Comparison of Body Shape Descriptors for Biometric Recognition using MMW Images

Ester Gonzalez-Sosa, Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia
Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{ester.gonzalezs, ruben.vera, julian.fierrez, javier.ortega}@uam.es

***Abstract*—The use of Millimetre wave images has been proposed recently in the biometric field to overcome certain limitations when using images acquired at visible frequencies. In this paper, several body shape-based techniques were applied to model the silhouette of images of people acquired at 94 GHz. We put forward several methods for the parameterization and classification stage with the objective of finding the best configuration in terms of biometric recognition performance. Contour coordinates, shape contexts, Fourier descriptors and silhouette landmarks were used as feature approaches and for classification we utilized Euclidean distance and a dynamic programming method. Results showed that the dynamic programming algorithm improved the performance of the system with respect to the baseline Euclidean distance and the necessity of a minimum resolution of the contour to achieve promising equal error rates. The use of the contour coordinates is the most suitable feature to use in the system regarding the performance and the computational cost involved when having at least 3 images for model training. Besides, Fourier descriptors are more robust against rotations, which may be of interest when dealing with few training images.**

I. Introduction

Many biometric characteristics are used to identify individuals: fingerprint, iris, voice, face, hand, signature, etc. The majority of these biometric traits are acquired with cameras working at visible frequencies of the electromagnetic spectrum. Such images are affected by, among others factors, lighting conditions and body occlusions (e.g. clothing, make up, hair, etc.). To overcome these limitations, researchers have proposed the use of images acquired at other spectral ranges: X-ray, infrared, millimeter (MMW) and submillimeter (SMW) waves [1]. The images captured beyond the visible spectrum circumvent, to some extent, some of the mentioned limitations; furthermore, they are more robust to spoofing than other biometric images/traits.

Among the spectral bands out of the visible spectrum, the millimeter waves (with frequency in the band of 30-300 GHz) present interesting properties that can be exploited in biometrics: ability to pass through clothing and other occlusions, innocuous to health, low intrusiveness, and the recent deployment and rapid progress of GHz-THz systems in screening applications.

In spite of the previous advantages, to date, there are just a few works on this field. Specifically, just one working with real data [2], and some others based on BIOGIGA database, which is a synthetic database [3], [4], [5]. In [2], Alefs *et al.* proposed a holistic recognition approach based on the texture information of the MMW images. On the other hand, the works by Moreno-Moreno *et al.* [4], [1], [3] put forward a biometric system based on geometric measures between different silhouette landmarks of the contour. There is also a previous work [5] in which some baseline techniques such as the contour coordinates, Euclidean distances and dynamic time warping algorithm are utilized in order to build a biometric system based on the contour information. This shortage of biometric recognition research based on MMW images is mainly due to the lack of databases of images of people acquired at 94 GHz. This lack is a consequence of: *i)* the privacy concerns these images present, and *ii)* most of the imaging systems working at the MMW/SMW band are either in prototype form or not easily accessible for research.

In this paper, we extend the previous work [5] by comparing multiple shape descriptors of the body contours. This is inspired by previous works, which show that recognition through the shape and boundary of traits such as the hand [6], [7] or the signature are fairly reliable [8], [9]. For this comparative study, we use some shape feature that have a large dimensionality whereas others such as silhouette landmarks present fewer points. Our aim with this work consists of assessing all these features regarding its performance, computational time and robustness.

Figure 1 draws a simple diagram explaining the different stages of the considered body shape biometric system. As can be seen, there are three principal stages: the contour extraction, the feature extraction stage and the comparison stage.

This paper is structured as follows. The database, and the procedure carried out to obtain the contours of people is explained in Section II. Section III describes the different feature extraction and similarity computation approaches used to compare the contours. The evaluation of these methods is performed in Section IV, and conclusions are finally drawn in Section V.

II. Database and Contour Extraction

The corpus of the BIOGIGA database consists of synthetic images at 94 GHz of the body of 50 individuals (25 males and 25 females). The images are the result of simulations carried out on corporal models at two types of scenarios (outdoors and indoors) and with two kinds of imaging systems (passive and active). These corporal models are previously generated using the software MakeHuman[1] based on body measurements taken from real subjects. Then, these models are imported to

[1] http://makehuman.org/

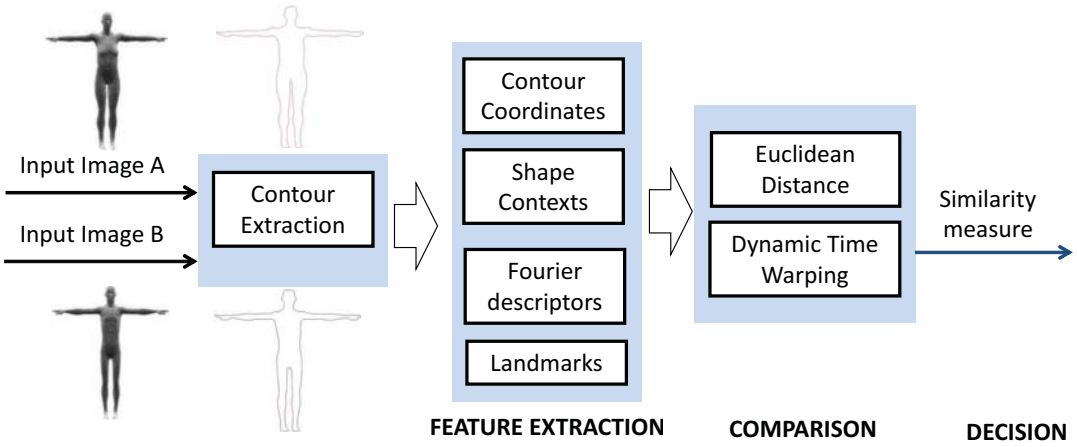


Fig. 1. General scheme of the body shape biometric system.

Blender[2], which simulates the effect of the 94 GHz radiation over the human models. A more detailed description of the generation of the BIOGIGA database can be found in [4].

In this paper, only passive images at outdoor scenarios are considered similarly as the previous work using a real database [2]. This subset of the database is comprised of 50 subjects, with 6 images per user. Three of them were simulated with clothes, and the other three were simulated without clothes to analyse the effect of clothing and have some variability between the images of the same person. For this, three angles between the subject and the camera were considered, having images with -10, 0 and +10 degrees. Figure 2 shows an example of the images from a single subject of the database. As can be seen, images with and without clothes are very similar as the 94 GHz band is almost transparent to clothes; however, the pixel intensity is a bit darker in the images with clothes and small parts of the clothes are still noticeable in the waist and neck.

The first processing step was to binarize the images, separating the background from the body. A characteristic of the images simulated by passive systems is the different grey level they present in different parts of the body. For instance the abdomen is much darker than the feet. This fact hinders the segmentation process. This problem was overcome performing the segmentation in two steps: *i)* border detection, and *ii)* morphological detection.

A Canny border detector is first applied to the image. The parameters of this detector has been empirically tuned. After that, various morphological operations are conducted on the resulting border image. These morphological operations consist of closing operations with different structural elements in different areas of the image (head, arms, from arms to calf, and feet). Finally, another set of morphological closing removes spurious irregularities, and leads to the final contour of the human body, which is used in the following experimental sections. Figure 3 shows an example of the process of segmentation and contour extraction for user 1.

[2]http://blender.org/

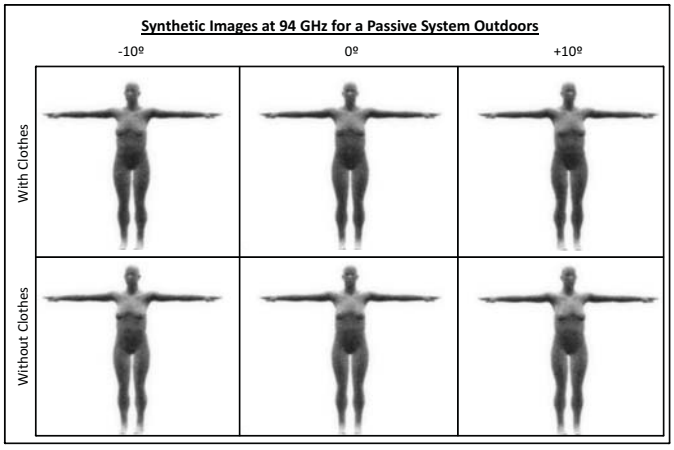


Fig. 2. Synthetic images of one user simulated at 94 GHz with a passive system and outdoors. The figure shows the three different camera angles available, and images with clothes and without clothes.

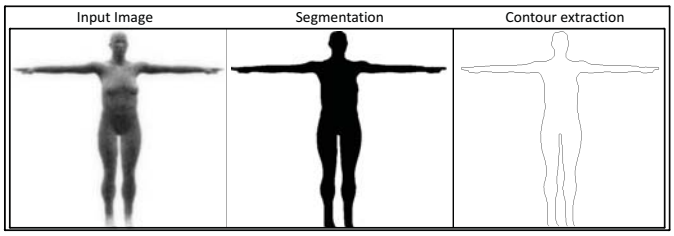


Fig. 3. Main steps followed in our system to extract the contour. From left to right: Original image (of a subject with clothing and a camera angle of +10 degrees), segmented image, contour extraction.

III. Shape-based Body comparison

This section describes the approaches based on the body shape followed in order to build the biometric system based on the contour information. Within the feature extraction and the classification stage, there are several techniques that may be applied. In this section, we will proceed to outline every technique proposed either for the feature extraction or the classification stage.

A. Shape descriptors

We have selected for the feature extraction stage four different approaches: *i)* contour coordinates themselves, *ii)* shape contexts (a complex descriptor proposed by [10]), *iii)* Fourier descriptors of the coordinates [11] and *iv)* silhouette landmarks which are a reduced set of points which describe the most discriminative parts of the silhouette [4].

1) Contour coordinates (CC): Contour coordinates are used as the baseline feature approach. Concretely, we mean by coordinate the 2-dimensional vector which specifies the x and y position of every single point within the silhouette of the body. The resolution of the contour is defined by the number of coordinates being the original resolution of the contours extracted from the MMW images of around 2800 points. The starting point of the sequence is the middle point of the head. Through subsampling techniques, we will be able to use different contour resolution ranging from 100 points up to 2800 points.

2) Shape contexts (SC): Shape contexts descriptors are first introduced by Belongie *et al.* [10]. This technique describes a specific point considering the relative distance and angle of the rest of the points within a shape. This method considers the set of vectors originating from a point to all other sample points on a shape. The number of radial bins and theta bins are the main parameters of this descriptor. As a result, the shape contexts of a shape with N points forms a vector of size $(N * r\_bins * \theta\_bins)$.

For a point p_i on the shape, we compute a coarse histogram h_i of the relative coordinates of the remaining $n-1$ points of the shape:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{1}$$

Equation 1 specifies that component k of histogram h_i contains the number of points different to point p_i that lies in $bin(k)$. The basic idea of shape contexts is illustrated in Figure 4, which shows an example of a shape contexts descriptor for two points in the eight digit. Note that the log-polar histogram used in this case has a dimension 12×5 (we decide to use the same configuration of parameters the author originally proposed), where 12 accounts for the number of theta bins and 5 accounts for the number of radial bins. Dark colours mean a high density of points within a bin, while ligther colours imply less density of points. In both cases the majority of points are quite distant from the specific point. Regarding the angle distance, in the first case (see Figure 4 (a) and (c)) the major density of points relies on the farthest angle distances, while for the second case (see Figure 4 (b) and (d)) there are approximately as many points in a medium angle distance as in the far angle distance.

In order to compute the similarity between two shape contexts, different distance methods or standard statistical methods may be applied.

3) Fourier descriptors (FD): Although Fourier descriptors [11] are a 40-year-old technique, they are still considered as a good description tool [13].

These descriptors are simple to compute and robust against translations and rotations since the effect these transformations cause on the descriptors is completely known. To apply this technique to our system, first we need to represent the contour coordinates as complex numbers (see Equation 2). Secondly, we apply the Fourier Transform to this complex numbers to obtain the Fourier description (see Equation 3).

Let (x_k,y_k), $K = 0,1,...N-1$ be the coordinates of N samples on the boundary of an image region. For each pair (x_k,y_k) we define the complex variable:

$$u_k = x_k + jy_k \tag{2}$$

For the N u_k points we obtain its DFT f_l

$$f_l = \sum_{k=0}^{N-1} u_k exp(-j\frac{2\pi}{N}lk), l = 0, 1, ..., N-1 \tag{3}$$

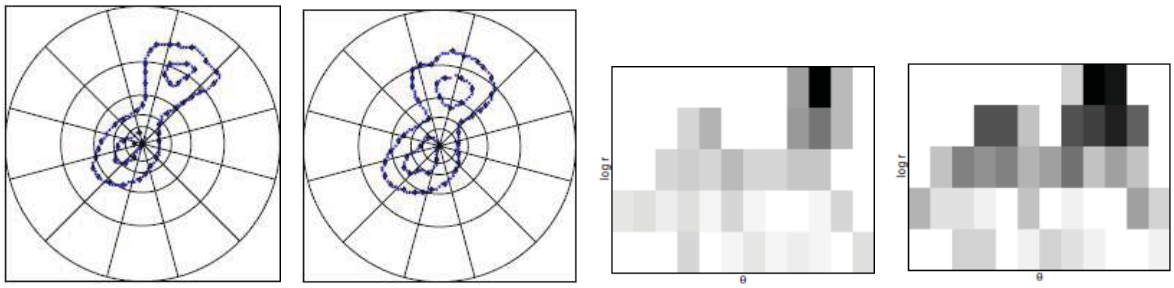
The coefficients f_l are also known as Fourier descriptors of the boundary. Once the f_l are available, the u_k can be recovered and the boundary can be reconstructed. The accuracy of the reconstructed boundary will depend on the number of Fourier coefficients used. If, instead of using all Fourier descriptors, we use only the first P coefficients in computing the inverse transformation, this is equivalent to setting $f_l = 0$ for $l > P-1$. This way the result of the inverse transformation will be an approximation to u_k. Although only *P* terms are used to obtain each component of $\hat{u}_k$, k still ranges from *0* to *N-1*. Hence, the same number of points exists in the approximate boundary, but not as many terms are used in the reconstruction of each point. Bearing in mind that high-frequency components account for fine detail, and low-frequency components determine the global shape, the discriminatory information of the shape is not lost in this approximation.

4) Landmarks (LM): This last approach is proposed as a possible feature for the feature extraction stage of the system. These landmarks consist of a reduced set of key points obtained in the work by [4]. Figure 5 shows an example of the situation of these 14 points. In particular, they depict the most singular parts of the people silhouette, among them: head, neck, hands, underarms, waist, hip, pubis and feet. Each landmark is caracterized by its position coordinates (x and y).

In this work, we also use these landmarks as features, evaluate the results obtained and compare them with the results achieved with the other approaches. Note that the dimensionality of these features is much smaller than in the other approaches.

B. Similarity computation

Regarding the classification stage, two types of distances are employed: the Euclidean distance and the dynamic time warping (DTW) algorithm - a more complex technique to compute a similarity between sequences [14].



(a) Point a (b) Point b (c) Histogram point a (d) Histogram point b

Fig. 4. Example of the computation of a shape contexts descriptor for two single points within the eight digit. Image a) and c) represent a point within the eight digit and its respective log-polar histogram while Image b) and d) a different point within the same digit and its associated log-polar histogram. Images extracted from [12].

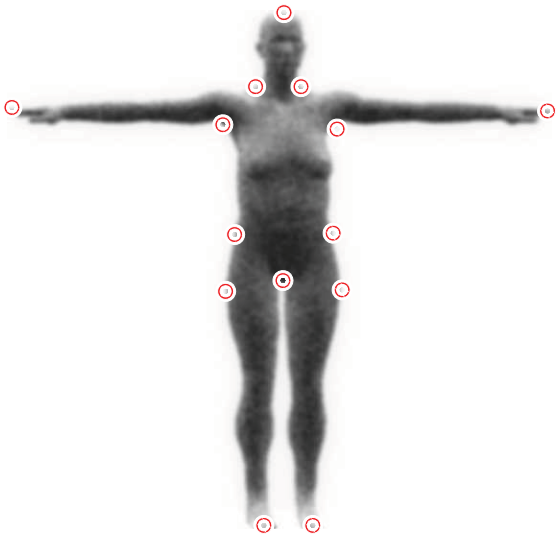


Fig. 5. Set of 14 points (landmarks) describing the silhouette of user 1.

1) Baseline technique: Euclidean distance (ED): This simple approach consists in computing a dissimilarity measure between the contour coordinates of two silhouette images. The only restriction of this method is the fact that distances need to be computed between sequences of the same length. Therefore, a normalization of the length of the sequences must be applied. Then, the Euclidean distance between the two normalized contours is computed following equation 4,

$$ED = \sum_{j=1}^{N} \sum_{i=1}^{2} \sqrt{((a_{ij} - b_{ij}))^2} \tag{4}$$

where a, b represent the sequence of contour coordinates of images a and b respectively; $i = 1, 2$ represents the number of coordinates describing each contour point, in this case: x and y axes, and $j = 1, ..., N$ defines every point of a contour, assuming that every contour is characterized by N points.

2) Dynamic Programming: dynamic time warping (DTW): The goal of DTW is to find an elastic match among samples of a pair of sequences that minimize a given distance measure. In the biometric field, it was first used for signature verification [8].

In this work, DTW is used to obtain a cumulative distance between two strings of coordinates that is known to be minimal. Equation 5 shows the transformation of this minimal distance into a similarity score where K is a normalization factor that takes into account the number of aligned points between two sequences. One of the main advantages of this algorithm is the possibility of dealing with sequences of point that do not have the same dimensionality. In our case, contours from different images do not have the same amount of points.

$$score = e^{-\frac{DIST}{K}} \tag{5}$$

IV. Experiments

This section describes the experimental work carried out to analyse the performance of the different approaches described in Section III. The aforementioned methods are tested with the contour coordinates of the BIOGIGA database previously described in Section II.

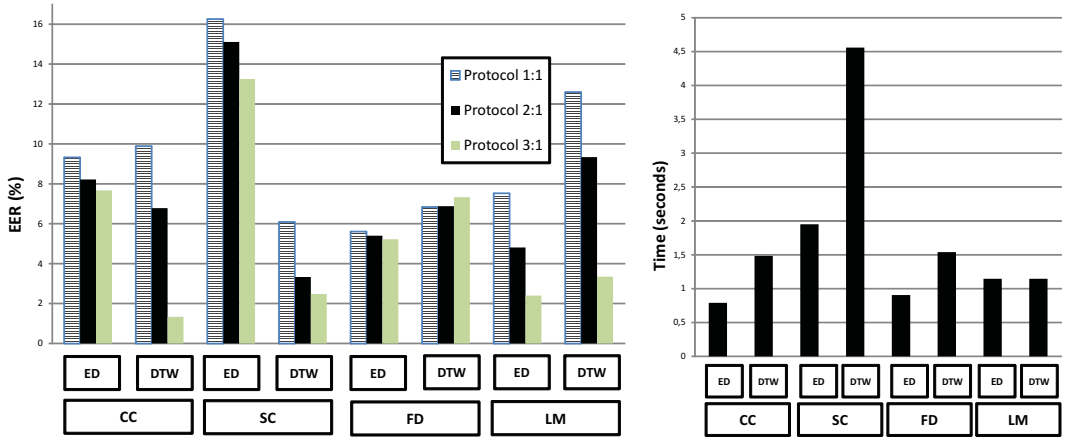
In this work, three different experimental protocols are considered: *i)* protocol 1:1, *ii)* protocol 2:1 and *iii)* protocol 3:1, where the first number refers to the number of training images considered per user, and the second number to the number of test images per user, one in all cases. In order to have the most challenging scenario with severe mismatch between enrolment and testing regarding clothes, the database is divided into two sets, where the images with clothes are used for the training and the images without clothes are used for the test.

It is worth noting that when having 2 or 3 images for training, the fusion of the information contained in the images is carried out at the score-level, i.e., all single comparisons between training and test are done image by image, and then the scores are fused using a sum rule.

As mentioned in Section III, experiments based on DTW are analysed with all contours having their original size. ED experiments are carried out with contours normalized to the same size.

A. Results

In this experiment, we compare the performance of the different approaches. Bearing in mind that there are four differ-



(a) Performance (b) Computational time

Fig. 6. Performance (EER%) of the eight different approaches for 2800-points contours and for Protocols 1:1, 2:1 and 3:1. ED=Euclidean Distance, DTW=Dynamic Time Warping, CC=Contour Coordinates, SC=Shape Contexts, FD=Fourier Descriptors and LM=Landmarks.

ent feature extraction approaches and two different similarity measures, we have a total of eight possible biometric system configurations. Figure 6 a) shows the performance in terms of equal error rate (EER) for each of these approaches and the three different protocols (P1:1, P2:1 and P3:1).

First, we can observe from Figure 6 a) that the EER of the system decreases as the number of training images increases (Protocol 3:1). It is also worth noting the remarkable improvement of performance when applying the DTW algorithm instead of the baseline Euclidean distance to the contour coordinates and shape contexts approaches (specially in P3:1). Applying DTW to the Fourier descriptors do not result in better performance of the system since we deal with transformed features that do not represent the shape of the body anymore. In the case of the landmarks features, it is not worth applying DTW to a feature vector of such a small dimensionality.

Regarding the average performance between all protocols, the best approaches are CC-DTW (6%), SC-DTW (3.96%), FD-ED (5.41%) and LM-ED (4.91%). For the 3:1 protocol, a 1.33% of EER for the CC-DTW approach is achieved. However, considering the other two protocols, it can be seen that the performance of the CC-DTW approach is worse compared to other cases. For example, the SC-DTW produces lower EER rates for protocols 1:1 and 2:1. In the case of the FD-ED approach, we observe that the number of training images do not imply considerable variations in the performance. Regarding the LM-ED approach, we see that using a vector with a dimensionality quite smaller than the vectors used in any of the previous approaches, can produce comparable results to the best approaches. However, in a real database we believe the localization of these landmarks points would be not as robust as in this synthetic database.

The computation cost invested in each approach should not be forgotten during the assessment process (Intel i7-3770 CPU @ 3.4 Ghz RAM 8GB in Matlab R2012b). Figure 6 b) depicts the time in seconds spent during all the process, taking into account the time needed to compute the features and the time to compare a pair of feature-vectors. The main conclusion we can extract from this time-comparison is the fact that DTW algorithm implies an increment of the computational time. This fact is magnified when dealing with shape contexts features, mainly caused by the huge dimensionality of this vector. Aiming to reduce the computational time, we propose for future work to apply PCA to shape contexts. Doing this, the shape contexts approach would be much more feasible regarding real time applications.

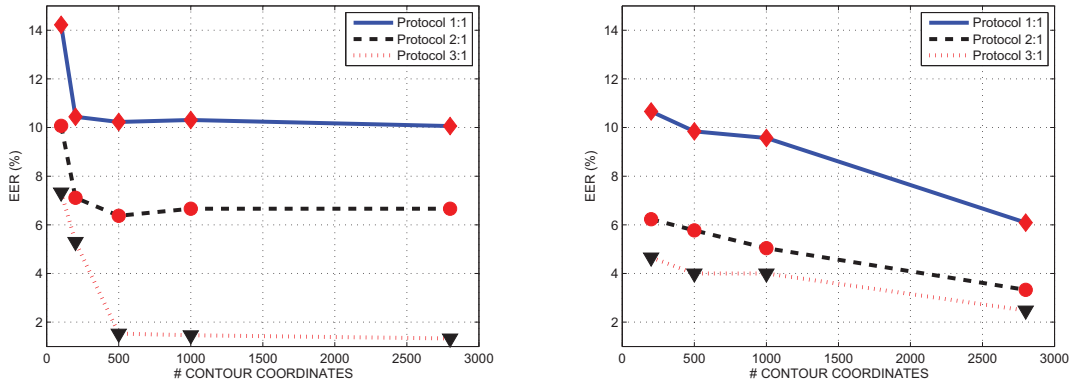
B. *Effect of contour resolution*

A second experiment is carried out to analyse the effect of the contour size in the recognition performance, for the cases of contour coordinates and shape contexts using DTW. Figure 7 a) and b) represent the performance of the system against the resolution of the contour. Considering the case of CC-DTW it is very interesting to note a notable drop of EER when dealing with a contour resolution of more than 500 points. Even though the EER obtained with the bigger resolution (2800 points) is slightly better that the EER obtained with 500 points, there is a big difference in terms of computional time between using a 2800-CC-DTW approach rather than a 500-CC-DTW one. Concretely, the computational time invested drops from 1.48 to 0.81 seconds when reducing the resolution of the contour down to 500 points. This is an important issue to bear in mind for real-time applications.

Conversely, when using shape contexts descriptors the EER drops as the resolution of true contour increases but there is no clear knee point as in the previous case. In this case, we need to work with the highest possible resolution.

V. Conclusions

In this paper, a complete body shape biometric system has been developed for MMW body images (BIOGIGA database). The use of MMW images instead of images acquired at other spectral bands presents some advantages, mainly the transparency of clothing at that frequency, allowing to extract easily the contours from the images. Different approaches have been analysed ranging from naive approaches such as contour coordinates for the feature extraction stage or the Euclidean distance for the classification stage to complex schemes such as shape contexts or Fourier descriptors for the feature extraction stage or the dynamic time warping algorithm for the similarity



(a) Contour coordinates-DTW

(b) Shape contexts-DTW

Fig. 7. Effect of the resolution with Contour Coordinates and Shape Contexts from 100 points up to 2800 points.

computation stage. The best results are obtained when using the DTW algorithm directly to the contour coordinates for the contours with the highest resolution for protocol 3:1 (1.33%). For other protocols, the use of Fourier descriptors or Shape contexts may be quite reasonable bearing in mind the computational cost of Shape contexts.

The limitations of this work are related to the special characteristics of the database used explained previously.

Results from previous work [2], achieve similar performance results but using only texture information of the torso image. We believe that a fusion of both approaches, i.e, body shape and texture would improve the results significantly. We also propose for future work to carry out further studies so as to kwow whether there is any difference in performance when gender is taken into account.

ACKNOWLEDGMENTS

This work has been partially supported by projects TeraSense (CSD2008-00068), Bio-Shield (TEC2012-34881), Contexts (S2009/TIC-1485) and "Cátedra UAM-Telefónica". E. Gonzalez-Sosa is supported by a PhD scholarship from Universidad Autonoma de Madrid.

REFERENCES

[1] M. Moreno-Moreno, J. Fierrez, and J. Ortega-Garcia, "Biometrics beyond the visible spectrum: Imaging technologies and applications," in *Proc. Biometric ID Management and Multimodal Communication*. Springer, 2009, pp. 154–161.

[2] B. Alefs, R. den Hollander, F. Nennie, E. van der Houwen, M. Bruijn, W. van der Mark, and J. Noordam, "Thorax biometrics from millimetre-wave images," *Pattern Recognition Letters*, vol. 31, no. 15, pp. 2357–2363, 2010.

[3] M. Moreno-Moreno, J. Fierrez, R. Vera-Rodriguez, and J. Parron, "Simulation of millimeter wave body images and its application to biometric recognition," in *Proc. of SPIE Vol*, vol. 8362, 2012.

[4] ——, "Distance-based feature extraction for biometric recognition of millimeter wave body images," in *Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST*, 2011, pp. 1–6.

[5] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Body shape-based biometric recognition using millimeter wave images," in *Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST*, 2013.

[6] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1803–1815, 2006.

[7] J. Burgues, J. Fierrez, D. Ramos, and J. Ortega-Garcia, "Comparison of distance-based features for hand geometry authentication," ser. LNCS-5707. Springer, 2009, pp. 325–332.

[8] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern Recognition*, vol. 35, no. 12, pp. 2963 – 2972, 2002.

[9] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, "Hmm-based on-line signature verification: feature extraction and signature modeling," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2325–2334, December 2007.

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[11] E. Persoon and K.-S. Fu, "Shape discrimination using fourier descriptors," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 7, no. 3, pp. 170–179, 1977.

[12] H. Zhang and J. Malik, "Learning a discriminative classifier using shape context distances," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. I–242.

[13] M. Yang, K. Kpalma, J. Ronsin *et al.*, "A survey of shape feature extraction techniques," *Pattern recognition*, pp. 43–90, 2008.

[14] M. Yasuhara and M. Oka, "Signature verification experiment based on nonlinear time alignment: a feasibility study," *IEEE Trans. Systems Man Cybernet*, vol. 17, pp. 212–216, 1977.