# Facial gender recognition using multiple sources of visual information

Federico MATTA, Usman SAEED, Caroline MALLAURAN and Jean-Luc DUGELAY

*Multimedia Communications Department, Eurécom*
*2229 route des Cretes, 06904, Sophia Antipolis cedex, France*
{Federico.Matta, Usman.Saeed, Caroline.Mallauran, Jean-Luc.Dugelay}@eurecom.fr

*Abstract*—**In this article we present a novel multimodal gender recognition system, which successfully integrates the head and mouth motion information with facial appearance by taking advantage of a unified probabilistic framework. In fact, we develop a temporal subsystem that has an extended feature space consisting of parameters related to head and mouth motion; at the same time, we introduce a complementary spatial subsystem based on a probabilistic extension of the eigenface approach. In the end, we implement an integration step to combine the similarity scores of the two parallel subsystems, using a suitable opinion fusion (or score fusion) strategy. The experiments show that not only facial appearance but also head and mouth motion possess a potentially relevant discriminatory power, and that the integration of different sources of biometric information from video sequences is the key strategy to develop more accurate and reliable recognition systems.**

## I. INTRODUCTION

Human face contains a variety of information for adaptive social interactions amongst people. In fact, individuals are able to process a face in a variety of ways to categorize it by its identity, along with a number of other demographic characteristics, such as gender, ethnicity, and age. In particular, recognizing human gender is important since people respond differently according to gender. In addition, a successful gender classification approach can boost the performance of many other applications, including person recognition and smart human-computer interfaces.

In this article, we address the problem of automatic gender recognition by exploiting the physiological and behavioural aspects of the face at the same time. We have already investigated the use of the head and mouth motion information for person recognition in an earlier research study [1]. Currently, comforted by the promising results obtained by this previous approach, we explore the possibility of using head motion, mouth motion and facial appearance in a gender recognition scenario. Hence, we propose a multimodal recognition approach that integrates the temporal and spatial information of the face through a probabilistic framework.

The remainder of this article is organised as follows: in section II we propose a short review of related works, and then in section III we detail our recognition system; afterwards we report and comment the experiments in section IV, and finally we conclude this paper with remarks and future work in section V.

## II. RELATED WORKS

There exists a vast literature in social and cognitive psychology describing the impressive capabilities of humans at identifying familiar faces; though, most works deal with person recognition, and only few studies are focused on gender recognition. The automatic gender recognition from human faces has been studied since the late '80s [2][3][4][5][6], but only in the new millennium it has received significant attention from the scientific community: here we propose a short review of the latest approaches for gender recognition.

Sun et al. [7] applied principal component analysis (PCA) to represent each image as a feature vector in a low dimensional space; genetic algorithms (GA) were then employed to select a subset of features form the low dimensional representation that mostly encodes the gender information. Four different classifiers were compared in this study: the Bayesian decision making, a neural network (NN), support vector machines (SVM) and a classifier based on linear discriminant analysis (LDA). The SVM achieved the best performance in the comparative experiments.

Gutta et al. [8] considered a hybrid classifier for gender determination of human faces that consisted of an ensemble of radial basis functions (RBFs) and decision trees (DTs).

Nakano et al. [9] focused on the edge information and exploited a neural network (NN) classifier for gender recognition. In particular, they computed the density histograms of the edge images, which were successively treated as input features for the NN.

Lu et al. [10] exploited the range and intensity information of human faces for ethnicity and gender identification using a support vector machine (SVM). They firstly dealt with the ethnicity discrimination between Asian and non-Asian, where they exploited the relationship between the 3D shape of the human face and the ethnicity. The 3D scans were firstly registered by manually specifying six landmark points, and then they applied a grid to crop the face area and generate the feature vectors. As a similarity measure for classification they computed the posterior probabilities from the SVMs; in the end, they identified the gender of the user by comparing the probability to be a man with that of being a woman.

Moghaddam et al. [11] also proposed to classify gender from facial images (of 21x21 pixels) using support vector machines (SVMs). They tested the SVMs by implementing different kernels and they obtained the best experimental results with the Gaussian kernel, followed by the cubic polynomial kernel.

Saatci and Town [12] have proposed a combined gender and expression recognition system using facial images. First the face was modelled using an Active Appearance Model (AAM), then features were extracted. Linear, polynomial and RBF based SVM were then used to classify gender and expression. Furthermore gender specific differences in appearance were exploited for expression recognition and vice versa.

Kim et al. [13] base their gender recognition system on a Gaussian Process Classifier (GPC). Facial images are first normalized to a standard dimensions and background and hair information was removed. Parameters for the GPC are learned using Expectation Maximization (EM) - Expectation Propagation (EP) algorithm. Finally GPC is used for classification.

Zhiguang et al. [14] have focused on improving gender classification results by texture normalization. After scale normalization, affine fitting and Delaunay triangulation warping is employed to get a shape free texture. Lastly SVM, FLD and Adaboost are used for classification.

Rowly and Baluja [15] have proposed using adaboost using low resolution greyscale images. Several weak classifiers were build based on pixel value comparisons which lead to results just better than random chance. Then these are combined using adaboost to create a strong classifier. Tests were carried out on Feret database with an overall accuracy of 90 %.

Caifeng et al. [16] have fused gait and face features for gender classification. Gait Energy Images were used for gait feature representation and normalized pixel values were used facial feature representation. These were then fused at feature level using canonical correlation analysis. Lastly support vector machines were used for classification.

## III. Proposed Method

Our system is composed by two parallel complementary subsystems and a score integration step. The first recognition subsystem (identified by light yellow boxes in Fig. 1) is exploiting the temporal video information and is based on unconstrained head and mouth motion. The second recognition subsystem (identified by tan boxes in Fig. 1) works with the spatial information and exploits facial appearance; more precisely, it is a probabilistic extension of the original eigenface technique presented in [17] by Turk and Pentland. For a consistent integration of this heterogeneous biometric information (motion and appearance) into a unified recognition approach, both subsystems share the same probabilistic framework: a Gaussian mixture model (GMMs) approximation to represent the biometric features of each client, and Bayesian inference to calculate the similarity between tests and models. In the end, the similarity scores of the two parallel subsystems are combined in the last step (identified by a gold box in Fig. 1), which operates the final identification and verification decisions after a suitable opinion fusion (or score fusion).

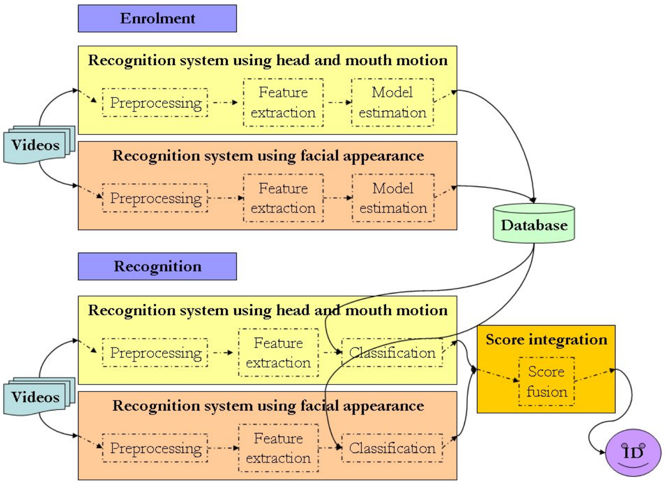

Fig. 1. Architecture of the multimodal recognition system.

The two subsystems and the integration step are detailed in the following three subsections.

### A. Temporal recognition subsystem exploiting head and mouth motion

This subsystem exploits the unconstrained head and mouth motion information for gender recognition, by tracking a few facial landmarks in the image plane and by segmenting the outer lip contour in each video frame. Here we provide a short summary of this temporal subsystem: more details can be found in a previous article [1] (in section III).

#### 1) Pre-processing

Each video sequence is firstly pre-processed by semi-automatically detecting the face, and then by automatically tracking the facial landmarks over time using a template matching strategy in the RGB colour space. Next, the lip segmentation algorithm applies a series of image processing techniques, to locate the outer lip contour in every video frame, and several additional operators, to isolate and enhance the shape of the resulting contour (mainly morphological operators and convex hull approximations).

#### 2) Feature extraction

The feature extraction step is divided in two phases: the geometrical normalisation of the tracking signals, and the calculation of the feature vectors. The former part centres and scales the tracking signals; this way, after clearing the features from any dependence on absolute head location and size, the head motion information is isolated and the inter-video variation is reduced. Then the feature matrix, $\mathbf{X} \in \Re^{D\times N}$, which retains the whole head and mouth discriminative information extracted from the corresponding video sequence, is generated by concatenating the following distinct features:

- Head positions: the location of the head over time is included using the normalised tracking signals.
- Centred major axis of the outer lip contour.
- Centred minor axis of the outer lip contour.

*3) Model estimation*

The model estimation step approximates the class conditional probability density functions (PDFs) of each client, by using Gaussian mixture models (GMMs):

$$p(\mathbf{x} \mid \mathbf{\Theta}) \equiv \sum_{c=1}^{C} \alpha_c \aleph(\mathbf{x} \mid \boldsymbol{\mu}_c, \mathbf{\Sigma}_c)$$

where $\aleph(\mathbf{x} \mid \boldsymbol{\mu}_c, \mathbf{\Sigma}_c)$ is a singular multivariate normal distribution of a random variable, $\mathbf{x} \in \Re^D$, $\mathbf{\Theta} = \{\alpha_c, \boldsymbol{\mu}_c, \mathbf{\Sigma}_c \mid c = 1, \ldots, C\}$ is the parameter list, and $\alpha_c \in [0,1]$ is the weight of the $c$-th Gaussian component. In addition, each $\alpha_c$ corresponds to the a priori probability that an observation $\mathbf{x}$ has been generated by the $c$-th normal source, and its value is normalised such as: $\sum_{c=1}^{C} \alpha_c \equiv 1$. Then, for each client $k$, his/her model parameters $\mathbf{\Theta}_k$ are obtained by solving a maximum likelihood problem through the expectation-maximisation (EM) algorithm.

*4) Classification*

Finally, the classification step computes the similarity scores of the temporal subsystem by applying the probability theory and the Bayesian decision rule (also called Bayesian inference). The similarity score for the identification task is the video log-posterior probability:

$$S^{(ID)}(\mathbf{X}, \mathbf{\Theta}_k) = \ln p(k \mid \mathbf{X})$$

while the similarity score for the verification mode is the video log-posterior probability ratio:

$$S^{(VER)}(\mathbf{X}, \mathbf{\Theta}_k) = \ln \left[ \frac{p(k \mid \mathbf{X})}{p(\bar{k} \mid \mathbf{X})} \right]$$

where $p(\bar{k} \mid \mathbf{X})$ is the posterior probability of the alternative hypothesis $\bar{k}$, and $p(\mathbf{x}_n \mid \bar{k})$ is the impostor model (the class conditional PDF for $\bar{k}$).

### B. Spatial recognition subsystem exploiting facial appearance

For a consistent integration of the facial appearance information in our multimodal person recognition system, we developed a probabilistic extension of the original eigenface technique [17]. In particular, the pre-processing and feature extraction steps are kept pretty close to the standard eigenface approach, while the model estimation and classification steps are adapted to share the same probabilistic framework of the other recognition subsystem that exploits head and mouth motion.

*1) Pre-processing*

The pre-processing step applies some image processing techniques to a set of colour pictures belonging to a video sequence. The first transformation is a histogram equalisation colour component by colour component, which is useful to reduce the impact of inter-image illumination and colour variations. Then, this step converts the image signal into the most discriminative representation, which in our case is the NTSC (luminance, hue and saturation) colour space. Finally, the image pixels are arranged in long vectors through a process called image vectorisation.

*2) Feature extraction*

The feature extraction step isolates the discriminative information that characterises the individual and discards the irrelevant one, by applying a linear transformation from the high dimensional image space to a lower dimensional space (called the face space), which is much smaller. More precisely, each vectorised image $\mathbf{s}_n \in \mathrm{N}^{3RC}$ is approximated with its projection in the face space $\mathbf{v}_n \in \Re^D$ by the following linear transformation:

$$\mathbf{v}_n = \mathbf{W}^T(\mathbf{s}_n - \boldsymbol{\mu})$$

where $\mathbf{W} \in \Re^{3RC \times D}$ is a projection matrix with orthonormal columns, and $\boldsymbol{\mu} \in \Re^D$ is the mean image vector of the whole training set.

The optimal projection matrix $\mathbf{W}$ is computed using the principal component analysis (PCA) (also called the Karhunen-Loeve transform (KLT)), which has the property of optimally representing the distribution of data in the root mean squares sense; the details on the calculation of $\mathbf{W}$ can be found in [17].

Once the image data set is projected into the face space, the vectors in the feature matrix, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \Re^{D \times N}$, are generated by computing the whitened projections in face space.

*3) Model estimation*

The model estimation step adopts the same probabilistic approach of the parallel subsystem using head and mouth motion for recognition. In fact, the distribution of the feature vectors of each client is modelled with a GMM, which approximates the class conditional probability density function of each user, $k$, in feature space:

$$p(\mathbf{x}_n \mid k) \cong p(\mathbf{x}_n \mid \mathbf{\Theta}_k) \equiv \sum_{c=1}^{C_k} \alpha_{k,c} \aleph(\mathbf{x}_n \mid \boldsymbol{\mu}_{k,c}, \mathbf{\Sigma}_{k,c})$$

*4) Classification*

Finally, the classification step also closely resembles to the one in the temporal recognition system; in fact, it also computes the similarity scores by applying the probability theory and the Bayesian decision rule (also called Bayesian inference). In our implementation, we select only one key frame to test a given video sequence; hence, the related feature matrix contains only one feature vector, $\mathbf{X} = \mathbf{x} \in \Re^D$, and the video posterior probability is equal to the frame posterior probability.

As before, the similarity score for the identification task is the video log-posterior probability:

$$S^{(ID)}(\mathbf{x}, \mathbf{\Theta}_k) = \ln p(k \mid \mathbf{x})$$

while the similarity score for the verification mode is the video log-posterior probability ratio:

$$S^{(VER)}(\mathbf{x}, \mathbf{\Theta}_k) = \ln\left[\frac{p(k \mid \mathbf{x})}{p(\bar{k} \mid \mathbf{x})}\right]$$

### *C. Score integration step*

The score integration step combines the similarity scores of the two parallel subsystems by applying a suitable opinion fusion (or score fusion) strategy; after that, it takes the final identification and verification decisions using this extended measure of similarity.

The score integration step calculates the multimodal similarity scores between the $j$-th test sequence $\Phi_j$, and the $k$-th client model by applying the weighted summation fusion (also called sum rule), which has the following general formula:

$$\xi_{j,k}^{(i)} \equiv S^{(i)}(\Phi_j, \mathbf{\Theta}_k) = a_{j,k}\eta_{j,k}^{(i)} + b_{j,k}\rho_{j,k}^{(i)}$$

where $a_{j,k}$ and $b_{j,k}$ are the weighting values, $\eta_{j,k}^{(i)}$ and $\rho_{j,k}^{(i)}$ are the similarity scores of the temporal and spatial subsystems, and $i$ specifies the identification or verification case. In particular, we selected the equal weighting of modalities, which is obtained by taking the average of the separate similarity scores, or equivalently by setting the weights as:

$$a_{j,k} = b_{j,k} = 0.5$$

for $\forall j, k$. This choice has an interesting probabilistic interpretation; in fact, if we assume that the features related to facial motion $\mathbf{X}_j^{(mtn)}$ and those to facial appearance $\mathbf{x}_j^{(app)}$ are statistically independent, then the multimodal similarity scores for the identification task are equal to the joint log-posterior probabilities of $\mathbf{X}_j^{(mtn)}$ and $\mathbf{x}_j^{(app)}$:

$$\xi_{j,k}^{(ID)} \equiv S^{(ID)}(\Phi_j, \mathbf{\Theta}_k) = \frac{1}{2}\log p\left(k \mid \mathbf{X}_j^{(mtn)}, \mathbf{x}_j^{(app)}\right) + \frac{1}{2} p(k)$$

except for an irrelevant translating factor, the a priori probability $p(k)$, which is not dependent on the test itself and it is already known before the recognition process.

## IV. Experiments and results

### *A. Video database*

Unfortunately, the existing standard video databases (like XM2VTS, Valid or My IDea) do not match the requirements for efficiently testing the proposed system. In fact, even if they are generally intended for multimodal approaches, they do not consider the head and facial motion information, and they do not present observable and characteristic movements; moreover, they do not contain enough video recordings per user to allow the exploitation of the behavioural information of the user.

#### *1) Our video database of Italian TV speakers*

For these reasons, we have been collecting a set of 208 video sequences belonging to 13 different persons, and we have trained and tested our system with this data. The video chunks present TV speakers announcing the news of the day: their motion is natural and no capricious events are occurring, like a scene change, a discussion with a guest, etc. A typical sequence has a spatial resolution of 352x288 pixels and a temporal resolution of 23.97 frames per second, and lasts 13 seconds. The videos are of low quality, acquired using a fixed camera and compressed at 118 Kbits per second, and they have been collected during a period of 21 months. It is important to notice that: there is no camera motion, because the camera is fixed, there are no zooms or changes in scale, and that the depth variation due to the in-depth movement of the speaker is insignificant, because the camera is far. Hence, all motion that can be extracted from these sequences is relative to the behaviour of the announcers. Fig. 2 illustrates our data set by showing the first 7 frames of 4 TV speakers.

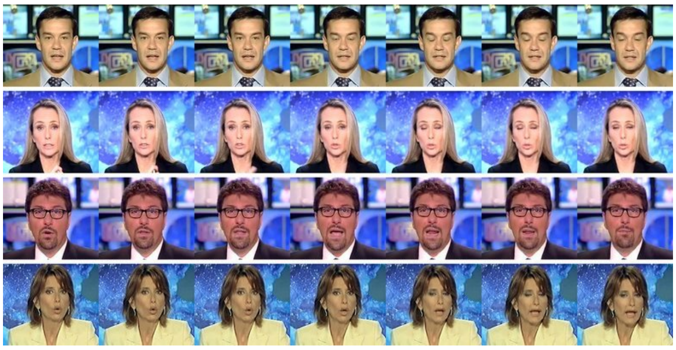

Fig. 2. Illustration of our video database with the first 7 frames of 4 TV speakers.

#### *2) Image version of the database*

Due to the well known high sensitivity of PCA-based recognition algorithms to facial alignment, variation in pose and scale, we derived a special version of the video database of Italian TV speakers by sub sampling and manually normalising some video frames. More precisely, for the enrolment subset we extracted 28 frames from each sequence, at a frame rate of 2 frames per second, whereas for the recognition subset we retrieved only the first key frame. After that, to normalise the video frames we firstly (in-plane) rotated the heads to horizontal eye position, then we cropped the face regions, and finally we aligned the images using the locations of the pupils.

### 3) *Enrolment and recognition sets*

For the evaluation of our recognition system, we split the whole database (both the video and image versions) into an enrolment subset and a recognition subset: 104 video sequences (8 for each of the 13 clients) are employed for the training of our system (enrolment), and the remaining 104 (8 for each of the 13 clients) are used for its testing (recognition).

## *B. Experimental set-up*

### 1) *Optimal configuration of the temporal subsystem*

In the optimal configuration, the head motion of each individual is represented through 8 tracking signals of 4 facial landmarks: the two eyes, the nose and the mouth. To improve the robustness of the tracking and reduce the intra-video variation, all frames are pre-processed using a histogram equalisation, colour component by colour component. During the head tracking step, the algorithm generates a starting template of 19 pixel rows and 25 pixel columns for each landmark; then, the similarity scores of each colour component are based on the city-block distance measure. After that, the feature extraction consists of centring the tracking signals, by applying a zero mean transformation, and using the normalised head positions and the centred major and minor axes of the outer lip contour as features for recognition; in the end, the dimensionality of the feature space is 10. Then, the client models are approximated using GMMs with 4 Gaussian components, and their parameters are estimated through the EM algorithm, which is initialised with: cluster means (computed using K-means), uniform weights and covariances. Finally, the impostor models for verification are approximated by taking the average of the best 2 background (or cohort) models.

### 2) *Optimal configuration of the spatial subsystem*

In the optimal configuration of the subsystem using facial appearance, all images are firstly pre-processed with a histogram equalisation, colour component by colour component, to reduce the mismatches due to illumination variations. Next, the data set is represented by using the NTSC colour space, because it empirically provides more discriminative signals than the RGB does. Due to the well known problem of approximating high dimensional distributions with a limited amount of data, we are obliged to adopt serious restrictions on the dimension of the face space and the number of Gaussian components for a reliable GMM parameter estimation. In fact, with 228 images per person in the enrolment subset, we should use an eigenspace of dimension 10 or less for being able to reliably train 2 components, and 8 or less for 3, which is excessively constraining because too much discriminative information is lost with such a reduced space. Hence, the client models are estimated using a single Gaussian component (which reverts on using a multivariate normal distribution), in a small face space of dimension 27, and the feature vectors are calculated by whitening the projection coefficients. Finally, the impostor models for verification are approximated by taking the average of the best 2 background (or cohort) models.

### 3) *Implementation of the eigenface approach for comparison*

In our implementation of the original eigenface approach [17], we firstly pre-process all images with a histogram equalisation, colour component by colour component, to reduce the mismatches due to illumination variations. Next, we represent the data set by using the NTSC colour space, because it empirically provides more discriminative signals than the RGB does. Once the colour components are rearranged into large vectors, we apply the PCA to the enrolment subset to compute a reduced face space of dimension 243, and we calculate the feature vectors by whitening the projection coefficients in the eigenspace. Then, the client models are registered into the system using their centroid vectors, which are calculated by taking the average of the feature vectors in the enrolment subset; in the end, recognition is achieved using a nearest neighbour classifier with cosine distances in (the whitened) face space.

## *C. Recognition results*

We express the identification results by reporting the correct identification rates (CIRs), and by plotting the cumulative (correct) match scores (CMSs) as a function of the $M$ best matches retained. For the verification scenario, we report the equal error rates (EERs) and we show the receiver operating characteristic (ROC) curves, which offer a global description of the system from low to high security applications.

The experimental results show that the integration of multiple sources of biometric information clearly improves the performance of the individual modalities. In fact, a biometric system exploiting only head motion obtains poor gender recognition scores with a CIR of 84.6% and an EER of 15.4%, we observe that the addition of mouth motion and then of facial appearance in our multi-biometric system increases the CIR to 96.2% and 99.0%, and decreases the EER to 3.8% and 1.0% respectively.

We also compared our gender recognition approach with the eigenface technique; our system performs in between the perfect recognition scores (100.0% of CIR and 0.0% of EER) of eigenfaces when applied to perfectly normalised images, which is an unrealistic and too favourable condition, and the poor results of eigenfaces when applied to not normalised frames: a CIR of 89.4% and an EER of 10.6%.

By looking at Fig. 3 we can visually evaluate the benefit of integrating multiple sources of biometric information. These graphs show that, even if head motion (aqua curves, “HM”) is not such an important discriminative identifier for gender recognition applications, it can still achieve excellent results if supported by the mouth motion information (violet curves, “HM+MM”), and particularly by the facial appearance one (blue curves, “HM+MM+FA”).

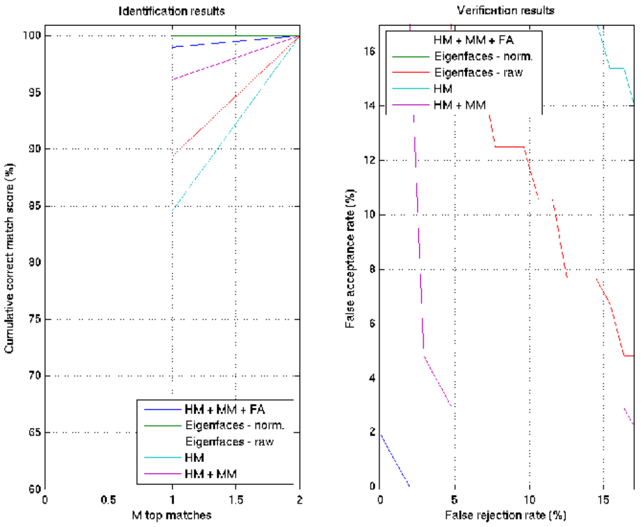

Fig. 3. Comparison of gender recognition results between: the proposed method with multiple sources of biometric information, and eigenfaces.

In fact, these experiments are a clear demonstration of the advantage of multi-biometrics, in which complementary sources of biometric information can increase the accuracy and augment the reliability of the resulting multimodal system, by taking advantage of their redundant and richer information to compensate their individual weaknesses.

However, due to the noisy tracking signals and the noisy mouth parameters in the temporal subsystem, along with the reduced dimensionality of the face space and complexity of the GMM modelling in the spatial one, our recognition system is still far from its optimal working condition, so the potential discriminatory power of facial appearance and head and mouth motion for gender recognition is probably higher than the one established in these experiments.

## V. Conclusion and Future Works

In this article we presented a novel multimodal gender recognition system, which successfully integrates the head and mouth motion information with facial appearance by taking advantage of a unified probabilistic framework. By looking at the results of our experiments, we believe that not only facial appearance but also head and mouth motion possess a potentially relevant discriminatory power, and that the integration of different sources of biometric information from video sequences is the key strategy to develop more accurate and reliable recognition systems.

There are different ways to improve our present multimodal approach. First of all, our result should be validated by using larger datasets and different scenarios to evaluate the impact of stress or of varying emotional states. Then, the temporal subsystem can be improved by increasing the accuracy of the tracking signals, for example by implementing a more robust tracker. We should also increase the precision of the segmented lip contours, either by exploiting a database of higher quality, or by improving the robustness of the segmentation algorithm. In addition, we could extend the feature space by extracting more parameters related to facial motion, like: the lip curvature, the lip area, the motion of the pupils, eyebrows, etc. Afterwards, the static subsystem can apply a more discriminating space reduction technique, like LDA, CCA, etc., or a more performing strategy if compatible with our probabilistic framework. Finally, more elaborate fusion techniques, like a post-classifier, might improve the integration of the discriminating information of our system, but the amount of required data could be a problem.

A possible future application could be in the domain of impostor detection, when a male is disguised as a woman. In particular, it could be interesting to study the incoherencies between the static subsystem, working on facial appearance, and the temporal one, based on head and facial motion.

## References

[1] Saeed U., Matta F. and Dugelay J.-L., "Person recognition based on head and mouth dynamics", in *IEEE Proceedings on Multimedia Signal Processing*, pag. 29-32, October 2006.

[2] Golomb B.A., Lawrence D.T. and Sejnowski T.J., “Sex-Net: a neural network identifies sex from human faces”, in *Proceedings of Advances in neural information processing systems*, pag. 572-577, 1990.

[3] Cottrell G.W. and Metcalfe J., “EMATH: face, emotion, and gender recognition using holons”, in *Proceedings of Advances in neural information processing systems*, pag. 564-571, 1990.

[4] Brunelli R. and Poggio T., “HyperBF networks for gender classification”, in *Proceedings of the DARPA Image Understanding Workshop*, pag. 311-314, 1992.

[5] Wiskott L., Fellous J.M., Kruger N. and Von der Malsburg C., “Face recognition and gender determination”, in *Proceedings on Automatic Face and Gesture Recognition*, pag. 92-97, 1995.

[6] Tamura S., Kawai H. and Mitsumoto H., “Male/female identification from 8x6 very low resolution face images by neural network”, in *Pattern Recognition*, vol. 29, iss. 2, pag. 331-335, February 1996.

[7] Sun Z., Bebis G., Yuan X. and Louis S.J., “Genetic feature subset selection for gender classification: a comparison study”, in *IEEE Proceedings on Applications of Computer Vision*, pag. 165-170, 2002.

[8] Gutta S., Huang J.R.J., Jonathon P. and Wechsler H., “Mixture of experts for classification of gender, ethnic origin, and pose of human faces”, in *IEEE Transactions on Neural Networks*, vol. 11, iss. 4, pag. 948-960, July 2000.

[9] Nakano M., Yasukata F. and Fukumi M., “Age and gender classification from face images using neural networks”, in *Signal and Image Processing*, 2004.

[10] Lu X., Chen H. and Jain A.K., “Multimodal facial gender and ethnicity identification”, in *Advances in Biometrics*, vol. 3832/2005, pag. 554-561, December 2005.

[11] Moghaddam B. and Yang M.-H., “Gender classification with support vector machines”, in *IEEE Proceedings on Automatic Face and Gesture Recognition*, pag. 306-311, March 2000.

[12] Saatci Y. and Town C., “Cascaded classification of gender and facial expression using active appearance models", in *Automatic Face and Gesture Recognition*, pag. 393-398, April 2006.

[13] Kim H.-C., Kim D., Ghahramani Z. and Bang S.Y., “Appearance-based gender classification with Gaussian processes”, in *Pattern Recognition Letters*, vol. 27, iss. 6, pag. 618-626, April 2006.

[14] Zhiguang Y., Ming L. and Haizhou A., “An Experimental Study on Automatic Face Gender Classification”, in *IEEE Proceedings on Pattern Recognition*, pag. 1099-1102, August 2006.

[15] Baluja S. and Rowley H.A., “Boosting Sex Identification Performance”, in *International Journal of Computer Vision*, vol. 71, iss. 1, pag. 111-119, January 2007.

[16] Caifeng S., Shaogang G., and McOwan P.W., “Learning gender from human gaits and faces”, in *IEEE Proceedings on Advanced Video and Signal Based Surveillance*, pag.505-510, September 2007.

[17] Turk M.A. and Pentland A.P., "Face recognition using eigenfaces", in *IEEE Proceedings on Computer Vision and Pattern Recognition*, pag. 586-591, June 1991.