Image and Vision Computing 32 (2014) 1080–1089

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Editor's Choice Article

Using a Discrete Hidden Markov Model Kernel for lip-based biometric identification ☆

Carlos M. Travieso [a], Jianguo Zhang [b], Paul Miller [c], Jesús B. Alonso [a]

[a] Signal and Communications Department, Institute for Technological Development and Innovation in Communications, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, sn, Ed. de Telecomunicación, Pabellón B, Despacho 111, E35017 Las Palmas de Gran Canaria, Spain
[b] School of Computing, University of Dundee, Scotland, United Kingdom
[c] The Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Northern Ireland Science Park, Queen's Road, Queen's Island, BT3 9DT Belfast, United Kingdom

ARTICLE INFO

Article history:
Received 17 October 2013
Received in revised form 31 May 2014
Accepted 2 October 2014
Available online 22 October 2014

Keywords:
Discrete Hidden Markov Model Kernel
Image processing
Lip-based biometrics
Pattern recognition

ABSTRACT

In this paper, a novel and effective lip-based biometric identification approach with the Discrete Hidden Markov Model Kernel (DHMMK) is developed. Lips are described by shape features (both geometrical and sequential) on two different grid layouts: rectangular and polar. These features are then specifically modeled by a DHMMK, and learnt by a support vector machine classifier. Our experiments are carried out in a ten-fold cross validation fashion on three different datasets, GPDS-ULPGC Face Dataset, PIE Face Dataset and RaFD Face Dataset. Results show that our approach has achieved an average classification accuracy of 99.8%, 97.13%, and 98.10%, using only two training images per class, on these three datasets, respectively. Our comparative studies further show that the DHMMK achieved a 53% improvement against the baseline HMM approach. The comparative ROC curves also confirm the efficacy of the proposed lip contour based biometrics learned by DHMMK. We also show that the performance of linear and RBF SVM is comparable under the frame work of DHMMK.

1. Introduction

Personnel security is becoming increasingly important in today's modern world [1]. Biometric-based access control is one of the most important technologies for cyber–physical security, and has received increasing attention over the past two decades. In the competitive business world of today, the need and demand for a biometric physical security solution have never been higher. The biometric market is increasing each year and this trend is set to continue, due to the increasing need for security at borders, and in buildings, airports, etc. [2]. At its core, it aims to identify a person with one or more of their body features, such as their face, hand, fingerprint, or voice [1–3]. These biometric modalities can be deployed for different applications including; searching for people, remote access control, and secure corridors in airports.

To date, there has been a large amount of work done in biometrics, with most of it focusing on using a single biometric mode. Recently, the trend has been to build robust person identification systems based on multimodal approaches, i.e., a combination of biometric features. However, to obtain a robust multimodal solution, it is of benefit to employ individual modalities which have good performance in isolation. Furthermore, there is still much room for improvement with respect to single mode approaches.

Human recognition through distinctive facial features supported by an image database is still an appropriate subject of study as already mentioned. We should not forget that this problem still presents various difficulties. For example, what will happen if an individual's haircut is changed? Is make-up a determining factor in the process of verification? Would it significantly distort facial features? For these reasons, the study of different parts of a face still merits investigation in order to improve identification. Consequently, the analysis of lip contours is receiving greater attention [4,5], as it is particularly well-suited to deployment on mobile phone platforms. The importance of lip features as biometrics is reported in [6], where numerous lip-based features are evaluated. Therefore in this work, an approach based on the shape of lips is presented.

1.1. Related work

In this section, we briefly review work closely related to ours. Early work in this area involved the tracking of lips, using features extracted from color distributions around the lip area [7]. The resulting feature dimensionality was reduced using principal component analysis (PCA), and classification was performed by linear discriminant analysis. By combining this approach to lip movement analysis with speech analysis, a significant improvement in speaker verification in noisy conditions

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Josef Bigun.

E-mail addresses: carlos.travieso@ulpgc.es (C.M. Travieso), jgzhang@computing.dundee.ac.uk (J. Zhang).

http://dx.doi.org/10.1016/j.imavis.2014.10.001
0262-8856/

was demonstrated. In [8], the modeling of lip movements by hidden Markov models (HMMs) is presented. Each lip movement clip is represented by 2D discrete cosine transform (DCT) coefficients of the optical flow vectors within the mouth region. In [9], speech, lip movements and face images are combined to give robust person identification. In this work, DCTs of intensity normalized mouth images were employed to provide static features. These were then combined with an HMM to classify the speaker via log-likelihood.

Rather than recognizing a speaking person, research by Newman and Cox tries to determine the language a person is talking in by recognizing their lip movements when speaking a specific passage of text [10]. For this, they use Active Appearance Models (AAM) to locate the face and mouth, and produce a vector that represents the lip shape for each video frame. They obtain recognition results of 100% for seventy-five different languages for a single speaker. Subsequently, Newman and Cox [11] modified the classification system to obtain speaker independent language recognition, obtaining 100% classification accuracy for five bilingual speakers — even with a viseme classification accuracy of as low as 40%.

Another field in which lip contour extraction is used is in facial expression recognition as described by Raheja et al. [12]. They studied three facial expressions by processing an image of a face. To do this, they extract the lip contour by edge detection, generate a binary image, post-process to fill in holes, and perform a histogram analysis of the binarized image for classification. Using this system, they achieve a recognition rate of up to 95%.

There have been various investigations into recognizing a person from their lips. In one of these by Mehra et al. [13], PCA is used to obtain feature vectors of reduced dimension, which are then input to a neural network for classification. They achieve an accuracy rate of 91.07%. In [14], a novel ordinal contrast measure, called Local Ordinal Contrast Pattern, is proposed for representing video of the mouth region of a speaker while talking. This has been used in a three orthogonal plane configuration as input to a speaker verification system. Verification was accomplished using the chi-squared histogram distance or LDA classifiers, obtaining a half total error rate of less than 1%. Wang and Liew [6] studied the roles of different lip features, related to both physiological and behavioral properties of lips, in personal identification, and demonstrated that though dynamic features achieve higher recognition accuracy, both dynamic and static features are promising biometrics for verification. In [15], a new approach to speaker verification using video sequences of lip movements is proposed, in which a Motion History Image is used to provide a biometric template of a spoken word for each speaker. A Bayesian classifier is used for classification, obtaining an average recognition rate of 90% at a false alarm rate of 5%. In another work, a new motion based feature extraction technique for speaker identification using orientation estimation in 2D manifolds is reported [16]. The motion is estimated by computing the components of the structure tensor from which normal flows are extracted. By projecting the 3D spatiotemporal data to 2-D planes, projection coefficients are obtained which are used to evaluate the 3-D orientations of brightness patterns in TV like image sequences. An implementation, based on joint lip movements and speech, is presented along with experiments demonstrating a recognition rate of 98% on the publicly available XM2VTS database.

There exist a number of approaches for lip contour extraction, or lip corner detection, for visual speech/speaker recognition. For example, [17] used a monochrome image histogram to detect lip corners. However, it is more common to use color images, such as RGB [6] and HSV images [18]. Prewitt and Sobel operators are employed to detect lip edges in [19]. In [20], a manifold based approach is introduced to extract the lip contour. The red exclusion method [17] is widely used, due to its simplicity and efficiency. Similar to the approach in [21,22] is based on an RGB transformation of the lip regions. The resulting transformation and the b component of the CIELAB color space, are then used for the clustering phase. The task is formulated as finding the optimum partitioning of a given color image into lip and non-lip regions. The partitioning utilizes multispectral information in a color image, instead of just the limited information in a single component gray-scale image.

In [23], an approach is presented for a system based on lip reading. In [24,25], HMM based visual speech recognition is described. Several studies have found that a multimodal approach to speech recognition, combining speech, mouth and face features, gives significantly improved performance over a single mode approach under trying test conditions [1,3,26–28].

1.2. Innovation: the present work

To date, most of the existing work concentrates either on lip movement, or the combination of lip movement and static features. While multimodal lip biometrics could be likely made more successful [6], there is little dedicated attention sufficiently paid into static lip shape alone. It would be easy to get a high accuracy via a multimodal approach with more biometrical information; however, developing a high performance system based on solely one modality is usually a more challenging task. In cases when another type of biometric features is not accessible, e.g., only a static lip image present while lip dynamics are unavailable, the high performance of a single biometric modality would be vital to maintain the good performance of the whole system.

In this work, we focus on the single mode approach, based on static shape information. Motivated by our previous work on modeling the shapes of offline signatures using HMMs [29], and the recent successful use of shape for other applications [30–33], we present a novel biometric identification approach based on lip contours encoded by an HMM kernel, and learned by an SVM. Though the proposed technique is somewhat incremental, the difference in results is significant. We would like to emphasize that the proposed approach achieves an improvement in classification accuracy of up to 53%, using only two training images, compared to the standard HMM approach. To the best of our knowledge, a study in lip-based biometrics using DHMMK has not been carried out before.

The rest of this paper is organized as follows. Section 2 introduces the lip extraction method, and Section 3 presents the lip descriptor. Section 4 describes our classification scheme and our kernel. Experiments and results are given in Section 5. Discussions and conclusions are presented in Section 6.

2. Preprocessing: lip extraction

Our approach consists of three main parts; lip extraction, DHMMK parameterization, and classification. A block diagram of our method is presented in Fig. 1. The first stage of our approach is to extract lips from an image by firstly detecting the face, then the mouth and finally the lips. Fig. 2(a) shows a typical color face image from one of our datasets. Face detection is achieved using the popular Viola and Jones face detector [34], which gives detection rates of 99% over the three datasets we use.

Once the face is detected, our next step is to localize the mouth region. It has previously been shown that face parts such as the eye, nose and mouth usually have very strong geometrical relationships [26]. For example, the mouth region is always located in the lower part of a detected near-frontal face. This fact makes it possible to use a simple heuristic approach to localize the mouth region without the need to build an advanced mouth detector. So, we take the lower half region of the detected face as a rough estimate of the mouth region. To remove any boundary effects, we further exclude a few pixels from the face boundary. The segmented mouth region is indicated by the red bounding box shown in Fig. 2(b). Though this 'guessing' is heuristic, it does remove some background facial regions in preparation for the next step, lip contour extraction.

Motivated by the work in [5,17], we use an RGB transformation to enhance the lip region based on the fact that lip color usually has a

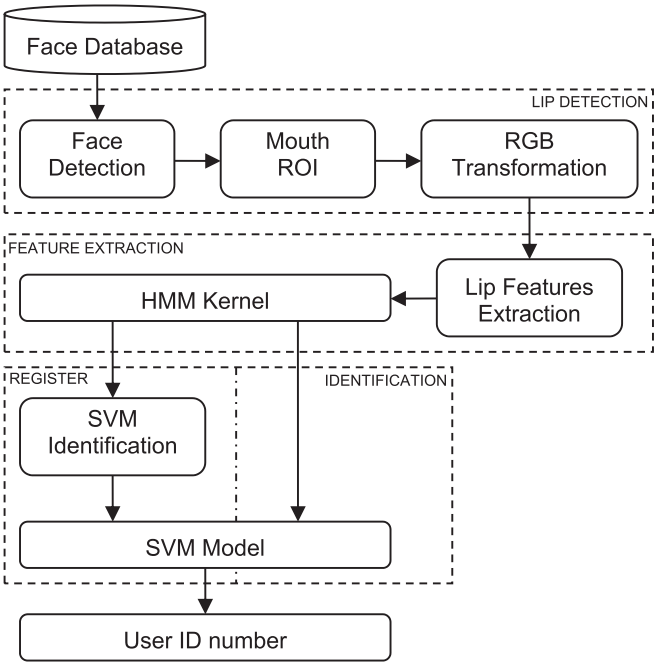

Fig. 1. Block diagram of the proposed approach.

stronger red component than other parts of the mouth [17]. Therefore, we apply the following simple transformation on the mouth region to convert the color image into gray scale [35]:

$$I = R - 2.4 \cdot G + B. \quad (1)$$

The effect of this transformation is clearly shown in Fig. 2(c). The lip region becomes much brighter than the background pixels of the mouth region. We then employ the Otsu binarization method [36] to segment the lips in the enhanced image. The lip contour is obtained by dilating the segmented lips with the following 3×3 morphological operator $se = [1\ 0\ 1;\ 1\ 1\ 1;\ 1\ 0\ 1]$. The undilated segmented image is then subtracted from the dilated one leaving only the lip contour, Fig. 2(d). Figs. 2 and 3 show the whole of the lip contour extraction process using exemplar face images from each of the three datasets. From these examples, it is clear that our simple lip extraction approach works very well, even on faces with extensive facial hair, as shown in Fig. 3.

3. Feature extraction

Following extraction of the lip contour, we create a lip descriptor. For this we consider two different geometrical–sequential features on two grid layouts; rectangular and polar. This step transforms the 2D contour into a one-dimensional feature vector. The rectangular grid features are the Euclidean distances from sample points along the vertical and horizontal axes to points on the lip contour. For the vertical axis, we equally sample 180 points, while for the horizontal axis we sample 300 points (Fig. 4a). Thus, the resulting normalized feature vector has 480 elements after concatenation.

The polar grid features are motivated by the work of [29]. These have been proven to be a powerful descriptor for offline signature verification. The set of polar features is calculated from the centroid of the lip contour. The orientation is sampled from 0 to 360° with a sampling interval of one degree. For each angle, the radius is computed as the distance between the point of intersection on the contour and the center (Fig. 4b). Hence, the resulting normalized feature vector consists of 360 elements after concatenation. We have found that this feature size gives better accuracy rates.

4. Methods: classification systems and kernel

In this section, classification approaches are described. We used two classification approaches Hidden Markov Model (HMM) [37] and support vector machine (SVM) [38]. HMM has been deployed in order to model the sequential information from our features, in a similar fashion to [37]. Finally, these features have been transformed based on an HMM kernel learnt by SVM.

4.1. Hidden Markov Model

HMMs have become increasingly popular over the past two decades. They are theoretically sound and usually perform very well in real applications such as speech recognition [39]. In this section we review the theoretical aspects relating to our work. An HMM has two associated stochastic processes; *dynamics* and *observations*. The former is not visible and is usually modeled by the probability of transition between hidden states. The observation process is modeled by the probability of obtaining an observed value given a hidden state (see [37,40] for a complete treatment of HMM). In our paper, we use a discrete HMM (DHMM) [37] since it forms the basis of our DHMMK in the next section. A DHMM consists of the following parameters:

1) The number of states N
2) The number of different observations M

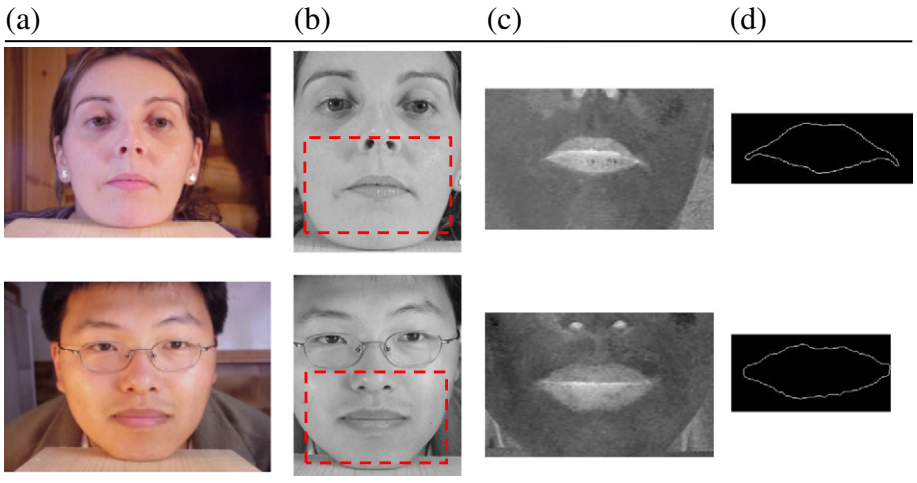

Fig. 2. Examples of the lip detection process. Original images, (a), face and mouth ROI, (b), enhancement, (c), extracted contour, (d).

Fig. 3. Examples of the lip detection process for PIE face database. Original images, (a), extracted contour, (b).

3) The transition probability matrix $a(N,N)$
4) The initial state probability $\pi(N,1)$
5) The observation probability matrix b (N,M).

A DHMM is very powerful at modeling time sequences where events at different times have some causal relationship. However, it is also possible to extend this concept to modeling the dependencies between parts of a shape. In particular, N is going to represent a little sequence or segments of widths and/or heights. An example of this usage is for offline signature verification, where the shape of a signature is modeled by a DHMM [29]. We view the offline signature shape and lip shape is a similar problem. The online signature clearly offers both temporal information, i.e., a clear temporal ordering of the points, and the geometrical dependence of those points. However, the offline signature is a shape, thus only offers the geometrical dependence of those points. This has been verified by our previous work of using HMM for offline signature verification using the polar grid features [29]. Under the sample umbrella, lip contour can be considered as a signature of face, and we consider that there exists potentially dependence between the points on the lip contour. Lip contour could also be considered as a one-dimensional manifold, and is an ordered sequence in this space. All of this motivates us to use the DHMM to model our lip contour descriptors. Hence, "left to right" DHMMs turn out to be especially appropriate for lip contours because the transition through the states is produced in a single direction. This equips the model with the ability to maintain a certain ordering with respect to the observations produced, where each sequential element of geometry distance is among the most representative changes (see Fig. 4).

In the DHMM approach, the first step is to quantize the feature vector elements. We use the K-Means algorithm for clustering to create a set of symbols that are required by the DHMM as in [41]. These symbols construct a set of states (with a "left to right" structure) that a discrete observation at a step t could be taken from. For our case, t represents the step between the feature vector elements. This approach is similar to the one used by [29,40]. The number of observed symbols, M, is determined by experimentation. We use thirty-two in all of our experiments. Based on Markov assumptions, given an observation sequence $X = [x_1, x_2, x_3, \ldots, x_t]$ with step t, and a trained DHMM $\lambda = [a,b,\pi]$, the probability of X given λ can be calculated as follows:

$$P(X/\lambda) = \sum_{S} \prod_{i=1:t} P(x_i/s_i, a) \cdot P(s_i/s_{i-1}, b) \tag{2}$$

where $S = [s_1, s_2, s_3, \ldots, s_t]$ denotes the variables for the hidden states. Thus for a given sequence, the optimal HMM can be selected by maximizing the posterior probability over a set of C trained HMMs:

$$\lambda_0 = \max_{j \in C} P\left(\lambda_j / X\right). \tag{3}$$

4.2. Discrete Hidden Markov Model Kernel

Though the HMM has achieved great success in many applications, the learning of this model is based on maximizing the marginal probability of the observations over the hidden states. Thus, it is a generative method and does not fully utilize the inter-class discriminative information presented in the training set. A natural way to address this weakness is to incorporate it into a discriminative learning framework, such as an SVM. This can be achieved by computing the Fisher score of an observation sequence over the learned DHMM parameters, which is calculated by the gradient of the parameter space [42].

From [42], we can calculate that gradient of the logarithm of the probability in Eq. (2) with respect to the HMM parameter λ, as follows (we call this a DHMM kernel):

$$U_{(X/\alpha)} = \nabla_{\alpha_{i,j}} \log P(X/\lambda) = \frac{\xi\left(x, s_j\right)}{\alpha_{i,j}} - \xi\left(s_j\right) \tag{4}$$

$$U_{(X/\beta)} = \nabla_{\beta_{i,j}} \log P(X/\lambda) = \frac{\xi\left(s_i, s_j\right)}{\beta_{i,j}} - \xi\left(s_j\right) \tag{5}$$

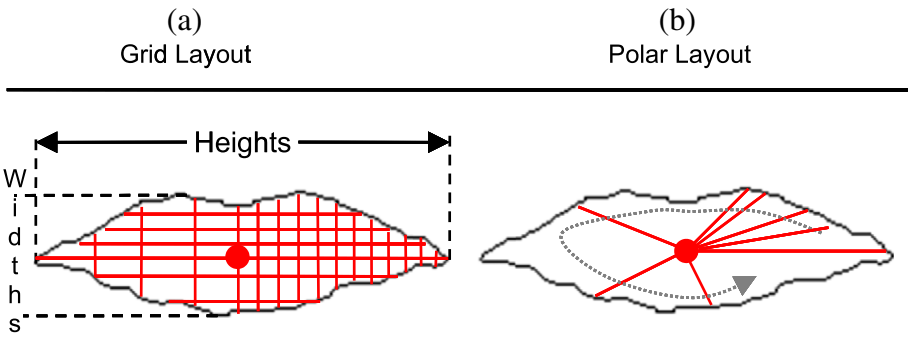

Fig. 4. Geometrical–sequential features: rectangular grid (a) and polar grid (b).

where $\xi(x, s_j)$ represents the number of times a certain symbol x is generated by a state s_j in a sequence, while $\xi(s_i, s_j)$ denotes the frequency of the joint occurrence of two states s_i and s_j at two adjacent time intervals over a sequence, and α and β are the forward and backward variables, respectively. These are in fact the sufficient statistics for the emission probability $a_{i,j}$ and transition probability $b_{i,j}$ given a sequence. $\xi(s_j)$ represents the frequency of state s_j occurring in a sequence [37,42]. These values can be directly obtained from the forward–backward algorithm [42].

The DHMM kernel vector U_X for a given sequence X is simply the concatenation of the two gradient vectors calculated from Eqs. (4) and (5) respectively. The length of the resulting feature vector for a sequence is $N \times (M + N)$.

Thus, the similarity of two sequences with a learned HMM could be evaluated in a kernel fashion using the corresponding Fisher score vector as follows

$$K(X, Y) = K(U_X, U_Y). \tag{6}$$

where $K(\cdot)$ could be any type of standard kernels for an SVM. (We have used the gpdsHMM tool [41].)

5. Experiments and results

5.1. Datasets

Our study is carried out on three different datasets; GPDS-ULPGC [30], the PIE dataset [43] and the RaFD database [44]. The GPDS-ULPGC dataset was collected by us specifically for this study. It consists of fifty users with ten samples per user (thus 500 images in total). The database is composed of 54% males and 46% female, with ages ranging from ten to sixty. Each sample is a color image of size 768 × 1024 pixels. It is available for downloading from [30]. The PIE dataset is a publicly available dataset [43] composed of sixty-eight subjects, with eleven samples per subject (thus giving 748 images in total), where each sample is a color image of size 200 × 300. The main characteristic of the dataset is it contains illumination changes and different hair styles (e.g., bearded and beardless, as shown in Fig. 3). The RaFD Face Dataset is composed of sixty subjects with nine samples per subject (thus giving a total of 540 images) [44]. The image resolution is 1024 × 681 and the database contains eight facial expressions for each subject. Since our study focuses on static lip features, only three of the eight expressions present in the database (neutral, sadness and indifference) suit our purpose, and are used in the experiments. Furthermore, images containing non-frontal poses for subjects with their mouth open have been removed.

5.2. Experimental methodology

We use a multi-class SVM for classification, which is built using the one-versus-all strategy. The SVM_*light* [45] implementation is utilized in our experiment. We tried two different kernels in our experiment; linear and radius bias function (RBF). It is well-known that the RBF kernel usually works better than the linear kernel [46]. In our study, we show that the recognition performance of the linear SVM is comparable to RBF SVM with DHMMK, while the linear SVM is usually much faster than RBF. Specifically, we use the DHMMK output, U_X, from Eqs. (4) and (5), as the input to the SVM. For the RBF kernel, the optimal value of the gamma parameter is found by a grid search.

Our experiments are performed using the well-known hold-out cross validation [47]. We report our results in terms of classification accuracy based on ten different runs/splits separately on each of the three different datasets. For each run/split, we randomly choose a number of images per subject for training and the rest for testing. The sampling is carried out without replacement to ensure there is no overlap between the training and test sets. In this way, we have reduced the risk of over-optimistic results from traditional cross validation experiments on small sample domains [47]. The mean and standard deviation of the classification accuracy of each run over all classes are reported.

We test the performance of our algorithm with respect to three different parameter settings: 1) different features, rectangular and polar grids; 2) number of states, N; 3) number of training images. We vary the number of training images from one to five per class. Combinations of these settings are comparatively tested with DHMM, SVM and DHMMK–SVM, respectively. Results on the GPDS-ULPGC dataset are summarized in Tables 1–6, while Tables 7–12 give the results obtained with the PIE and RaFD datasets. The idea is to validate our approach, which has been constructed using the GPDS-ULPGC dataset, using two independent, or blind, datasets — PIE and RaFD, thereby, demonstrating the robustness of our approach.

5.3. DHMM experiments

In order to highlight the performance of the proposed DHMM kernel based approach, we first perform experiments using the standard DHMM method without the kernel trick. Specifically, we use a similar training approach as in [29], since it proved very successful for offline signature verification. For the DHMM, we compare the performance of rectangular and polar grids by varying the number of states from 40 to 140. We use half the sample subjects for training and the rest for testing. i.e., five training samples per class. Results are shown in Table 1.

From Table 1, we can see that rectangular grid feature clearly outperforms the polar grid feature. This might be because the geometry constraints of the rectangular grid are considerably stronger than for the polar configuration. Table 1 also shows that increasing the number of states improves the accuracy for both feature types, particularly the

Table 1
DHMM classification results for both the rectangular and polar grid features with different number of states, 5 training images per class.

Features	# training samples	State numbers	HMM success rate
Polar layout	5	40	32.22% ± 7.19
Polar layout	5	65	36.01% ± 7.34
Polar layout	5	90	45.51% ± 6.68
Polar layout	5	115	42.43% ± 7.13
Polar layout	5	140	39.52% ± 7.49
Grid layout	5	40	54.16% ± 5.41
Grid layout	5	65	72.96% ± 2.31
Grid layout	5	90	78.67% ± 2.71
Grid layout	5	115	81.53% ± 6.03
Grid layout	5	140	**80.26% ± 2.73**

The bold color indicates the best accuracy under the mean and standard deviation points of view. If the standard deviation is low means better stability of the mean value.

Table 2
DHMM classification results for rectangular grid feature with 140 states, different number of training images per class.

Features	# training samples	State numbers	HMM success rate
Grid layout	5	140	**80.26% ± 2.73**
Grid layout	4	140	79.52% ± 2.73
Grid layout	3	140	76.00% ± 4.56
Grid layout	2	140	65.20% ± 2.45
Grid layout	1	140	59.47% ± 5.38

The bold color indicates the best accuracy under the mean and standard deviation points of view. If the standard deviation is low means better stability of the mean value.

Table 3
SVM classification results for both the rectangular and polar grid features with different kernels, 5 training images per class.

Features	# training samples	Kernel	SVM success rate
Polar layout	5	Linear	87.61% ± 2.34
Polar layout	5	RBF	88.14% ± 2.14
Grid layout	5	Linear	95.41% ± 1.41
Grid layout	5	RBF	96.02% ± 1.89

Table 4
SVM classification results for the rectangular grid feature with 140 states, different number of training images per class.

Features	# training samples	Kernel	SVM success rate
Grid layout	5	Linear	95.41% ± 1.41
Grid layout	5	RBF	96.02% ± 1.89
Grid layout	4	Linear	94.33% ± 1.82
Grid layout	4	RBF	95.03% ± 2.07
Grid layout	3	Linear	89.35% ± 1.89
Grid layout	3	RBF	91.14% ± 3.05
Grid layout	2	Linear	85.31% ± 3.61
Grid layout	2	RBF	87.37% ± 2.95
Grid layout	1	Linear	64.71% ± 4.23
Grid layout	1	RBF	65.34% ± 4.52

rectangular grid feature. The highest accuracy, 80.26%, is achieved using rectangular grid features with a DHMM of 140 states and five training images per class.

Since the rectangular grid feature outperforms the polar grid, we choose the former to test the robustness of this approach to decreasing the number of training images from five to one per class. Table 2 shows that the classification accuracy decreases significantly from 80.26% with five training samples, to 59.47% with one training sample. We conclude therefore, that the DHMM based approach is not very robust to a reduction in training image number.

5.4. SVM experiments

We perform further experiments using an SVM without the DHMM modeling process, i.e., the rectangular and polar grid features are input directly to the SVM. We tested both the linear and RBF kernels, Table 3, and found no significant difference between them. (For the RBF kernel, the optimal parameter is determined by a grid search using cross validation). As before, the rectangular grid feature performed better. Table 4 shows the results obtained by varying the number of training images from five to one. From these results we can observe that the RBF kernel works slightly better than the linear kernel. As was the case for the DHMM, the SVM is not robust to a reduction in training image number. It is interesting to note that discriminative SVM works better than the generative DHMM.

5.5. DHMMK experiments

Tables 5 and 6 show the results of using the DHMM kernel under the same experimental settings used in Subsections 5.3 and 5.4. Table 5 shows the results of varying the number of states with five training images per class. Compared to the results in Table 1, we can see that the DHMMK significantly outperforms the DHMM for both rectangular and polar grid features. The rectangular grid feature again performs better than the polar grid. In particular, we achieve an accuracy of 100%

Table 6
DHMMK classification results for the rectangular grid feature with 140 states, different number of training images per class.

Features	# TS	SN	SVM — linear kernel success rate	SVM — RBF kernel success rate
Grid layout	5	140	**100% ± 0**	**100% ± 0**
Grid layout	4	140	100% ± 0	100% ± 0
Grid layout	3	140	99.95% ± 0.08	100% ± 0
Grid layout	2	140	99.83% ± 0.16	99.89% ± 0.16
Grid layout	1	140	83.78% ± 5.64	29.56% ± 16.99

The bold color indicates the best accuracy under the mean and standard deviation points of view. If the standard deviation is low means better stability of the mean value.

using the rectangular grid feature with five training samples and 140 states.

Using similar settings to the DHMM, we use the rectangular grid feature to test the robustness of the DHMMK against a reduction from five to one training images per class. Table 6 shows that the proposed approach can still achieve an accuracy of 99.83% using only two training images per class. This indicates that the DHMMK approach is less sensitive to the number of training images, than either DHMM or SVM approaches.

Overall, compared to the results in Subsections 5.3 and 5.4, the DHMMK approach outperforms both SVM and DHMM. It is also interesting to note that the proposed approach is fast when classifying a test sample. Using MATLAB on a computer with a 2.66-GHz CPU, and 2 GB RAM, the running speed of our approach is 400 milliseconds per sample, although the training time is longer than the running speed, being slightly over three hours.

5.6. Experiments with the PIE and RaFD face databases

In this section we test the performance of our approach using two independent public face datasets, PIE [43] and RaFD [44]. As stated before, the PIE dataset contains illumination changes with images of lower resolution, while the RaFD dataset contains images of different facial expressions but with a similar image resolution to the GPDS-ULPGC database. We have kept exactly the same experimental and parameter settings as in Subsections 5.3, 5.4, and 5.5 for the GPDS-ULPGC database. Results are summarized from Tables 7 to 12. From those results, we can draw the same conclusions as for the GPDS-ULPGC dataset. 1) Best results are obtained with the rectangular grid feature, and the DHMM based approach is not very robust to a reduction in the number of training images (Table 7). 2) The SVM works better than the DHMM. The performance gain of using a rectangular versus a polar grid with an SVM is much higher than the gain achieved with DHMM (Tables 9 and 10). This demonstrates that the SVM fully utilizes the discriminative potential of the rectangular grid feature. 3) The proposed DHMM kernel based approach works much better than either SVM or DHMM, and is less sensitive to the number of training images. With only two training

Table 5
DHMMK classification results for both the rectangular and polar grid features with different number of states, 5 training images per class.

Features	# TS	SN	SVM — linear kernel success rate	SVM — RBF kernel success rate
Polar layout	5	40	83.70% ± 1.05	84.00% ± 4.52
Polar layout	5	65	78.71% ± 5.05	78.43% ± 3.03
Polar layout	5	90	75.21% ± 4.58	75.59% ± 4.73
Polar layout	5	115	72.88% ± 4.13	73.24% ± 4.38
Polar layout	5	140	69.54% ± 5.82	70.04% ± 5.11
Grid layout	5	40	96.19% ± 1.71	97.08% ± 2.23
Grid layout	5	65	97.76% ± 1.22	98.24% ± 1.80
Grid layout	5	90	99.47% ± 1.12	99.33% ± 1.09
Grid layout	5	115	99.60% ± 0.80	99.60% ± 0.80
Grid layout	5	140	100% ± 0	100% ± 0

Table 7
DHMM classification results for both the rectangular and polar grid features with different number of states, 5 training images per class.

Features	State numbers	HMM success rate for PIE dataset	HMM success rate for RaFD dataset
Polar layout	40	27.53% ± 3.50	31.19% ± 7.51
Polar layout	65	26.63% ± 7.76	32.46% ± 7.87
Polar layout	90	24.84% ± 0.38	33.49% ± 8.55
Polar layout	115	22.88% ± 0.98	32.71% ± 4.21
Polar layout	140	23.45% ± 5.43	30.23% ± 6.92
Grid layout	40	25.82% ± 8.02	36.81% ± 6.90
Grid layout	65	32.84% ± 10.1	38.46% ± 7.42
Grid layout	90	36.19% ± 9.99	40.33% ± 7.57
Grid layout	115	33.82% ± 7.03	38.76% ± 6.06
Grid layout	140	33.17% ± 13.5	35.62% ± 8.31

Table 8
DHMM classification results for the rectangular grid feature with 140 states, different number of training images per class.

Features	# training samples	HMM success rate for PIE dataset	HMM success rate for RaFD dataset
Grid layout	5	33.17% ± 13.5	35.62% ± 8.31
Grid layout	4	30.49% ± 4.72	31.62% ± 5.86
Grid layout	3	26.52% ± 4.32	27.14% ± 4.29
Grid layout	2	25.92% ± 3.41	26.78% ± 5.91
Grid layout	1	17.43% ± 4.28	16.81% ± 4.57

Table 9
SVM classification results for both the rectangular and polar grid features with different kernels, 5 training images per class.

Features	Kernel	SVM success rate for PIE dataset	SVM success rate for RaFD dataset
Polar layout	Linear	25.35% ± 2.38	48.75% ± 2.91
Polar layout	RBF	39.02% ± 3.86	59.27% ± 2.83
Grid layout	Linear	56.01% ± 2.27	66.72% ± 2.13
Grid layout	RBF	70.24% ± 2.14	78.96% ± 1.98

Table 10
SVM classification results for rectangular grid feature with 140 states, different number of training images per class.

Features	# training samples	Kernel	SVM success rate for PIE dataset	SVM success rate for RaFD dataset
Grid layout	5	Linear	56.01% ± 2.27	66.72% ± 2.13
Grid layout	5	RBF	70.24% ± 2.14	78.96% ± 1.98
Grid layout	4	Linear	55.54% ± 2.27	64.81% ± 2.66
Grid layout	4	RBF	68.53% ± 2.35	75.26% ± 2.35
Grid layout	3	Linear	49.37% ± 4.70	60.32% ± 2.71
Grid layout	3	RBF	61.68% ± 3.72	71.12% ± 2.81
Grid layout	2	Linear	47.13% ± 2.05	66.72% ± 2.97
Grid layout	2	RBF	58.20% ± 2.89	57.03% ± 2.87
Grid layout	1	Linear	42.61% ± 2.51	66.72% ± 2.80
Grid layout	1	RBF	54.15% ± 2.01	51.84% ± 3.31

images per class, the DHMMK–SVM approach gets a classification accuracy of 97.13%.

The results in Table 7 show that when increasing the number of the states, the performance of the standard DHMM decreases (e.g., in the case of rectangular grid features, 36.19% and 40.33% with 90 states vs. 33.17% and 35.62% with 140 states for the PIE and RaFD datasets respectively). This is likely to be due to over fitting with the large number of states. However, the proposed approach with a DHMM kernel and SVM is much more robust to the change of state numbers for the PIE dataset (e.g., 99.43% with 90 states vs. 99.26% with 140 states). Lastly, similar to GPDS-ULPGC, the RaFD dataset is much more resistant to over fitting.

It is worth pointing out that the proposed approach is more robust to the change of scale and resolution than simply using HMM and SVM. This is demonstrated by cross-comparing the results achieved on the RaFD and PIE dataset from Tables 8 to 11. Note that one major difference between RaFD and PIE dataset is that the image size in the PIE dataset is much smaller (1/10) (size: 200×300) than those in the RaFD dataset (size: 1024×681). This scale and resolution change posed a significant challenge for the traditional method using HMM and SVM directly. For example, from Table 9, we can see that traditional approach using linear SVM and grid layout features with 5 training samples per class achieves the mean accuracy of 56.01% on the PIE dataset compared to the 66.72% on the RaFD dataset. The performance gap is up to 10% between those two datasets. While the proposed approach with an HMM kernel under the exactly the same experimental setting achieves the mean accuracy of 99.26% on the PIE dataset compared to 99.44% on the RaFD dataset, the performance gap (~0.2%) between the two datasets under the proposed approach is neglectable. This shows that the propose approach is robust to the change of scale and resolution to a large extent. Another point worth pointing out is that nowadays, the cameras equipped by the mobile phone are not of high quality, and it is not difficult to capture high resolution images that could well match the working conditions of our approach.

5.7. Receiver operating characteristic curves

Finally, we adopt a similar biometric experiment protocol to that used in [48] and evaluated the three approaches above for user authentication. Fig. 5 shows the receiver operating characteristic (ROC) curves for DHMMK–SVM, SVM and DHMM with the rectangular grid feature using two training images per class on the GPDS-ULPGC dataset. Figs. 6 and 7 show the corresponding curves for the PIE and RaFD databases respectively. It is quite clear that DHMMK significantly outperforms the SVM and DHMM approaches, especially on the challenging PIE and RaFD datasets. For the PIE dataset, the low resolution of the extracted lip contour tends to degrade the performance of all three approaches by differing degrees. As shown in the classification experiments, for the baseline approach using DHMM, the performance drops significantly from 65.20% on the GPDS-ULPGC dataset to 25.92% on the PIE dataset using the rectangular grid feature. For DHMMK, the performance is only slightly lower, dropping from 99.83% to 97.13%, with two training samples. As the number of sample is increased, this effect disappears, as verified from the ROC curves in Figs. 5 and 6. This confirms that the proposed DHMMK approach tends to be much more robust with respect to loss in resolution and illumination changes. The performance comparison between the RaFD and GPDS-ULPGC datasets reveals a similar trend, with the RaFD performance being slightly lower than that obtained with the GPDS-ULPGC dataset (98.10% vs. 99.83% using two training samples). When the number of training samples is increased, this difference is reduced, giving comparable performances for the two datasets (see Fig. 7).

Table 11
DHMMK classification results for the rectangular and polar grid features with different number of states, 5 training images per class.

Features	# TS	SN	SVM — linear kernel success rate for PIE dataset	SVM — RBF kernel success rate for PIE dataset	SVM — linear kernel success rate for RaFD dataset	SVM — RBF kernel success rate for RaFD dataset
Polar layout	5	40	97.38% ± 1.21	97.22% ± 1.10	97.56% ± 1.39	97.71% ± 1.56
Polar layout	5	65	95.59% ± 0.88	95.42% ± 0.79	97.82% ± 1.42	97.94% ± 1.31
Polar layout	5	90	95.75% ± 1.35	95.83% ± 0.98	98.08% ± 1.27	98.08% ± 1.27
Polar layout	5	115	97.14% ± 0.93	96.81% ± 0.88	98.29% ± 1.08	98.29% ± 1.08
Polar layout	5	140	96.49% ± 1.41	96.49% ± 1.41	98.25% ± 1.18	98.25% ± 1.18
Grid layout	5	40	99.18% ± 0.32	99.51% ± 0.22	98.62% ± 1.24	99.16% ± 0.71
Grid layout	5	65	98.94% ± 0.56	98.94% ± 0.56	99.14% ± 0.72	99.22% ± 0.46
Grid layout	5	90	99.43% ± 0.08	99.43% ± 0.08	99.38% ± 0.40	99.45% ± 0.38
Grid layout	5	115	98.69% ± 0.14	98.77% ± 0.18	99.41% ± 0.24	99.57% ± 0.32
Grid layout	5	140	99.26% ± 0.72	99.26% ± 0.72	99.44% ± 0.06	99.72% ± 0.23

Table 12
DHMMK classification results for the rectangular grid Layout feature with 140 states, different number of training images per class.

Features	# TS	SN	SVM — linear kernel success rate for PIE dataset	SVM — RBF kernel success rate for PIE dataset	SVM — linear kernel success rate for RaFD dataset	SVM — RBF kernel success rate for RaFD dataset
Grid layout	5	140	99.26% ± 0.72	99.26% ± 0.72	99.44% ± 0.06	99.72% ± 0.23
Grid layout	4	140	98.54% ± 0.58	99.05% ± 0.70	99.19% ± 0.40	99.49% ± 0.29
Grid layout	3	140	97.86% ± 0.59	98.03% ± 0.62	98.52% ± 0.64	98.79% ± 0.39
Grid layout	2	140	96.36% ± 0.41	97.13% ± 0.56	97.30% ± 0.81	98.10% ± 0.40
Grid layout	1	140	79.47% ± 1.84	80.57% ± 1.54	81.53% ± 1.90	83.17% ± 1.76

6. Discussions & conclusions

In this work we presented a novel and robust identification approach based on a DHMMK from static lip information. The main results are:

- The performance of the DHMMK is significantly superior to that of DHMM or SVM for all three datasets, proving that the use of static lip information is a good biometric modality for the identification.
- The DHMMK performance improvement is greater for the more challenging PIE and RaFD datasets, indicating a greater robustness to changes in illumination, resolution loss and variation in facial expression.
- The DHMMK requires fewer training images per subject.
- The rectangular grid feature was found to provide better performance than the polar grid feature.
- The proposed approach is less sensitive to HMM system parameters, such as the number of hidden states, *N*.
- The proposed approach is found less sensitive to scale and resolution changes.

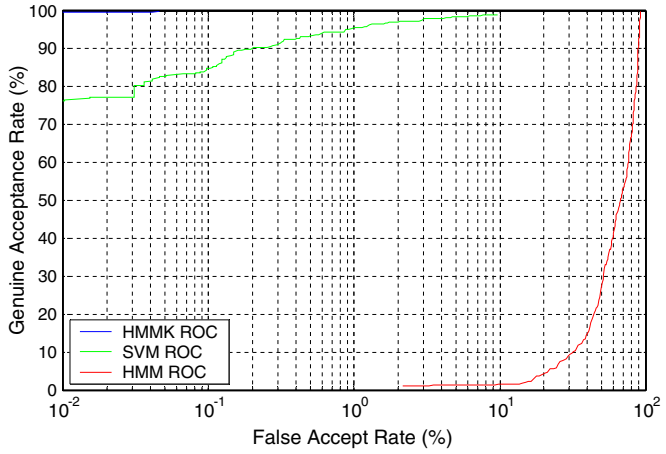

Fig. 5. ROC curves for the GPDS-ULPGC dataset using 2 training samples per class (better viewed in color).

Previously, using static lip features with an HMM has not been particularly promising with regard to subject authentication [14]. However, DHMMK overturns this perception and shows great promise for biometric based access control. This is due to the fact that the DHMMK captures valuable information from the gradient of the probability of having a certain feature vector in a particular state for both the rectangular and polar grid features. Furthermore, the fact that it works well while only requiring two training images is important with respect to subject registration — an important practical consideration for large scale access control.

Our lip extraction technique is robust to the change of skin colors to some extent. This was demonstrated by the lip contour extraction results on the PIE dataset illustrated in Fig. 3. The first image presents much darker skin color than the second image. The lip contour could be extracted relatively reliable in such cases. Much darker region such as a black face might affect the lip extraction results heavier. One possible idea to overcome this is to develop a possible face detector and adapt the lip extractor for black, dark, and Caucasian faces respectively.

It is worth pointing out that in all the three datasets tested, it is not necessary to conduct rotation correction since there are no rotated lips presented. To handle the moderate rotations of the lips, one possible way is to develop a simple algorithm in the pre-processing block to

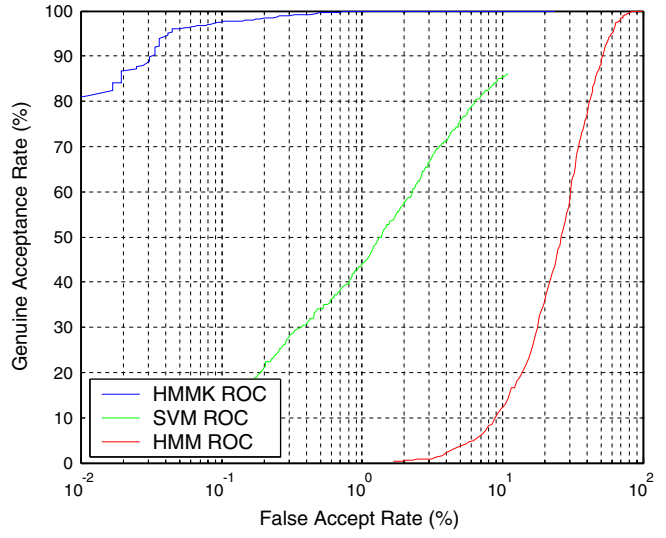

Fig. 6. ROC curves for the PIE dataset using 2 training samples per class (better viewed in color).

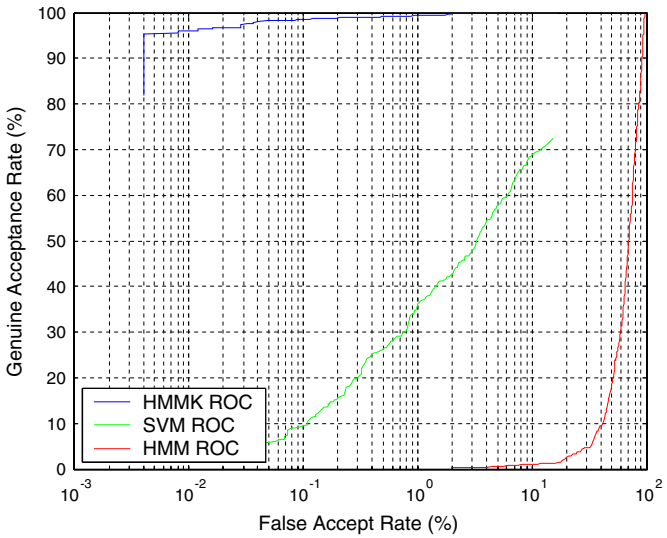

Fig. 7. ROC curves for the RaFD dataset using 2 training samples per class (better viewed in color).

detect the rotation of the lips. The final horizontal line can be detected as the principal axis and the slope of this line could be used for the rotation correction.

In future work, we propose to investigate the performance of DHMMK under different camera viewpoints, in more complex scenarios, and under complex illumination conditions. Finally, fusing static lip features with other biometric modalities such as the face [30], or lip dynamic features, to develop a robust multimodal system and tested on a larger dataset such as the FRGC [49] would be another interesting direction to investigate.

Acknowledgment

This work is partially supported by funds from The Royal Society of Edinburgh under RSE International Exchange Programme, 2012 to Carlos M. Travieso-González; and by the Spanish Government, under Grant MCINN TEC2012-38630-C04-02.

References

[1] Anil K. Jain, Sharath Pankanti, Salil Prabhakar, Lin Hong, Arun Ross, Biometrics: a grand challenge, ICPR (2004) 935–942.
[2] Acuity Market Intelligence, Available: http://www.securitydreamer.com/2007/06/new_clearheaded.html. September 30 2012.
[3] L. Bui, D. Tran, X. Huang, G. Chetty, Face gender recognition based on 2D principal component analysis and support vector machine, 4th International Conference on Network and System Security, 2010, pp. 579–582.
[4] N.A. Fox, R. Gross, J.F. Cohn, R.B. Reilly, Robust biometric person identification using automatic classifier fusion of speech, mouth and face experts, IEEE Trans. Multimed. 9 (2007) 701–714.
[5] A.K. Jain, P. Flynn, A.A. Ross, Handbook of Biometrics, Springer, 2008. (Ed.).
[6] S.L. Wang, A.W.C. Liew, Physiological and behavioral lip biometrics: a comprehensive study of their discriminative power, Pattern Recogn. 45 (2012) 3328–3335.
[7] T. Wark, S. Sridharan, V. Chandran, An approach to statistical lip modelling for speaker identification via chromatic feature extraction, IEEE International Conference on, Pattern Recognition, 1998, pp. 123–125.
[8] T.J. Wark, S. Sridharan, V. Chandran, The use of speech and lip modalities for robust speaker verification under adverse conditions, Proc. Int. Conf. Multimedia Computing and Systems, 1997, pp. 812–816.
[9] H. Çetingül, Y. Yemez, E. Erzin, A. Tekalp, Discriminative analysis of lip motion features for speaker identification and speech-reading, IEEE Trans. Image Process. 15 (10) (2006) 2879–2891.
[10] J.L. Newman, S.J. Cox, Automatic visual-only language identification: a preliminary study, IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4345–4348.
[11] J.L. Newman, S.J. Cox, Speaker independent visual-only language identification, IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 5026–5029.
[12] J. Lal-Raheja, R. Shyam, J. Gupta, U. Kumar, P.B. Prasad, Facial gesture identification using lip contours, Second International Conference on Machine Learning and Computing (ICMLC), 2010, pp. 3–7.
[13] A. Mehra, M. Kumawat, R. Ranjan, B. Pandey, S. Ranjan, A. Shukla, R. Tiwari, Expert system for speaker identification using lip features with PCA, Second International Workshop on Intelligent Systems and Applications (ISA), 2010, pp. 1–4.
[14] C.H. Chan, B. Goswami, J. Kittler, W. Christmas, Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication, IEEE Trans. Inf. Forensics Secur. 7 (2) (2012) 602–612.
[15] A. De la Cuesta, J. Zhang, P. Miller, Biometric identification using motion history images of a speaker's lip movements", Machine Vision and Image Processing International Conference, 2008, pp. 83–88.
[16] M.I. Faraj, J. Bigun, Motion Features from Lip Movement for Person Authentication, 18th International Conference on Pattern Recognition, vol. 3, 2006, pp. 1059–1062.
[17] R.A. Rao, R. Mersereau, Lip modeling for visual speech recognition, Conference Record of the Twenty-Eighth Asilomar conference on Signals, Systems and Computers, vol. 1, 1994, pp. 587–590.
[18] T. Coianiz, L. Torresan, D. Massaro, 2D Deformable Models for Visual Speech Analysis, NATO Advanced in Speechreading by Humans and Machines, Springer-Verlag, 1996. 391–398.
[19] R. Steifelhagen, J. Yang, U. Meier, Real time lip tracking for lip reading, Proceedings of Eurospeech '97, 1997.
[20] L. Yaling, D. Minghui, Lip contour extraction based on manifold, International Conference on MultiMedia and Information Technology, 2008, pp. 229–232.
[21] T.W. Lewis, D.M.W. Powers, Lip Feature Extraction Using Red Exclusion, ACM International Conference Proceeding Series, vol. 9, 2000, pp. 61–67.
[22] R. Rohani, S. Alizadeh, F. Sobhanmanesh, R. Boostani, Lip segmentation in color images, International Conference on Innovations in Information Technology, 2008, pp. 747–750.
[23] L. Dong, S.W. Foo, Y. Lian, A two-channel training algorithm for hidden Markov model and its application to lip reading, EURASIP J. Appl. Signal Process. 9 (2005) 1382–1399.
[24] S. Alizadeh, R. Boostani, V. Asadpour, Lip feature extraction and reduction for HMM-based visual speech recognition systems, 9th International Conference on, Signal Processing, 2008, pp. 561–564.
[25] S.W. Chin, L.M. Ang, K.P. Seng, Lips detection for audio–visual speech recognition system, International Symposium on Intelligent Signal Processing and Communications Systems, 2008, pp. 1–4.
[26] X. Yan, S. Guangda, Multi-parts and multi-feature fusion in face verification, Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–6.
[27] A. Salazar, G. Daza, S.L. Sánchez, F. Prieto, G. Castellanos, C. Quintero, Feature extraction & lips posture detection oriented to the treatment of CLP children, Proceedings of the 28th IEEE EMBS Annual International Conference, 2006, pp. 5747–5750.
[28] G. Chetty, M. Wagner, Audio visual speaker verification based on hybrid fusion of cross modal features, Proc. Pattern Recognition and Machine Intelligence (PReMI), 2007, pp. 469–478.
[29] M.A. Ferrer, J.B. Alonso, Carlos M. Travieso, Offline geometric parameters for automatic signature verification using fixed-point arithmetic, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 993–997.
[30] D. Yujie, D.L. Woodard, Eyebrow shape-based features for biometric recognition and gender classification: a feasibility study, International Joint Conference on Biometrics (IJCB), 2011, pp. 1–8.
[31] L. Zhe, L.S. Davis, Shape-based human detection and segmentation via hierarchical part-template matching, IEEE Trans. Pattern Anal. Mach. Intell. 32 (4) (2010) 604–618.
[32] C.M. Travieso, J. Zhang, P. Miller, J.B. Alonso, M.A. Ferrer, Bimodal biometric verification based on face and lips, Neurocomputing 74 (14–15) (2011) 2407–2410.
[33] P.S. Aleksic, A.K. Katsaggelos, Audio–visual biometrics, Proc. IEEE 94 (11) (2006) 2025–2044.
[34] P. Viola, M. Jones, Robust real-time object detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.
[35] J.A. Dargham, A. Chekima, Lips detection in the normalised RGB colour scheme, International Conference on Information and Communication Technologies, 2006, pp. 1546–1551.
[36] N. Otsu, A thresholding selection method from gray-level histogram, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[37] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[38] J. Cervantes, X. Li, W. Yu, K. Li, Support vector machine classification for large data sets via minimum enclosing ball clustering, Neurocomputing 71 (4–6) (2008) 611–619.
[39] J. Hernando, C. Nadeu, J.B. Mariño, Speech recognition in a noisy environment based on LP of the one-sided autocorrelation sequence and robust similarity measuring techniques, Speech Commun. 21 (1997) 17–31.
[40] L. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP Mag. 3 (1) (2003) 4–16.
[41] S. David, M.A. Ferrer, C.M. Travieso, J.B. Alonso, gpdsHMM: a hidden Markov model toolbox in the Matlab environment, International Conference Complex System Intelligence and Modern Technology Application, 2004.
[42] T. Jaakkola, M. Diekhans, D. Haussler, A discriminative framework for detecting remote protein homologies, J. Comput. Biol. 7 (1–2) (2000) 95–114.
[43] PIE Face Database, Available: http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261. September 12 2012.
[44] O. Langner, R. Dotsch, G. Bijlstra, D.H.J. Wigboldus, S.T. Hawk, A. van Knippenberg, Presentation and validation of the Radboud Faces Database, Cogn. Emot. 24 (2010) 1377–1388.
[45] T. Joachims, SVM_light Support Vector Machine, Available: http://svmlight.joachims.org/. September 2 2012.
[46] K. Tselios, E.N. Zois, E. Siores, A. Nassiopoulos, G. Economou, Grid-based feature distributions for off-line signature verification, IET Biom. 1 (1) (2012) 72–81.
[47] U.M. Braga-Neto, Edward R. Dougherty, Is cross-validation valid for small-sample microarray classification? Bioinformatics 20 (3) (2004) 374–380.
[48] A. Kumar, D. Zhang, Personal recognition using hand shape and texture, IEEE Trans. Image Process. 15 (8) (2006) 2454–2461.
[49] FRGC dataset: Face Recognition Grand Challenge, Available: http://www.nist.gov/itl/iad/ig/frgc.cfm. July 17 2013.

Carlos M. Travieso-González received the M.Sc. degree in 1997 in Telecommunication Engineering at Polytechnic University of Catalonia (UPC), Spain; and Ph.D. degree in 2002 at University of Las Palmas de Gran Canaria (ULPGC-Spain). He is an Associate Professor in ULPGC, teaching subjects on signal processing and learning theory from 2001. His research lines are biometrics, biomedical signals, data mining, classification system, signal and image processing, and environmental intelligence. He has researched in International and Spanish Research Projects, some of them as a head researcher. He is a co-author of books, and book chapters, a co-editor of Proceedings Books, and a Guest Editor for four international journals under IF-JCR. He has many papers published in international journals, conferences and Springer Series. He has been a reviewer in different international IF-JCR journals and top conferences since 2001. He is a member of IASTED Technical Committee on Image Processing and a member of IASTED Technical Committee on Artificial Intelligence and Expert Systems. He will be an InnoEducaTIC 2014 General Chair and IEEE-IWOBI 2015 General Chair, and was an IEEE-IWOBI 2014 General Chair, IEEE-INES 2013 General Chair, NoLISP 2011 General Chair, JRBP 2012 General Chair and Co-Chair on 39th Annual 2005 IEEE-ICCST. He has been a PC Member up to 50 international conferences. He is the Vice-Dean since March 2013 in Higher Technical School of Telecommunication and Electronic Engineers and was the Vice-Dean from 2004 to 2010 in Higher Technical School of Telecommunication Engineers in ULPGC.

Jianguo Zhang (BSc MSc DPhil) is a Senior Lecturer of Visual Computation in School of Computing, University of Dundee. Previously, he worked as a lecturer in Electronics, Electrical Engineering and Computer Science at Queen's University Belfast, UK (2007–2010),

and a researcher in the Department of Computer Science at Queen Mary University of London (2005–2007), with the Lear group of INRIA Rhône-Alpes in France (2003–2005), and in the School of Electrical and Electronic Engineering at Nanyang Technological University of Singapore (2002–2003). He received his DPhil in National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2002. He twice won the International Pascal Visual Object Classification Challenge for the categorization contest in 2005 and 2006. He was awarded the Present Prize of Chinese Academy of Sciences 2002. He is the founding co-chair and co-organizers of the International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR2009, 2010, 2011, 2012, and 2013). He is the guest editor of Pattern Recognition Journal 2012, and an editor of a book "Intelligent Video Event Analysis and Understanding" in Springer-Verlag series. He is an area chair of BMVC 20142011. He regularly served as PC members/referees for many international journals and top conferences, including IJCV, IEEE TPAMI, IEEE TIP, IEEE TSMC, IVC, PR, CVIU, PRL, ICCV, CVPR, and ECCV. He is a Senior Member of IEEE (SMIEEE) and a fellow of High Education of Academy United Kingdom (FHEA).

Jesús B. Alonso-Hernández received the Telecommunication Engineer degree in 2001 and the Ph.D. degree in 2006 from University of Las Palmas de Gran Canaria (ULPGC-Spain) where he is an Associate Professor in the Department of Signal and Communications from 2002. He has researched in different International and Spanish Research Projects. He has numerous papers published in international journals and international conferences. He has been a reviewer in different international journals and conferences since 2003. His research interests include signal processing in biocomputing, biometrics, nonlinear signal processing, recognition systems, audio characterization and data mining. He was a Guest Editor of Special Issues for Elsevier in Cognitive Computation and Neurocomputing. He was the head of excellent network in biomedical engineering in ULPGC. He is the Vice-Dean from 2009 in School of Telecommunications Engineering in ULPGC.

Dr. Paul Miller is a Senior Lecturer in the School of Electronics, Electrical Engineering and Computer Science at Queen's University Belfast (QUB). He is also a Research Director of the Intelligent Surveillance Systems group in the Centre for Secure Information Technology. He has published over sixty papers in image and video analysis, including a best paper award for his work on object recognition. Previously, he worked as a senior research scientist at the Defence, Science and Technology Organisation, Australia where he led a team providing science and technology advice on unmanned aerial surveillance systems. Before that he worked as a research fellow at QUB. He received his PhD in Optical Image Processing from QUB in 1989. Since returning to academia he has continued to work in video analytics for both defence and civilian CCTV applications, and also bio-medical image analysis. His research interests include image restoration, segmentation, multi-camera tracking and gender/age profiling of subjects in video. During his academic career he has constantly worked in close collaboration with industry, including both multinational and SME companies.