. INTRODUCTION

Hand gesture recognition systems are nowadays widely used in the field of Human-Computer Interaction (HCI). Video games [1], multimedia [2] or home appliances controlling applications [3] are some of the fields that incorporate these systems. In general terms, a process of hand gesture recognition can be divided into three main steps. First, the hand has to be properly segmented from the background. Then, it is necessary to track the path of the hand along the time and finally, depending on the trajectory performed by the hand, one gesture or another is recognized. Since these vision systems have to perform effectively in a wide variety of environments and under changing lighting conditions, the process of hand detection is still an unsolved challenge.

Different methods in the literature try to address these problems. Among these algorithms, the techniques used to detect the position of a hand are directly related to the kind of sensor chosen to acquire the images. Many systems are equipped with RGB cameras because of their low cost. In this case, the algorithms mostly depend on skin color to differentiate between hand region and background [4]. However, they have been proved not to be quite effective at identifying hands in some situations in which hand color detection is not possible, specially under uncontrolled lighting conditions. Other more robust systems use 3D images and rely on the distance to the sensor to segment the hand. Different sensors can be used to get depth images. There exist color and depth sensing devices (RGB-D), as Microsoft Kinect [5], Time-Of-Flight sensors (TOF) [6] and stereo vision systems [7]. In

our approach a single monochromatic camera as vision sensor with infrared LED as lighting source are used. Here the lens of the sensor is used in conjunction with a visible light filter. Therefore, invariance to external lighting conditions is ensured and situations with difficult skin color detection are overcome.

Once the hand candidate has been segmented from the background, it is necessary to determine the position of the hand. Ganapathyraju presents in [8] a specific method for hands detection. There, the convexity hull defects of the segmented candidate are computed. Despite the fact that this method performs well when the hand is perfectly segmented, in some cases in which more parts of the user's body appear, the convexity defects might be not properly extracted. Something similar occurs with techniques that rely on Neural Networks, as the proposed in [9]. Although they have been proved to be quite robust at identifying a wide variety of hand gestures, they are not suitable for detecting the hand region within open contours. It is also necessary to perform training processes, which is not the most efficient way to proceed. More general techniques can be found in the literature. These are not only suitable for hand contour matching but also for detection of other different kinds of objects. These methods can be classified depending on whether they use boundary (contour-based) or interior (region-based) shape information. Since contour-based methods have been proved to be more effective, we focus on this category.

Contour-based methods include both distances and angles computation. Xu et al. [10] compute the high curvature contour points to use them as shape descriptors. In the curvature scale space methods [11], the number of points with high curvature values at different levels of smoothed contour is used as main feature for shape matching. Since we are dealing with quite irregular contours, this method would not be accurate enough to determine the portion of the contour that corresponds to a hand. Wang et al. in [12] and Ling and Jacobs in [13] compute the relative distance among all the contour points. With this approach they achieve a dense representation of the contour. However, the complexity of the problem increases exponentially with the number of contour points, and it turns out computationally expensive to extract the contour descrip-

tor. Michel et al. tried to overcome the dimensionality problem with the method proposed in [14]. They compute the inner-distances from each contour point to the nearest contour point along a discrete number of directions pointing to the inner-side of the contour. To deal with open contours they label the distances along the directions pointing to those open contours in a different manner. They also propose reverting the inner-descriptor so as to deal with strongly concave contours. In such situations they compute the external distances.

In our approach we have implemented a variation of the above mentioned methods that takes into account both local and global information. We first compute the inner-distances in a similar way as in [14]. This reduces significantly the complexity of the problem in comparison with the algorithm proposed in [13]. However, we deal with open and concave contours in a more general way than in [14]. To manage strongly concave contours we rely on a variation of the inner-angles proposed in [13]. We assign to each contour point a set of the inner-angles of its neighbors.

Shortly, in this paper we address the problem of hand detection in NIR images with a new shape descriptor and a novel parametrization to perform elastic matching. Real experiments with NIR images show promising results. Additionally, our method is compared with related works with good performance in terms of accuracy, robustness and efficiency.

## II. CONTOUR EXTRACTION

The raw data acquired by the NIR sensor is composed of grayscale pixel values. The pixels values are directly related to the intensity of the light that each part of the scene reflects. Since lighting conditions depend both on hands position and hands distance to the sensor, a fixed threshold would not be adequate to binarize the acquired image. We propose a novel threshold algorithm that adapts its threshold value depending on the distribution of the associated histogram.

In particular, given a grayscale image (Fig. 1(left)), we compute the histogram of the intensity values (Fig. 2). Note that the biggest peak corresponds to background pixels. However, the specific value at which the background pixels end and the relevant information starts is not so easy to determine. Since in general the relevant information starts within homogeneous pixels values in the image, in our approach we have set the threshold to the value at which the derivative of the histogram values approaches to zero. In particular, for the image shown in Fig. 1(right), the calculated threshold is 60 over 255.

Once the initial image has been binarized, some morphological operations are performed. By means of erosion, dilation and Gaussian smoothing it is feasible to properly extract contours even in noisy images. It might occur that the number of final blobs is higher than one, so attending to the area of each blob we are able to filter all non-relevant blobs to finally get the one that contains the hand candidate (Fig. 1(right)). Since the size of the candidate that appears in the image can vary considerably depending on its distance to the sensor, a normalization of the extracted contour is performed here. This normalization directly contributes to a more effective result in

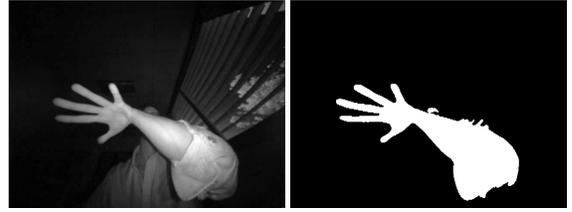


Fig. 1. Grayscale image captured by the sensor (left). In this type of images, the value of each pixel depends on the IR intensity that each part of the image reflects. Segmented image (right). The threshold value for this image is 60 over 255. Note that additional blobs have already been filtered so as to get a single candidate blob.

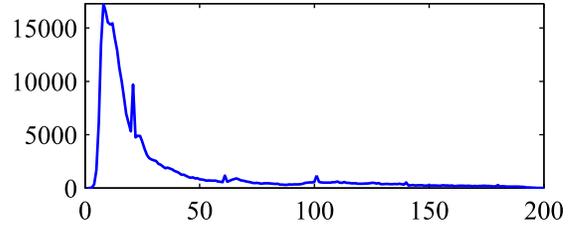


Fig. 2. Histogram of the pixel values in the grayscale image of Fig. 1. The biggest peak corresponds to background pixels in the image. The threshold value is set for the value at which the histogram derivative approaches to zero. Here this threshold value has been set to 60 over 255.

the matching step. So, given a set of  $N$  contour points, denoted by  $P^* = \{p_1^*, p_2^*, \dots, p_N^*\}$ , and being each contour point a set of two coordinates,  $p_i^* = (x_i^*, y_i^*)$ , the normalized contour will be denoted as  $P = \{p_1, p_2, \dots, p_N\}$  and each component  $p_i$  is computed as follows:

$$p_i = \left( p_i^* - \frac{\sum_{i=1}^N p_i^*}{N} \right) \cdot \left( \frac{N \cdot \sqrt{2}}{\sum_{i=1}^N \sqrt{p_i^{*T} \cdot p_i^*}} \right). \quad (1)$$

The first of the two terms centers the contour points around zero, whereas the second term re-sizes the contour so that all extracted contours have similar dimensions. Other considerations related to the process of contour extraction need to be taken into account. Here the contour points are forced to be counter-clockwise ordered, but any other convention could have been chosen. In addition, the extracted points have to be equally spaced to ensure an homogeneous shape descriptor computation. In posterior steps, a re-sampling process will be performed depending on the local scale of the shape. This permits to characterize more densely those areas of the contour that require more detail. Fig. 3(left) shows the extracted contour for the acquired image.

## III. SHAPE DESCRIPTOR

In this section, the shape descriptor **SD** defined to identify which part of a segmented contour corresponds to a hand is presented. First, it is necessary to take into account that the proposed descriptor has to be invariant to geometric transformations such as scaling, orientation and translation. Robustness under local non-rigid deformations is also an essential requirement in order to overcome difficult matching situations when self occlusions or partial segmentations occur.

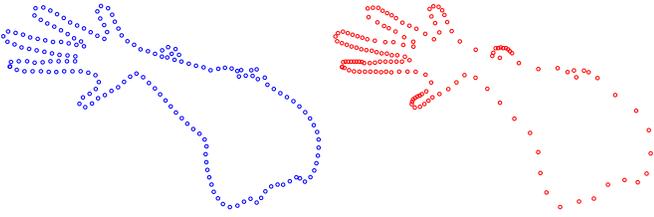


Fig. 3. Contour extraction and re-sampling. Initial contour (left). Locally scale dependent re-sampled contour (right).

### A. Shape Descriptor Computation

In the proposed approach two different measurements have been used to compose the final descriptor, the inner-distances and the inner-angles of the extracted contour. Given the extracted contour  $P$ , it is first necessary to check whether it is an open or a close contour. To do that, the first and the last contour points are compared. In the case it is an open contour, we will consider the starting point as the ending one so that  $p_1$  and  $p_N$  are the same point. Now for each contour point ( $p_i$ ), the inner-angle ( $\theta_i$ ) is computed as the angle defined by the evaluation point ( $p_i$ ), the next contour point ( $p_{i+1}$ ) and the point before the evaluation point ( $p_{i-1}$ ). Thus, being the vector  $\mathbf{r}_- = p_{i-1} \times p_i$  and the vector  $\mathbf{r}_+ = p_i \times p_{i+1}$ , where  $p_i, p_{i-1}, p_{i+1}$  are written here in homogeneous coordinates ( $p_i = (x_i, y_i, 1)$ ), the inner-angle  $\theta_i$  will be computed as follows:

$$\theta_i = \text{atan2} \left( (\mathbf{r}_- \times \mathbf{r}_+)_z, \mathbf{r}_-^T \cdot \mathbf{r}_+ \right), \quad (2)$$

where  $\theta_i$  is transformed such  $\theta_i \in (0, 2\pi)$ . Then, this inner-angle is equally divided into  $K$  directions, and the distances from the evaluation point to the intersection points of those rays with the contour are calculated. In Fig. 4 an example of this process is shown. Here the inner-angle is represented with a gray shadow. In the proposed approach, the  $\mathbf{SD}_P$  for the  $N$ -points contour  $P$  will be composed by  $Z$  inner-angles and by  $K$  inner-distances:

$$\mathbf{SD}_P = \begin{pmatrix} \mathbf{A}_P^{[Z \times N]} \\ \mathbf{D}_P^{[K \times N]} \end{pmatrix}, \quad (3)$$

where the sub-matrices denoted as  $\mathbf{A}_P$  and  $\mathbf{D}_P$  correspond to the inner-angles and inner-distances components, respectively. Therefore, the final dimension of the proposed descriptor will be  $[M \times N]$ , being  $M = Z + K$ . It is necessary to note that both the inner-angles and the inner-distances values require to be normalized by the maximum value of the computed inner-angles and inner-distances, respectively. Hence, the values of the  $\mathbf{SD}_P$  for the given contour will be all in the range  $(0, 1]$ . Here, for a single contour point ( $p_i$ ), each inner-angle value will be  $\mathbf{A}_P(z, i) = \theta_\xi$  (with  $z = \{1, \dots, Z\}$ ), where  $\theta_\xi$  is the inner-angle of the point that corresponds to index  $\xi$ , which is calculated as follows:

$$\xi = i - \frac{Z+1}{2} + z. \quad (4)$$

Since the proposed descriptor is calculated as a set of inner-angles and inner-distances, it is intrinsically invariant to

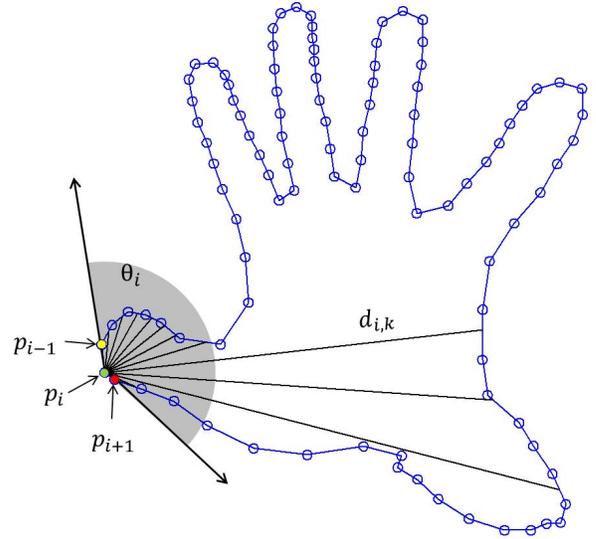


Fig. 4. Shape descriptor calculation scheme. Here the number of inner-distances has been set to  $K = 11$  and the inner-angle is represented by the gray shadow.

translations and rotations. Due to the fact that both the contour and the final descriptor values have been normalized it is also invariant to scale changes.

### B. Adaptive Local Scale Sampling

In the first step of contour extraction, all the points should be equally spaced. However, due to the fact that the hand candidate will probably be integrated in a bigger contour, such as the arm or the chest of the people, it should be desirable to perform a contour re-sampling process dependent on the inner-distances components of the computed  $\mathbf{SD}_P$ . The idea of estimating the local scale to define the sampling rate according to the level of detail of each portion of the contour was already developed in [14]. However, in that case the non-uniform re-sampling was performed without imposing any constraints. Here we define a variation of that method by limiting the maximum and the minimum sampling rate.

Let  $P$  be the target contour and let  $\mathbf{D}_P$  be the inner-distances components of the corresponding computed shape descriptor of  $[K \times N]$ . The local scale for the target contour ( $LS_P$ ) will be defined as  $LS_P = \{\phi_1, \phi_2, \dots, \phi_N\}$ , and the particular value for each contour point is computed as follows:

$$\phi_i = \left\lceil \frac{\sum_{j=1}^K \mathbf{D}_P(j, i)}{K} \cdot \lambda + \mu \right\rceil, \quad (\lambda, \mu > 0), \quad (5)$$

where parameters  $\lambda$  and  $\mu$  limit the maximum and the minimum sampling rate, respectively. The final contour  $Q = \{q_1, q_2, \dots, q_L\}$  will be composed by a subset of the initially extracted contour points so that the number of points discarded after every point added to the final contour will be its computed local scale value. This method samples more densely those parts of the contour in which little details appear, while the parts that do not experiment relevant variations are

more widely sampled. This fact directly contributes to a more efficient performance of the detection algorithm. Since  $L < N$ , the number of contour points is drastically reduced by the re-sampling method. Hence, the computation cost of the contour matching process will be also reduced. In Fig. 3(right) the locally scale dependent re-sampled contour of the segmented candidate is presented.

#### IV. SHAPE MATCHING

Given a grayscale image acquired with the sensor, the problem of identifying which part of the extracted contour corresponds to a hand has been addressed as a target to reference matching problem. Thus, it will be first necessary to extract the shape descriptor of a reference hand (Fig. 6(top-left)). This reference should be a hand correctly segmented so as to ensure that no information has been lost during the segmentation or descriptor computation processes. Before presenting the shape matching procedure some considerations need to be explained. Since the candidate hand can be integrated in a bigger contour, the reference and the candidate hands might have different number of points. Moreover, depending on the distance to the sensor of the hand, the inner-distances would be very different. Note also that the normalization performed in the previous section does not overcome this problem since the extracted shape descriptor has been normalized by the maximum inner-distance value of the entire blob, not just the hand. We propose to solve this issue with a novel parametrization of the reference descriptor and an elastic matching algorithm based on an appropriately defined cost function optimization.

##### A. Shape Descriptor Parametrization

To allow for an elastic matching between the reference hand and the hand integrated within the target contour, the computed shape descriptor for the reference hand,  $\mathbf{SD}_R$ ,  $[M \times U]$  dimensional, is parametrized with two parameters. The first parameter,  $\gamma$ , allows for amplitude scaling and the second parameter,  $\sigma$  permits length elasticity. It is necessary to notice here that the inner-angle components of the computed descriptor do not require to be scaled since angles do not depend on the size of the contour. Thus, the amplitude scaled reference shape descriptor  $\mathbf{SD}_R^*$  is computed as follows:

$$\mathbf{SD}_R^* = \begin{pmatrix} \mathbf{A}_R \\ \gamma \cdot \mathbf{D}_R \end{pmatrix} \quad \text{with} \quad \mathbf{SD}_R = \begin{pmatrix} \mathbf{A}_R \\ \mathbf{D}_R \end{pmatrix}, \quad (6)$$

where the amplitude scaling only has to be applied to the inner-distances components of the shape descriptor. Then to achieve length scaling, the final shape descriptor  $\mathbf{SD}_R^{**}$ , will be obtained by discarding one of every  $\sigma - i$  points present in the amplitude scaled descriptor ( $\mathbf{SD}_R^*$ ). Thus, the final reference shape descriptor  $\mathbf{SD}_R^{**}$  will be  $[M \times V]$  dimensional, being  $V = \sigma \cdot U$ . Notice that the proposed parametrization of the shape descriptor allows for elastic matching between a reference hand and the one we want to detect. Without the proposed parametrization it would be necessary to generate a number of templates of reference hands at different scales and with different number of contour points, implying higher computational cost.

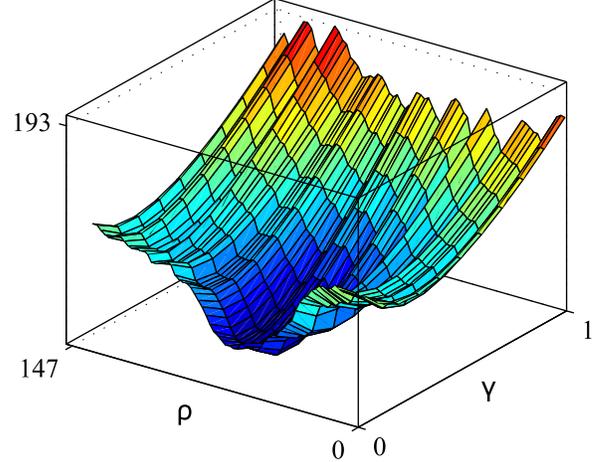


Fig. 5. Cost function. The representation of the cost function has been performed depending on the amplitude scale factor  $\gamma$  and the starting point  $\rho$ . The values for each factor are  $\gamma = \{0, 0.1, 0.2, \dots, 1\}$  and  $\rho = \{0, 2, 4, \dots, L - V\}$ , respectively.

##### B. Matching Algorithm

In the process of target to reference matching it is essential to define an appropriate cost function to determine which is the best correspondence between them. Let  $Q = \{q_1, q_2, \dots, q_L\}$  be the target and already re-sampled contour of  $L$  points and let  $R = \{r_1, r_2, \dots, r_U\}$  be the reference contour already re-sampled of  $U$  points. Let their corresponding shape descriptors be  $\mathbf{SD}_Q$  of  $[M \times L]$  and  $\mathbf{SD}_R^{**}$  of  $[M \times V]$ , respectively. The cost function  $C$  is defined as follows:

$$C = \sum_{i=1}^V \sum_{j=1}^M \frac{\left( \mathbf{SD}_R^{**}(j, i) - \mathbf{SD}_Q(j, \rho + i) \right)^2}{\text{Max} \left\{ \mathbf{SD}_R^{**}(j, i), \mathbf{SD}_Q(j, \rho + i) \right\}}. \quad (7)$$

Here, an additional parameter  $\rho$  has been introduced to determine at which point of the target contour the algorithm starts to compare both shape descriptors. In Fig. 5 the cost function for the target and reference descriptors of the previous example (Figs. 1 and 6(top-left)) is shown. It is interesting to notice here that despite the fact that the defined cost function presents a global minimum (darkest-blued area) there are also many local minima that can make the algorithm not to reach the optimum correspondence.

To overcome this issue and to be able to reach the global minimum of the cost function the *GlobalSearch* algorithm proposed by Ugray et al. in [15] has been used here. This method, instead of using a single starting point like other optimization methods do, uses a distributed set of points. Starting the searching algorithm from several initial points it is more likely to reach the global minimum. This method uses the scatter search algorithm proposed by Glover in [16] to generate the set of trial points.

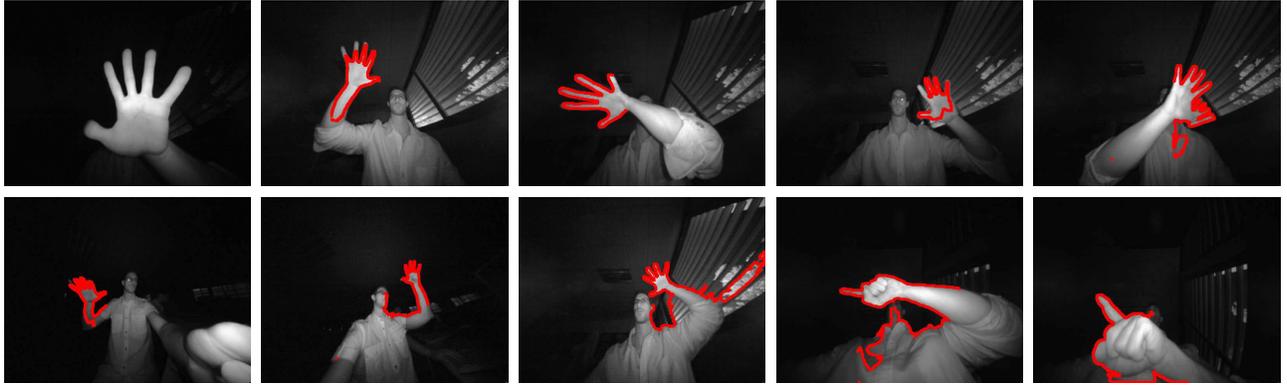


Fig. 6. Hand detection. The top-left image corresponds to a reference hand for an open-palm hand matching. The rest of the images illustrate hand detection with different hand poses, number of fingers and lighting conditions. Note last two images use a different reference for the case of one-finger hand detection.

## V. EXPERIMENTAL RESULTS

In order to demonstrate the effectiveness of the proposed approach two different validations have been performed. First, the method has been tested under real conditions. Then, the proposed shape descriptor has been tested with a widely used data set available online [17], the MPEG-7 database. Here the number of inner-angles  $Z$ , and inner-distances  $K$ , has been set empirically to 3 and 11, respectively and kept fixed for both experiments since those values result in a good compromise between accuracy and computation cost.

### A. Hand Detection in NIR images

The use of a NIR sensor with an attached NIR light source and a visible light filter permits to illuminate the scene with light of a particular wavelength at the same time that reaches invariance to visible external lighting sources as the sun or lamps. In order to demonstrate the robustness of the proposed method we have defined a set of experimental grayscale images acquired with the NIR sensor with size of  $480 \times 640$  pixel. Here several scenarios with different hand poses, number of fingers and under different lighting conditions have been studied. In Fig. 6 nine different situations are shown. The first image (Fig. 6(top-left)) corresponds to a reference image for an open-palm hand matching. A different reference for a single-finger hand matching was used for the last two examples. As can be observed, this method not only allows for the hand region detection, but it also properly identifies which is the contour that corresponds to the hand limits. This information will be essential to develop a hand gestures recognition system since it can provide not only the exact position of the hand in the image but also the number of fingers, the angle they form and other relevant hand properties.

Since this experiment focuses on detecting the contour that corresponds to a hand, it would be difficult to provide a quantitative performance evaluation value. To allow for comparison with other shape descriptors that can be found in the literature an additional experiment has been conducted with the MPEG-7 data set.

### B. MPEG-7 Dataset

The MPEG-7 data set [17] has been used to compare the performance of the proposed shape descriptor with other descriptors that can be found in the bibliography. This database consists of 1400 silhouettes of 2D objects grouped in 70 classes, 20 objects each. Fig. 7 shows six example images of three different categories. As can be observed, objects of the same class can have very different shapes, sizes and orientations. Local boundary and non-rigid deformations can also difficulty the matching of images with their correct category. The retrieval score for the MPEG-7 database is the so-called *Bull's eye score*. This retrieval score is computed as the rate between the sum of correct matches and the total number of possible correct matches, that is 28000 (1400 objects in total and 20 objects per class). The number of correct matches for each image is computed as the sum of the number of silhouettes of the same category as the target silhouette within the 40 most similar shapes.

In the experiment, the number of points for each image has been set to  $N = 100$ . Note that in the MPEG-7 data set the matching process complexity is reduced to a closed-to-closed contour matching whereas our approach can handle with open contours and shapes integrated in bigger shapes. Therefore, the matching carried out here is based on the dynamic programming procedure presented in [13].

The proposed approach achieves encouraging matching results, reaching a 83.7% for the MPEG-7 data set. In Fig. 8 the distance matrix for the experiment is represented. It is necessary to note here that each component of the distance matrix measures the dissimilarity between the row and the column image. The strong diagonal structure of this graph shows the high robustness and the accuracy achieved with the proposed descriptor. Note that the number of diagonal elements (number of categories) is 70 and each one of the darker diagonal elements is a square sub-matrix of  $[20 \times 20]$  (20 images per category). In Table I our approach is compared with several state-of-the-art methods. As can be observed, the shape descriptor proposed in this paper outperforms the algorithms proposed in [18] and [14] and achieves comparable results

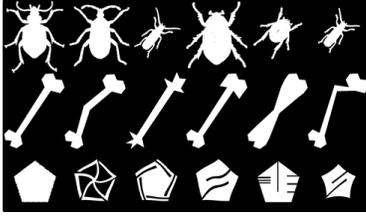


Fig. 7. MPEG-7 Database. Six sample images of three different categories.

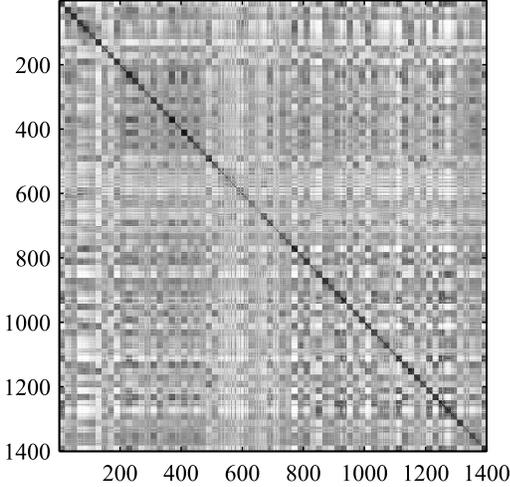


Fig. 8. Distance matrix. Each row and each column are one of the 1400 images of the MPEG-7 data set. The value of each component of the distance matrix measures the dissimilarity between the corresponding images.

TABLE I  
RETRIEVAL COMPARISON OF MPEG-7 Bull's eye score

Algorithm	$O_D$	$O_M$	Score
Height Func. [12]	$100 \times 100$	$20 \times 100$	89.66%
IDSC+DP [13]	$100 \times 100$	$96 \times 100$	85.4%
Our Approach	$14 \times 100$	$14 \times 100$	83.7%
Partial Match. [14]	-	-	83.4%
Distance Sets [18]	$100 \times 250$	$100 \times 250$	78.38%

with the algorithms proposed in [13] and [12]. Therefore, the main contribution of the proposed shape descriptor is a noticeable reduction of the complexity of the process of descriptor extracting ( $O_D$ ) and the complexity of the matching problem ( $O_M$ ), that will directly contribute to a more efficient performance of the system in a real context.

## VI. CONCLUSION

We have presented a shape descriptor with a novel parametrization. This parametrization of the shape descriptor in conjunction with a properly defined cost function and an adequate choice of the optimization algorithm permits elastic hand contour matching in NIR images. The main advantage of the proposed vision system is that it provides invariance under different external lighting conditions. The approach has been tested in a real setup and has demonstrated to be quite robust and accurate even in complex scenes. Comparisons with state-of-the-art algorithms demonstrate good performance in terms

of accuracy despite its low complexity. We consider the next step for future work the extension of the proposed approach to a hand gesture recognition algorithm.

## ACKNOWLEDGMENT

This work was supported by projects IPT-2011-1158-920000 of subprogram INNPACTO, RTC-2014-1847-6 of subprogram Retos-Colaboración, and by DPI2012-32100 from Ministerio de Economía y Competitividad/European Union.

## REFERENCES

- [1] Y. Zhu and B. Yuan, "Real-time hand gesture recognition with kinect for playing racing video games," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 3240–3246.
- [2] M. Lech, B. Kostek, A. Czyzewski, and P. Ody, "Gesture recognition framework for multimedia content viewer controlling," in *IEEE Conference Proceedings on Signal Processing Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2009, pp. 100–104.
- [3] R. Senanayake and S. Kumarawadu, "A robust vision-based hand gesture recognition system for appliance control in smart homes," in *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, 2012, pp. 760–763.
- [4] S. Zhao, L. Shen, and W. Tan, "A method of dynamic hand gesture detection based on local background updates and skin color model," in *IEEE International Conference on Computer Application and System Modeling (ICCASM)*, vol. 5, 2010, pp. 657–660.
- [5] S.-W. Lee, H. S. Yang, and Y.-H. Seo, "Foreground extraction algorithm using depth information for image segmentation," in *IEEE International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA)*, 2013, pp. 581–584.
- [6] H. S. Cho, K. Bae, K.-M. Kyung, S. Jeong, and T.-C. Kim, "Background subtraction based object extraction for time-of-flight sensor," in *IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, 2013, pp. 48–49.
- [7] Y. Guan, "Stereo vision based video real-time 3d pointing gesture recognition," in *IET Conference on Wireless, Mobile and Sensor Networks (CCWMSN07)*, 2007.
- [8] S. Ganapathyraju, "Hand gesture recognition using convexity hull defects to control an industrial robot," in *IEEE 3rd International Conference on Instrumentation Control and Automation (ICA)*, 2013.
- [9] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 8, pp. 1141–1158, 2009.
- [10] C. Xu, J. Liu, and X. Tang, "2d shape matching by contour flexibility," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 180–186, 2009.
- [11] F. Mokhtarian and M. Bober, *Curvature scale space representation: theory, applications, and MPEG-7 standardization*. Incorporated Springer Publishing Company, 2011.
- [12] J. Wang, X. Bai, X. You, W. Liu, and L. J. Latecki, "Shape matching and classification using height functions," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 134–143, 2012.
- [13] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.
- [14] D. Michel, I. Oikonomidis, and A. Argyros, "Scale invariant and deformation tolerant partial shape matching," *Image and Vision Computing*, vol. 29, no. 7, pp. 459–469, 2011.
- [15] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, "Scatter search and local nlp solvers: A multistart framework for global optimization," *INFORMS Journal on Computing*, vol. 19, no. 3, pp. 328–340, 2007.
- [16] F. Glover, "A template for scatter search and path relinking," in *Artificial evolution*. Springer, 1998, pp. 1–51.
- [17] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *IEEE Conference Proceedings on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2000, pp. 424–429.
- [18] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1274–1286, 2003.