Manifold Learning for Gender Classification from Face Sequences

Abdenour Hadid and Matti Pietikäinen

Machine Vision Group, P.O. Box 4500, FI-90014, University of Oulu, Finland

Abstract. We propose a novel approach to gender recognition for cases when face sequences are available. Such scenarios are commonly encountered in many applications such as human-computer interaction and visual surveillance in which input data generally consists of video sequences. Instead of treating each facial image as an isolated pattern and then combining the results (at feature, decision or score levels) as generally done in previous works, we propose to exploit the correlation between the face images and look at the problem of gender classification from manifold learning point of view. Our approach consists of first learning and discovering the hidden low-dimensional structure of male and female manifolds using an extension to the Locally Linear Embedding algorithm. Then, a target face sequence is projected into both manifolds for determining the gender of the person in the sequence. The matching is achieved using a new manifold distance measure. Extensive experiments on a large set of face sequences and different image resolutions showed very promising results, outperforming many traditional approaches.

1 Introduction

Determining the gender (i.e. man or woman) of a person in a given image or video is useful for many applications such as more affective Human-Computer Interaction (HCI), content-based image and video retrieval, restricting access to certain areas based on gender, collecting demographic information in public places, counting the number of women entering a retail store and so on.

First attempts of using computer vision based techniques to gender classification started in early 1990s. Since then, many approaches have been reported in literature. Among the most notable results to date are those obtained by Moghaddam and Yang [1], and also by Baluja and Rowley [2]. Moghaddam and Yang used raw pixels as inputs to Support Vector Machines (SVMs) and achieved a classification rate of 96.6% on FERET database of images scaled to 12×21 pixels [1]. Note that the considered FERET images were very clean and some persons may have appeared in both training and test sets. Comparable accuracy but at a higher speed was also reported by Baluja and Rowley who used AdaBoost to combine weak classifiers, constructed using simple pixel comparisons, into single strong classifier [2]. Note that both approaches are based on still images and assume well aligned faces. However, in many real-world applications (e.g. HCI, visual surveillance and content-based video retrieval) input

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 82–91, 2009.

data generally consists of video sequences and it is not always obvious to hold the face alignment assumption. So, the question which arises then is how to efficiently perform gender classification from face sequences? We believe that this issue is understudied in literature as most works tend to process each frame (or some selected ones) and then fuse the results at feature, decision or score levels. Obviously, such approaches ignore the correlation between the face images as they are only treated as "single" or "isolated" patterns in the image space while, in fact, they lie on a low-dimensional manifold.

From these observations, we propose to look at the problem of gender classification from manifold learning point of view. The goal of face manifold learning is to discover the hidden low-dimensional structure of the face images. This is a very useful but also a challenging unsupervised learning problem. It is a useful task because mapping the high-dimensional faces into low-dimensional coordinates would facilitate and speed-up the further analysis (i.e. gender classification in our case) and also may avoid "the curse of dimensionality". It is also a challenging problem since the face manifold is highly nonconvex and nonlinear and so is the face manifold of any individual under changes due to pose, illumination and other factors.

Therefore, instead of treating each facial image as an isolated pattern and then fusing the results (at feature, decision or score levels) as generally done in previous works, we propose to exploit the correlation between the face images and consider the problem of gender classification from manifold learning point of view. Our approach consists of first learning and discovering, from a set of training face sequences, the hidden low-dimensional structure of male and female classes, thus yielding in male and female face manifolds. Then, a target face sequence can be projected into both manifolds for determining the gender of the person in the sequence. The "closest" manifold (in terms of a newly introduced manifold distance measure) will then determine the gender of the person. In the rest of this paper, we further develop the proposed approach and validate it through extensive experiments and comparisons against traditional approaches. Among the novel contributions in this work are:

- → A new method to gender classification from face sequences is presented and extensively evaluated;
- → An extension to the Locally Linear Embedding algorithm to handle face sequences is proposed;
- → A simple yet efficient manifold to manifold distance measure is introduced.

2 Locally Linear Embedding

The first key issue in our proposed approach is to learn the face manifolds by embedding the face images into low-dimensional coordinates. For that purpose, there exist several methods. The traditional ones are Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). These methods are simple to implement and efficient in discovering the structure of data lying on or near linear subspaces of the high-dimensional input space. However, face images do

not satisfy this constraint as they lie on a complex nonlinear and nonconvex manifold in the high-dimensional space. Therefore, such linear methods generally fail to discover the real structure of the face images in the low-dimensional space. As an alternative to PCA and MDS, one can consider some nonlinear dimensionality reduction methods such as Self-Organizing Maps (SOM) [3], talent variable models [4], Generative Topographic Mapping (GTM) [5], Sammon's Mappings (SM) [6] etc. Though these methods can also handle nonlinear manifolds, most of them tend to involve several free parameters such as learning rates and convergence criteria. In addition, most of these methods do not have an obvious guarantee of convergence to the global optimum. Fortunately, in the recent years, a set of new manifold learning algorithms have emerged. These methods are based on an Eigen decomposition and combine the major algorithmic features of PCA and MDS (computational efficiency, global optimality, and flexible asymptotic convergence guarantees) with flexibility to learn a broad class of nonlinear manifolds. Among these algorithms are Locally Linear Embedding (LLE) [7], ISOmetric feature MAPping (ISOMAP) [8] and Laplacian Eigenmaps [9].

We adopt in our present work the LLE approach for its demonstrated simplicity and efficiency to recover meaningful low-dimensional structures hidden in complex and high dimensional data such as face images. LLE is an unsupervised learning algorithm which maps high-dimensional data onto a low-dimensional, neighbor-preserving embedding space.

In brief, considering a set of N face images and organizing them into a matrix X (where each column vector represents a face), the LLE algorithm involves then the following three steps:

1. Find the k nearest neighbors of each point X_i.
2. Compute the weights W_{ij} that best reconstruct each data point from its neighbors, minimizing the cost in Equation (1):

$$\epsilon(W) = \sum_{i=1}^{N} \left\| X_i - \sum_{j \in neighbors(i)} W_{ij} X_j \right\|^2 \tag{1}$$

   while enforcing the constraints $W_{ij} = 0$ if X_j is not a neighbor of X_i, and $\sum_{j=1}^{N} W_{ij} = 1$ for every i (to ensure that W is translation-invariant).
3. Compute the embedding Y (of lower dimensionality $d << D$, where D is the dimension of the input data) best reconstructed by the weights W_{ij} minimizing the quadratic form in Equation (2):

$$\Phi(Y) = \sum_{i=1}^{N} \left\| Y_i - \sum_{j \in neighbors(i)} W_{ij} Y_j \right\|^2 \tag{2}$$

   under constraints $\sum_{i=1}^{N} Y_i = 0$ (to ensure a translation-invariant embedding) and $\frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^T = 0$ (normalized unit covariance).

The aim of the first two steps of the algorithm is to preserve the local geometry of the data in the low-dimensional space, while the last step discovers the global

structure by integrating information from overlapping local neighborhoods. LLE is an efficient approach to compute the low-dimensional embeddings of high-dimensional data assumed to lie on a non-linear manifold. Its ability to deal with large sizes of high-dimensional data and its non-iterative way to find the embeddings make it attractive. However, in its actual formulation, the LLE method can only be used with still images but not with video sequences. In the next section, we introduce extensions to LLE to also handle face sequences and then present our proposed approach to gender classification from face sequences.

3 The Proposed Approach to Gender Classification from Face Sequences

Given a set of training face sequences of males and females. First, let us denote the male face sequences as $\{Seq_1, Seq_2, ..., Seq_M\}$. In order to discover the male face manifold in the low-dimensional space, we modify and apply the LLE algorithm as follows:

1′.a. Let $X_i^{(j)}$ be the column vector representing the i^{th} face image of the j^{th} sequence Seq_j.

1′.b. In every sequence Seq_j, find the k nearest neighbors of each point $X_i^{(j)}$ among all points which do not belong to the sequence Seq_j.

2′. Compute the weights on the newly constructed data and neighborhood in a similar way as in Step 2 of the original LLE algorithm.

3′. Compute the embedding in a similar way as in Step 3 of the original LLE algorithm.

Note that, in this modified version of LLE, we enforce (in Step **1′.b.**) the neighbors of each point to be taken from other sequences than the one to which the data point belongs. The aim of this constraint is to avoid constructing disjoint manifolds, each corresponding to an individual. The other reason is that we are not only focusing on the relationships between the image instances of the same individual but more importantly we are also looking at learning the relationships between the images of different subjects from the entire male class. In other terms, we are interested in discovering the features and structure that are shared among the male faces. The introduced changes to LLE are somehow inspired by the recent works on semi-supervised locally linear embedding for classification (e.g. [10]). From now and on, we only consider LLE with the proposed changes. Therefore, for the clarity of the presentation, we continue using the term LLE to refer to the modified version of the algorithm in the rest of this paper.

After embedding the face sequences $\{Seq_1, Seq_2, ..., Seq_M\}$ using LLE, we obtain the coordinates of each face image in the low-dimensional space, thus defining the face manifold of the male class. Let us denote then the obtained embedding result as M_Y. Note that the calculation of M_Y involves only two free parameters which are the number of neighbors (k) and the dimension of the embedding space (d). We discuss the values of these two parameters in the experimental analysis in Section 4.

Similarly to the analysis of the male sequences, we also apply our described procedure to the female training face sequences to compute the face manifold F_Y corresponding to the female class. Thus, we finally obtain two embeddings (face manifolds M_Y and F_Y) that are extremely useful to gender classification of unseen individuals in target face sequences, as the next section explains.

Classifying a New Face Sequence?

To determine the gender of an unknown person in a given face sequence

$$\{Face_{frame(1)}, Face_{frame(2)}, ..., Face_{frame(L)}\}$$

we first project every face instance $Face_{frame(i)}$ into both male and female face manifolds in the low-dimensional space. The "closest" manifold will then determine the gender of the person in the sequence. The projection of the target face sequence into the male manifold, for instance, is done using the following steps:

a. Let now X_i be the column vector representing the face image ($Face_{frame(i)}$) from the new sequence.
b. Find the k nearest neighbors of each point X_i among the male training face samples.
c. Compute the weights W_{ij} that best reconstruct each data point X_i from its neighbors using Equation (1).
d. Use the obtained weights W_{ij} to compute the embedding Y_i of each point X_i (i.e. $Face_{frame(i)}$) as:

$$Y_i = \sum_{j \in neighbors(X_i)} W_{ij} \, M_Y^j \tag{3}$$

where M_Y^j refers to the embedding point in the male manifold of the j^{th} neighbor of the point X_i.

As a result, we obtain the embedding Y of the new face sequence in the male manifold. Then, we compute how close is the embedding Y to the male manifold M_Y using:

$$D_{male} = \frac{1}{L} \sum_{i=1}^{L} \left\| Y_i - M_Y^{r(i)} \right\| \tag{4}$$

where L is the length of the target face sequence, Y_i is the embedding of the point X_i in the low-dimensional space and $M_Y^{r(i)}$ is the closest point (in term of Euclidean distance) from the manifold M_Y to Y_i as shown in Fig. 1.

In the same way as for the male manifold, we also compute the embedding of the new face sequence into the female manifold and then calculate the distance D_{female}. Finally, the smallest value between the two distances (D_{female} and D_{male}) defines the "closest" manifold and thus determines the gender of the person in the new sequence, as formulated in Equation (5).

$$Gender(Face_{frame(1)}, ..., Face_{frame(L)}) = \begin{cases} Man, & if \ D_{male} \prec D_{female} \\ Woman, & otherwise \end{cases} \tag{5}$$

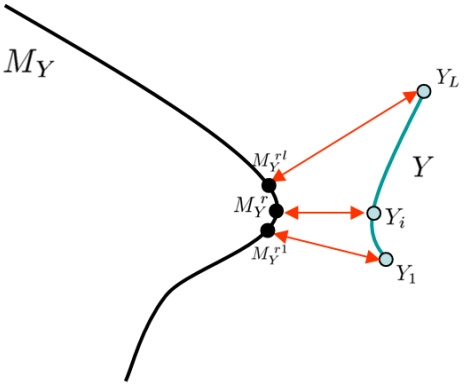

Fig. 1. Calculation of the distance between the manifold of a new face sequence (represented by its embedding Y) and the male manifold represented by its embedding M_Y

4 Experimental Analysis

4.1 Experimental Data

For experimental analysis of our proposed approach, we considered three different publicly available video face databases (namely CRIM [11], VidTIMIT [12] and Cohn-Kanade [13]) containing several face sequences subject to changes caused by different factors including face image resolution, illumination variations, head movements and facial expressions. CRIM is a large set of 591 face sequences showing 20 persons (10 female and 10 male) reading broadcast news for a total of about 5 hours. There are between 23 and 47 video sequences for each individual. The VidTIMIT database consists of audio recordings and video sequences of 43 subjects (19 female and 24 male), reciting ten short sentences

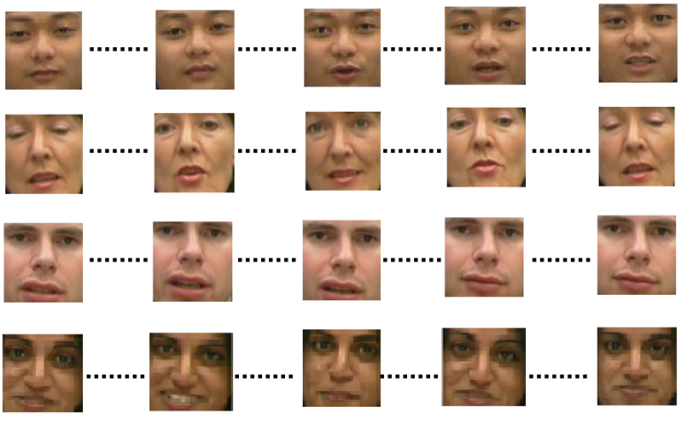

Fig. 2. Examples of face images from some sequences from the considered datasets

in three sessions with an average delay of a week between sessions, allowing for appearance and mood changes. Cohn-Kanade database consists of 100 subjects expressing different emotions like anger, disgust, fear, joy, sadness, and surprise. Sixty-five percent of the subjects were female, 15 percent African-American, and three percent Asian or Latino. We randomly segmented the datasets and extracted over 1000 video shots of 15 to 80 frames each. From each shot or sequence, we automatically detected the eye positions from the first frame. The determined eye positions are then used to crop the facial area in the whole sequence, yielding in nonaligned face images. Finally we scaled the resulted images into 3 different resolutions: 20×20, 40×40 and 60×60 pixels. Examples of face images from some sequences are shown in Fig. 2.

4.2 Results and Analysis

For evaluation, we adopted a 5-fold cross validation test scheme by dividing the 1000 sequences into five groups and using the data from four groups for training and the last group for testing. We repeated this process five times and we report the average classification rates.

As the proposed approach involves two free parameters which are the number of neighbors (k) and the dimension of the embedding space (d), we started then by analyzing the effect of these two factors on the system performance. Fig. 3 plots the gender classification results when varying k and d (These results are obtained using 40×40 pixels as face image resolution). For k, we can notice that the best rates are obtained using around 15 nearest neighbors and then the performance is decreasing for too small ($k \leq 6$) or too large ($k \geq 25$) values. This can be explained by the fact that too small values of k tend to make the reconstruction of each data point from its neighbors weak (see Equation (1)), while too large values tend to break the embedding as the locality assumption in LLE calculation may not hold. However, for a wide range of values between 10 and 20, the results remain good and stable. For the dimension of the embedding space (d), we can notice that the best results are obtained using values between $10 \leq d \leq 15$. Smaller values of d tend to decrease the performance while larger values ($d \geq 15$) do not alter the results. In fact, d corresponds to the intrinsic dimension of the training face sequences. An alternative to estimate the optimal value of d is to analyze the residual variance which characterizes how well the low-dimensional Euclidean embedding captures the geodesic distances of the input data. The lower the residual variance is, the better high-dimensional data are represented in the embedded space. We computed the residual variance as $1 - \rho^2(D_X, D_Y)$, where ρ is the standard linear correlation coefficient taken over all entries of D_X and D_Y; D_X is the matrix of geodesic distances between pairs of points in X (in the high-dimensional space) and D_Y is the matrix of Euclidean distances in Y (in the embedding space). Our experiments showed that the residual variance does not further decrease for dimensions ≥ 12 which consolidates our earlier findings.

Setting $k = 15$ and $d = 12$, we obtained gender classification rates of **96.8**%, **97.1**% and **97.2**% for resolutions of 20×20, 40×40 and 60×60 pixels,

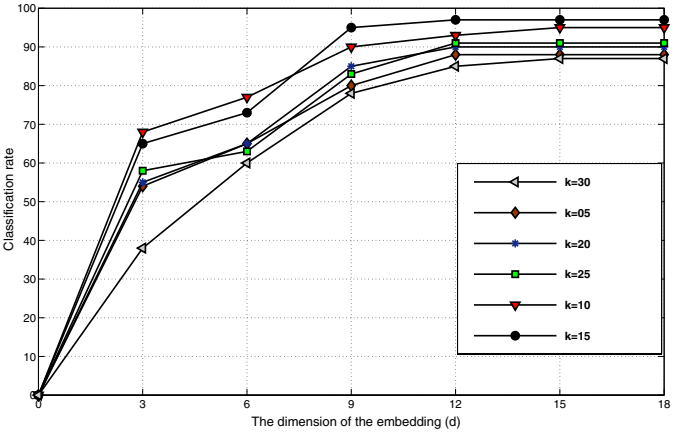

Fig. 3. Gender classification rates for different values of the number of neighbors (k) and the dimension of the embedding space (d)

Table 1. Comparison of gender classification results of several approaches on our test sequences with three different face image resolutions

Method	Average Gender Classification Rate (%)		
	20×20 Pixels	40×40 Pixels	60×60 Pixels
Pixels + SVM + Fusion	88.0	89.2	88.5
LBP + SVM + Fusion	90.5	91.0	92.1
VLBP + SVM	82.3	83.9	84.5
EVLBP + AdaBoost	83.1	84.4	84.6
Manifold Learning (Proposed Approach)	**96.8**	**97.1**	**97.2**

respectively. Although we did not perform any face alignment, the obtained results are very good and compare favourably against what is generally reported in literature even for well aligned faces [1][2]. For extensive and more objective comparison, we also implemented some state-of-the-art methods including **(i)** Use of raw pixels with SVM and fusion over the face sequence at score level; **(ii)** Use of Local Binary Pattern features (LBP) [14] with SVM and fusion over the face sequence at score level [15]; **(iii)** Use of Volume LBP with SVM [16] and **(iv)** Use of Extended Volume LBP with AdaBoost [15]. The obtained results using all these methods are summarized in Table. 1. From the table, we can clearly see that our proposed approach significantly outperforms all other considered methods.

Perhaps, the key reason behind the success of our approach lies in the effective use of the correlation between the face images through manifold learning. Our experimental analysis also showed that most state-of-the-art methods fail to handle faces under severe illumination changes and miss alignments, while our proposed approach seems to perform much better under such conditions.

Our results also indicated that image resolution does not affect very much gender classification performance and this confirms the conclusions of many other researchers.

5 Discussion and Conclusion

We proposed a novel approach to gender classification from face sequences. Our approach consisted of first learning the hidden low-dimensional structure of male and female manifolds using an extension to the original LLE algorithm. Then, a target face sequence is projected into both manifolds for determining the gender of the person in the sequence. The matching is achieved using a new manifold distance measure. Under natural conditions and without face alignment, a high classification rate of 97.2% is obtained for face sequences of 60×60 pixels, while most state-of-the-art methods fail to handle faces under miss alignments and illumination changes.

It is worth noting that, in addition to its efficiency, our approach involves only two free parameters which are quite easy to determine. Our future work consists of experimenting with much larger databases and further analyzing the few misclassification errors made by our system. It is also of interest to extend our approach to explicitly incorporate the temporal information during manifold learning.

Note that the proposed methodology can also be applied to other face analysis tasks such as face recognition, facial expression recognition and age classification from videos.

Acknowledgment

The financial support of the Academy of Finland is gratefully acknowledged.

References

1. Moghaddam, B., Yang, M.H.: Learning gender with support faces. IEEE Transactions on PAMI 24(5), 707–711 (2002)
2. Baluja, S., Rowley, H.: Boosting sex identification performance. International Journal of Computer Vision 71, 111–119 (2007)
3. Kohonen, T. (ed.): Self-Organizing Maps. Springer, Berlin (1997)
4. Bartholomew, D.J., Knott, M. (eds.): Latent variable models and factor analysis multidimensional scaling. Arnold Publication (1999)
5. Bishop, C.M., Svensen, M., Williams, C.K.I.: GTM: The generative topographic mapping. Neural Computation 10(1), 215–234 (1998)
6. Sammon, J.: A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 18(5), 401–409 (1969)
7. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
8. Tenenbaum, J.B., DeSilva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)

9. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, pp. 585–591. MIT Press, Cambridge (2002)
10. Kouropteva, O., Okun, O., Pietikäinen, M.: Semisupervised visualization of high-dimensional data. Pattern Recognition and Image Analysis 15(4), 645–648 (2005)
11. CRIM, http://www.crim.ca/
12. Sanderson, C., Paliwal, K.K.: Noise compensation in a person verification system using face and multiple speech feature. Pattern Recognition 36(2), 293–302 (2003)
13. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: FG 2000, pp. 46–53 (2000)
14. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI 24, 971–987 (2002)
15. Hadid, A., Pietikäinen, M.: Combining motion and appearance for gender classification from video sequences. In: Proc. 19th International Conference on Pattern Recognition (ICPR 2008) (2008) (in press)
16. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE TPAMI 29(6), 915–928 (2007)