Neurocomputing 74 (2011) 2407–2410

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Bimodal biometric verification based on face and lips

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

[a] Signals and Communications Department, Institute for Technology Development and Innovation in Communication, University of Las Palmas de Gran Canaria, Campus de Tafira, Edificio de Telecomunicación, Pabellón B, E-35017 Las Palmas de Gran Canaria, Spain

[b] School of Computing, University of Dundee, Dundee DD1 4HN, 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 28 May 2010
Received in revised form
4 January 2011
Accepted 10 March 2011
Communicated by M.T. Manry
Available online 30 March 2011

Keywords:
Biometrics
Lips-based identification
Fusion score
Image processing

ABSTRACT

In this paper, we present a novel approach to person verification by fusing face and lip features. Specifically, the face is modeled by the discriminative common vector and the discrete wavelet transform. Our lip features are simple geometric features based on a lip contour, which can be interpreted as multiple spatial widths and heights from a center of mass. In order to combine these features, we consider two simple fusion strategies: data fusion before training and score fusion after training, working with two different face databases. Fusing them together boosts the performance to achieve an equal error rate as low as 0.4% and 0.28%, respectively, confirming that our approach of fusing lips and face is effective and promising.

1. Introduction

Over the past decade, face recognition has become one of the main biometrics showing potential for deployment in modern access control systems due to its ease of capture and the relatively low level of cooperation required. Numerous approaches to face recognition have been proposed [1,2]; however, up to now, there has been very little work performed on using lips, which are a very important feature of the human face in their own right. Effective biometric verification using lip movement has been experimentally demonstrated [2,3]. Recently, researchers have shown that static lips are an additional cue for person verification [4]. To the best of our knowledge, little work has been done on fusing face and static lip shape features. In this paper we propose that combining face with simple lip features could outperform each of them individually, and we demonstrate the validity of this argument by experiment.

2. Methodology

In this section we introduce our feature extraction strategy for face and lips (see Fig. 1).

2.1. Face detection

Our system starts by detecting human faces from a natural image. Though many face detectors are available in the literature, we use a simple front face detector, similar to the Viola–Jones cascade detector [5], due to its simplicity, speed, and effectiveness.

2.2. Face descriptor

We use the discriminative common vector (DCV) [6–8] and discrete wavelet feature transform (DWT) as our face descriptors, since they are widely reported in the existing literature [9]. The DCV feature is a subspace method, while DWT is a function transform based approach. The DCV feature was first proposed in [7] and has demonstrated superior performance over other subspace based methods, including Eigenface, Fisherface, and Direct-LDA, in terms of accuracy, speed, and storage [8]. In principle, the DCV is a feature mining approach that has the capability to extract common features of each class. A common vector is computed by eliminating all feature vectors that are in the direction of eigenvectors corresponding to the non-null eigenvalues of the scatter matrix of its own class. After its computation, a DCV for each class is obtained by the product between the common vectors and a training sample of that class. The resulting DCV is then used for verification. The following gives a detailed description of the computation process.

* Corresponding author. Tel.: +34 928458264; fax: +34 928451243.
E-mail addresses: ctravieso@dsc.ulpgc.es (C.M. Travieso), jgzhang@computing.dundee.ac.uk (J. Zhang).
URL: http://www.gpds.ulpgc.es (C.M. Travieso).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.03.012

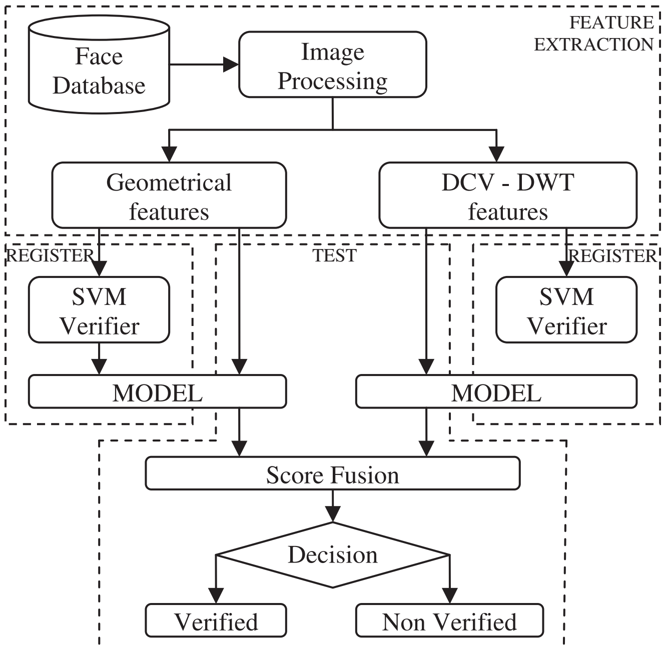

Fig. 1. Proposed approach.

We define a within-class scatter matrix (S_W) as;

$$S_W=\sum_{i=1}^{C}\sum_{m=1}^{N_i}(x_m^i-\mu_i)(x_m^i-\mu_i)^T \text{ where } \mu_i=\frac{1}{N_i}\sum_{m=1}^{N_i}x_m^i \tag{1}$$

where C represents the total number of classes, i the ith class. N_i is the number of samples of class i, x_m^i denotes the mth sample from the ith class and μ_i is mean vector of the ith class.

Suppose that the dimension of an original sample space is d, r is the rank of S_W ($V\epsilon R^r$), and $d-r$ is the null space rank of S_W ($V^{\perp}\epsilon R^{(d-r)}$), where V is the range space of S_W and $V^{\perp}$ is the null space of S_W. Let $Q=[\alpha_1\ldots\alpha_r]$ be a set of eigenvectors corresponding to nonzero eigenvalues and $\overline{Q}=[\alpha_{r+1}\ldots\alpha_d]$ be a set of eigenvectors corresponding to eigenvalues of zero. It has been proven that a unique common vector x_{com} in the same class can be obtained [7];

$$x_{com}^i=x_m^i-QQ^Tx_m^i=\overline{Q}\,\overline{Q}^Tx_m^i \tag{2}$$

Now we can get the common vector x_{com} from (2) (note, x_{com} is independent of the sample index m). It shows that we can calculate x_{com} by the eigenvectors spanning the range space or those spanning the null space. By performing the above steps, we obtain C common vectors. We then find the principal components between C classes by the covariance matrix of x_{com}^i:

$$S_{com}=\sum_{i=1}^{C}(x_{com}^i-\mu_{com})(x_{com}^i-\mu_{com})^T \text{ where } \mu_{com}=\frac{1}{C}\sum_{i=1}^{C}x_{com}^i \tag{3}$$

Here S_{com} is a d-by-d matrix that is the covariance matrix of the common vectors. Finally, we obtain a projection matrix $W=[\alpha_1^{com}\ldots\alpha_{C-1}^{com}]$ from α^{com}, where α^{com} is the eigenvector of S_{com}.

Because we have only C common vectors the projection matrix contains at most $C-1$ eigenvectors corresponding to nonzero eigenvalues, since the last eigenvector is the linear combination of the other eigenvectors. When a new test sample is given (x_{test}), we compute the test vector using the project matrix W as $\Omega_{test}=W^Tx_{test}$; after his projection, we have obtained the DCV features. Because W was obtained from S_{com}, which is in the rank space of x_{com}, we can also project samples of the same class onto one point by the projection matrix W directly.

For the purpose of comparison, our second face descriptor is DWT [9], a well-known function transform based approach. Mathematically, it is defined as follows:

$$C[j,k,l]=\sum_{n,m\in Z}f[n,m]\psi_{j,k,l}[n,m] \tag{4}$$

The same symbol C is used for the number of classes, where $f[n,m]$ is our image and $\psi_{j,k,l}$ is the transform function:

$$\psi_{j,k,l}[n,m]=2^{-j/2}\psi[2^{-j}n-k,2^{-j}m-l] \tag{5}$$

In our experiment, we have used "bio5.5" [10]; it is a discrete biorthogonal wavelet family with three pyramid levels for our feature extraction, similar to [9], obtaining 19×15 DWT coefficients.

2.3. Lip descriptor

To build the lip descriptor, we first need to extract the lip region. To do this, we use a simple transformation to convert a color image to gray scale image using $R-(2G)+B$ to enhance the lip region. Fig. 2 shows examples of face images after this transformation, where enhanced lip regions can be clearly identified as the interior middle region from the face detection. A binarization method by Otsu [11] is further applied on the enhanced image to segment the lips. We then extract the lip shapes using morphological operators, subtracting the binary image and its dilated version (see Fig. 2—there are no images of this in Fig. 2) obtained with a mask size of 3×3. We would like to clarify that the dataset we have studied does include users with beard or mustache as shown in Fig. 3. From this, we can see that our approach is still capable of extracting relatively reliable lip contour. Open lips issue does occur frequently in the research of lip reading. However, we would like to argue that it is very rare for a normal person to leave his/her mouth open when being quiet, such as not speaking, laughing, crying, etc. We would like to categorize this open issue into the research area of biometrics based on lip dynamics; see [2] for some initial work on this.

A set of geometric features are extracted based on the distances from equally sampled points along the vertical and horizontal centroid axes, to points on the lip contour. For the vertical centriod axis, we equally sample 300 points from a right initial point to the other left end of the mouth and while sampling

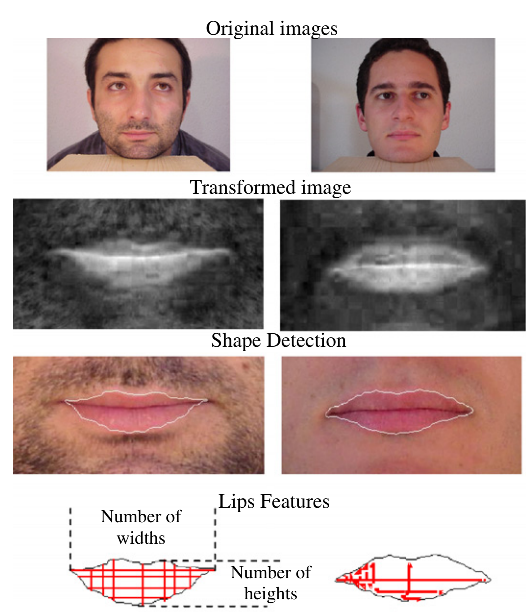

Fig. 2. Lips feature extraction.

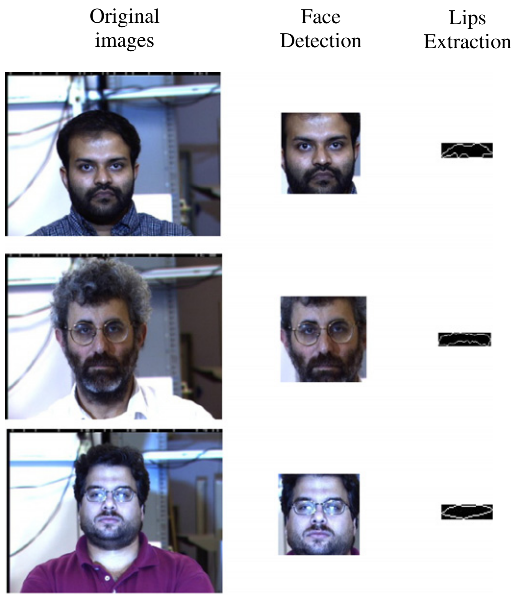

Fig. 3. Lips extraction for users with beard.

180 points for the horizontal centriod axis, from the top point to other bottom point of the mouth (see Fig. 2). We used different numbers of sampling points, but the best result was reached with 300 points vertically and 180 points horizontally. These values are the normalization of the geometric lip features, in order to get scale invariance for different depths of field for each photo. Therefore, the size of the mouth can be different for each photo, and we normalize its size. The resulting normalized feature vector is of size 480 elements by concatenation.

3. Classification and fusion system

A support vector machine (SVM) with radial basis function kernel is used for our experiment as it has been demonstrated to be a good classifier for face verification [9]. Specifically, we used the SVM light implementation [9] with a RBF kernel:

$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\delta}\right) \tag{6}$$

We have used two different strategies for fusion, based on data and score. For data fusion, we concatenated each feature vector (lips and face features) into one feature vector before input to the classifier. For the score fusion we have implemented two different fusion rules, based on the likelihood function of individual verifiers as follows:

$$\text{a) } \max \sum_{i=1}^{2} p(X'/\lambda_i) \text{ and b) } \max \prod_{i=1}^{2} p(X'/\lambda_i) \tag{7}$$

here λ_i is the model per user (verification approach) from the SVM of the feature modality i (face and lips), and $p(X'/\lambda_i)$ is the likelihood function of the feature vector from a test sample (X') against the model λ_i.

4. Experiments and results

We constructed a face database for our experiments, named the GPDS-ULPGC Face Database, which consists of 50 users with 10 samples per user (500 images in total) with the average detected face size of 800×600 pixels, and besides, we have used a public database, the PIE Face Database [12], which consists of 68 users with 11 samples per user (748 images in total), with the average detected face size of 200×200 pixels. We report our results in terms of the equal error rate (EER) based on 10 runs/splits in a similar fashion with cross validation. For each run/split, we randomly select 50% of the samples for training and the rest for testing. This test is repeated 10 times, thus constructing the 10 runs/split protocol. The mean accuracy and standard deviation of EERs of those runs among all classes are reported.

We tried six different cases: (1) lips only, (2) DCV face only, (3) DWT face only, (4) combining lips and face by feature fusing before training, (5) combining by score fusing (sum rule), and (6) combining by score fusing (product rule). For each case, our system contains two parameters: the EER threshold for verification (it is the decision value at the point of EER—equal error rate in the ROC curve) and the width of the RBF kernel (δ, see Tables 1 and 2), which is automatically determined by the SVM using grid search on training images only. Those best parameters are selected based on a grid search. Tables 1 and 2 show the corresponding result of each case.

From Table 1, we can see that lip shape and face appearance perform well individually, with lip shape giving slightly better results for the GPDS-ULPGC Face Database. For face verification only, the DCV feature performs better than the DWT feature (3.48% vs. 3.71%). Based on these results, we select the DCV as the face descriptor when combining face and lip together (for comparison, we also include the results of fusing face DWT feature and lips shape features). Table 1 also shows that data fusing (2.32%) by feature concatenation performs worse than score fusing (0.44% and 0.43%). From Table 2, using the PIE Face Database, we have obtained similar or even better results (see Table 2). We achieved the error rate as low as 0.28% using the score fusing, for lips and faces. This observation is consistent with the experiments on our own dataset, i.e., fusing the two modalities together boosts the performance.

This is because perhaps, in the case of feature concatenation, the SVM assigns a single weight to each sample in the training process for the two types of features, whilst in the case of score fusion the two trained SVMs assign different weights to each sample, thus reflecting the different roles of lip features and face in discrimination. We further notice that the probabilistic product rule performs similar to the sum rule, which indicates that lip geometric features are likely to be independent from face DCV

Table 1
EER for different approaches with lip and face data for GPDS-ULPGC Face Database.

Kind of feature	EER	Threshold	δ	Kind of fusion
Lips	2.59% ± 0.56	−0.21	6×10^{-6}	Not fusion
DCV faces	3.48% ± 0.37	−0.20	2×10^{-5}	Not fusion
DWT faces	3.71% ± 0.58	−0.16	5×10^{-9}	Not fusion
Lips + DCV faces	2.32% ± 0.57	−0.33	5×10^{-7}	Feature fusion
Lips + DWT faces	2.55% ± 0.82	−0.35	2×10^{-9}	Feature fusion
Lips + DWT faces	1.63% ± 0.22	−0.42	—	Score fusion (sum)
Lips + DWT faces	1.67% ± 0.20	−0.32	—	Score fusion (product)
Lips + DCV faces	0.43% ± 0.17	−0.40	—	Score fusion (sum)
Lips + DCV faces	0.44% ± 0.11	−0.52	—	Score fusion (product)

Table 2
EER for different approaches with lip and face data for PIE Face Database.

Kind of feature	EER	Threshold	δ	Kind of fusion
Lips	12.75% ± 0.76	−0.76	2×10^{-5}	Not fusion
DCV faces	1.74% ± 0.49	−0.37	7×10^{-6}	Not fusion
DWT faces	0.71% ± 0.38	−0.06	1×10^{-6}	Not fusion
Lips + DCV faces	9.17% ± 0.39	−0.62	2×10^{-5}	Feature fusion
Lips + DWT faces	9.57% ± 0.57	−0.43	9×10^{-6}	Feature fusion
Lips + DWT faces	0.42% ± 0.02	−0.53	—	Score fusion (sum)
Lips + DWT faces	0.40% ± 0.03	−0.45	—	Score fusion (product)
Lips + DCV faces	0.29% ± 0.05	−0.41	—	Score fusion (sum)
Lips + DCV faces	0.28% ± 0.05	−0.35	—	Score fusion (product)

features and that they are complementary. Overall, fusing the lip shape features with the face DCV features performs best.

5. Conclusion

This paper presents an effective and novel personal verification approach by fusing face appearance and lip shape. The face appearance feature is described by DVC, while lip shape is represented by simple geometrical features. Our study shows that both types of features are effective and complementary. Fusing them under the probability score product strategy reduced the EER to as low as 0.44% and 0.28%, according to GPDS-ULPGC and PIE Face Databases, respectively. Our approach suggests a promising direction to improve the appearance-based face recognition performance in future human verification applications.

It is worth pointing out that we have not considered rotation in our experiments. In those scenes where faces are rotated, we could employ some rotation invariant face detector available in literature to detect and correct the rotated face. In this way the lips features could be made rotation invariant.

Acknowledgment

This work was supported by the Canary Government with funds from "Movilidad de Investigadores" in 2009.

References

[1] G. Goudelis, S. Zafeiriou, A. Tefas, I. Pitas, Class-specific Kernel-discriminant analysis for face verification, IEEE Transactions on Information Forensics and Security 2 (3) (2007) 570–587 Part 2.
[2] A. De la Cuesta, J. Zhang, P. Miller, Biometric identification using motion history images of a speaker's lip movements, in: Proceedings of the Machine Vision and Image Processing International Conference, 2008, pp. 83– 88.
[3] M.I. Faraj, J. Bigun: Motion features from lip movement for person authentication, in: Proceedings of the 18th International Conference on Pattern Recognition, 2006, 3, pp. 1059–1062.
[4] X. Yan, S. Guangda, Multi-parts and multi-feature fusion in face verification, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1– 6.
[5] P. Viola, M. Jones, Robust real-time object detection, International Journal of Computer Vision 57 (2) (2004) 137–154.
[6] C . Zhong, T. Tan, C. Xu, J. Li: Automatic 3D face recognition using discriminant common vectors, in: Proceedings of the LNCS Springer Conference on Advances in Biometrics, vol. 3832, 2005, pp. 85–91.
[7] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana, Discriminate common vectors for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (1) (2005) 4–13.
[8] H. Cevikalp, M. Neamtu, A. Barkana, The kernel common vector method: a novel nonlinear subspace classifier for pattern recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part B 37 (4) (2007) 937–951.
[9] C.M. Travieso, J.B. Alonso, M.A. Ferrer, Arbitrary Illumination Conditions for Facial Identification, in: Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology, 2007, pp. 93–98.
[10] R.C. González, R.E. Wood, Digital Image Processing, Prentice Hall, 2002.
[11] N. Otsu, A thresholding selection method from gray-level histogram, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66.
[12] T. Sim, S. Baker, M. Bsa, The CMU pose, illumination, and expression database, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1615–1618.

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 from 2001 in ULPGC, teaching subjects on signal processing and learning theory. His research lines are biometrics, data mining, classification system, environmental intelligence, signal and image processing, and environmental intelligence. He has researched in more than 23 International and Spanish Research Projects, some of them as head researcher. He has over 150 papers published in international journals and conferences. He has been reviewer in different international journals and conferences since 2001. He is Image Processing Technical IASTED Committee Member.

Jianguo Zhang (D.Phil. (2002), M.Sc. (1999), B.Sc. (1996)) is currently a Senior Lecturer of Visual Computation at University of Dundee. Previously, he worked as a lecturer with School of EEECS at Queen's University Belfast (2007–2010), a researcher with Department of Computer Science at Queen Mary University of London (2005–2007), the LEAR Group of INRIA Rhone-Alpes in France (2003–2005), Nanyang Technological University of Singapore (2002–2003). He received D.Phil. in National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2002, M.Sc. (1999) and B.Sc. (1996) from Shandong University of Technology (currently Shandong University), China. He won the Best Paper Award in International Machine Vision and Image Processing Conference 2008. 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 co-organizer of the 1st, 2nd and 3rd International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR09, 10, 11). He is also an editor of a book "Intelligent Video Event Analysis and Understanding" in Springer-Verlag series. He is an area chair of BMVC 2011. He served as PC members/referees for many international journals and top conferences, including IEEE TPAMI, IEEE TIP, IEEE TSMC, IVC, PR, CVIU, PRL, ICCV, CVPR, ECCV, and BMVC, etc. He is a senior member of IEEE (SMIEEE) and a Fellow of High Education of Academy UK (FHEA).

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 newly formed £30M funded Centre for Secure Information Technology. 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 Ph.D. 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 biomedical image analysis. He has published over 60 papers in image and video analysis, including a best paper award for his work on object recognition. During his academic career he has constantly worked in close collaboration with industry, including both multinational and SME companies.

Jesús B. Alonso 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. His research interests include signal processing in biocomputing, biometrics, nonlinear signal processing, recognition systems, audio characterization, and data mining. He is head of excellent network in biomedical engineering in ULPGC. He is Vice-Dean from 2009 in School of Engineering in Telecommunication and Electronic of ULPGC.

Miguel A. Ferrer received his M.Sc. degree in Telecommunications in 1988 and Ph.D. in 1994, both from the Universidad Politécnica de Madrid, Spain. He is an Associate Professor at Universidad de Las Palmas de Gran Canaria, where he has taught since 1990 and heads the Digital Signal Processing Group there. His research interests lie in the fields of biometrics and audio quality evaluation where he has published more than 100 papers. He is a member of the IEEE Carnahan Conference on Security Technology advisory Committee.