Research Article

Enhanced Asymmetric Bilinear Model for Face Recognition

Wenjuan Gong, 1 Weishan Zhang, 1 Jordi Gonzàlez, 2 Yan Ren, 1 and Zhen Li 3

1 China University of Petroleum, Qingdao 266580, China
2 Computer Vision Center, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain
3 Sichuan Water Conservancy Vocational College, Chongzhou, Sichuan 61231, China

Correspondence should be addressed to Wenjuan Gong; wenjuangong@upc.edu.cn

Received 30 November 2014; Revised 7 February 2015; Accepted 8 February 2015

Academic Editor: Houbing Song

Copyright © 2015 Wenjuan Gong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bilinear models have been successfully applied to separate two factors, for example, pose variances and different identities in face recognition problems. Asymmetric model is a type of bilinear model which models a system in the most concise way. But seldom there are works exploring the applications of asymmetric bilinear model on face recognition problem with illumination changes. In this work, we propose enhanced asymmetric model for illumination-robust face recognition. Instead of initializing the factor probabilities randomly, we initialize them with nearest neighbor method and optimize them for the test data. Above that, we update the factor model to be identified. We validate the proposed method on a designed data sample and extended Yale B dataset. The experiment results show that the enhanced asymmetric models give promising results and good recognition accuracies.

1. Introduction

Face recognition has been an active research area due to its applications in security entry systems and automatic law enforcement system. Pose variances, illumination changes, and noise from observations make it hard to define a model with promising recognition accuracies. Previous work [1–5] has proven that image variations of faces under varying illuminations could be modeled by linear spaces while view spaces which resulted from varying head poses are considered nonlinear [6]. Manifold learning [6, 7], bilinear models [8–10], linear dimensionality reduction [11, 12], and tensor analysis [13–15] have all been applied to solve face recognition problems. Recently deep learning techniques are also employed to solve face recognition problems and have achieved good performances [16–18].

Usually, faces need to be aligned before recognition. There has been lots of works which have been done on face alignment [19–30]. And it is well known that misalignment will result in recognition accuracy drop [20]. In our work, we explore the possibility of face recognition without alignment. In the proposed algorithm, we concentrate on separating factors that might affect recognition accuracy. To be specific, we use bilinear model to separate a varying factor from recognition target. And we prove in experiments that this separation is effective and required.

Bilinear models [8] are usually used to model systems with two factors and each factor itself is linear given the other factor fixed. Bilinear models for 2D image data have been widely used to solve face recognition problems due to their simplicity in formulation. In this category of solutions, factors introduced in a system are modeled as a symmetric bilinear model or asymmetric bilinear model. In this work, we also introduce asymmetric bilinear model due to its simplicity in representation.

Among all variance factors in face recognition problems, illumination change is one influential and intensively studied factor. Example works on illumination variations in face recognition problem include [9, 10], of which work in [10] is closely related to our work. Authors in [10] combine ridge regression into symmetric bilinear model to separate illumination and identity factor. Symmetric bilinear model decomposes input data into two factor matrices representing two factors and an interaction matrix denoting interactions between two factors. Instead of symmetric models, we utilize asymmetric models which decompose input data into a factor matrix and a matrix containing the other factor and interaction matrix. With fewer parameters to optimize,
the asymmetric model does not suffer from convergence problem as in symmetric model. But, due to noise captured in image features, the original asymmetric model is not able to cope with illumination changes in real-world face recognition problems.

In this work, we propose a modified asymmetric model for accurate face recognition. As in the original asymmetric bilinear model, we calculate probability predictions with Bayesian rules and expectation maximization method (EM). But instead of initializing the factor probabilities randomly, we initialize them with nearest neighbor method and optimize them for the test data. And we update the factor model to be identified. The enhanced asymmetric model is composed of three modules: calculating factor matrices from training data, initialization step for test data, and optimization and classification step for test data. In the first step, we use singular value decomposition to separate the factor to be identified from other factors based on training data. In the initialization step, a nearest neighbor method is introduced to initialize the joint probabilities of all factors which is further updated in optimization step. In the third step, we update the factor model to be identified. We validate the proposed method on a manually created data and cropped images from extended Yale B dataset. The experiment results show that the enhanced asymmetric models give promising results and good recognition accuracies.

The rest of this paper is organized as follows: in Section 2 we explain the proposed method in detail, including how to use Bayesian rules and EM method to optimize test factors and how to achieve illumination-robust face recognition and identity-robust illumination recognition goal with the proposed method; in Section 3, we validate the proposed method on two types of data: designed data and public available extended Yale B dataset; in Section 4, we will conclude our work and discuss possible future works.

2. Enhanced Asymmetric Bilinear Model

Bilinear models [8] separate the input data into two factors, described as style and content. Content is the factor to be identified; for example, identity and style are the factors that vary and need to be separated. Bilinear models reduce to the linear model when one factor is fixed. It contains symmetric and asymmetric models. Authors in [10] use symmetric model and take the identity as content and the illumination as style. In a symmetric bilinear model, style and content interact using an interaction matrix. But actually we are not interested in both style and content matrices in a single task classification problem. For simplicity, we choose an asymmetric model and modify it to accurately recognize factors.

Suppose we represent content \( c \) with a vector of parameters \( b^c \) of dimension \( J \) and use \( y^c \) to denote \( K \) dimensional input feature vector of style \( s \) and content \( c \); an asymmetric bilinear model represents the observation vector \( y \) as

\[
y_k^c = \sum_j a_{jk}^c b_j^s.
\]

This formulation can also be written as

\[
y^c = A^c b^c,
\]

where \( A^c \) denotes \( K \times J \) matrix with entries \( a_{jk}^c \) and

\[
a_{jk}^c = \sum_i a_{ij}^c a_i^c.
\]

In a symmetric model, \( \omega \) and \( a \) in (3) are represented separately thus resulting in one more variable.

As in [8], we use singular value decomposition to fit asymmetric model to training data. First we stack \( SC \times K \)-dimensional input features as follows:

\[
Y = \begin{bmatrix} y^1 \cdots y^K \\ \vdots \ddots \vdots \\ y^S \cdots y^{SK} \end{bmatrix};
\]

then we apply singular value decomposition to the observation:

\[
Y = USV^T.
\]

The style matrix \( A \) is initialized as the first \( J \) columns of \( US \) and the content matrix \( B \) is initialized as the first \( J \) rows of \( V^T \).

2.1. Optimizing Factor Matrix for Test Data. Given test data \( \tilde{Y} \) of an unseen style \( \tilde{s} \), the goal is to find the matrix containing style factor \( A^\tilde{s} \) and content factor \( b^\tilde{c} \) approximating test data as precious as possible, that is, to minimize the approximation error \( E \),

\[
E = \| \tilde{Y} - A^\tilde{s} b^\tilde{c} \|.
\]

In original bilinear model, optimization is under several assumptions: style matrix of the test data \( b^\tilde{c} \) equals style matrix of the training data \( B^\tilde{s} \); the probability of test observation of a specific style and a specific content does not need special way of initialization. Due to captured noise in input feature, the problem in reality is not precisely a linear model. So we add initialization step and update content matrix in optimization. According to experimental results, updating the content matrix \( b^\tilde{c} \) instead of keeping it unchanged gives more accurate recognition results.

2.1.1. Bayesian Rules for Probability Predictions. We define the probability of a captured feature given a style and a content assignment \( p(\bar{y} | \bar{s}, \bar{c}) \) as a Gaussian distribution; that is,

\[
p(\bar{y} | \bar{s}, \bar{c}) = \exp \left\{ -\frac{1}{2} \| \bar{y} - A^\bar{s} b^\bar{c} \|^2 \right\}.
\]

And we use Bayesian rules to calculate probabilities \( p(\tilde{s}, \tilde{c} | \tilde{y}) \), which is later on utilized to update matrix \( A^\tilde{s} \):

\[
p(\tilde{s}, \tilde{c} | \tilde{y}) = \frac{p(\tilde{y} | \tilde{s}, \tilde{c}) p(\tilde{y} | \tilde{s})}{\tilde{y}}.
\]

This variable is introduced as a weight in optimization step. Instead of the random initialization, we set the initial \( p(\tilde{y} | \tilde{s}, \tilde{c}) \) as 1 if \( \tilde{y} \) has a nearest neighbor \( y_n \) with a style \( \tilde{s} \) and a content \( \tilde{c} \) and 0 otherwise.
2.1.2. The Algorithm. To minimize the error between the test features and its approximation $A^\tilde{s}_b\tilde{c}$, we use an EM and add an extra update of the matrix containing content $\tilde{b}$ which proves to give more accurate recognition results.

1. Initialize the matrix containing style factors $A^\tilde{s}$ as being equal to that computed from training data $A$, and initialize $p(\tilde{y} | \tilde{s}, \tilde{c})$ as 1 if $\tilde{y}$ has a nearest neighbor $y_n$ with a style $\tilde{s}$ and a content $\tilde{c}$, and otherwise 0.
2. Expectation step: calculate $p(\tilde{s}, \tilde{c} | \tilde{y})$ with (8).
3. Maximization step: to maximize the total log-likelihood of the test data $L^* = \sum_y \log p(y)$, calculate the gradient of $L^*$ with respect to the matrix containing style, that is, solving $\frac{\partial L^*}{\partial A^\tilde{s}} = 0$. And we get $A^\tilde{s} = \left(\sum_c m^\tilde{c} b^\tilde{c}^T\right)\left(\sum_y n^\tilde{s} \tilde{b} \tilde{c}^T\right)^{-1}$, where $m^\tilde{c} = \sum_y p(\tilde{s}, \tilde{c} | \tilde{y})y$ and $n^\tilde{c} = \sum_y p(\tilde{s}, \tilde{c} | \tilde{y})$.
4. Update content matrix $\tilde{B} = (Y^T A^\tilde{s})/(AA^\tilde{s})$.

2.1.3. Illumination-Robust Face Recognition and Identity-Robust Illumination Recognition. In this section, we explain the possible applications of the proposed method in face recognition problems. First, we define identity as content and illumination as style to show the application of linear models to illumination-robust face recognition problem; see Figure 1. Then we switch content and style and set the illumination as content and identity as style. It is widely acknowledged that the illumination changes could be modeled with linear space but it is not clear if the identity space is also the same case. We design the second experiment to show the application of the proposed method in identity-robust illumination recognition and to test the expansibility of the system. In the following section, we show experimental results of these two experiment settings.

3. Experimental Results

We validate the proposed method with two categories of data: one is a manually defined sample data and another is a publicly available face recognition dataset. The aim of the designed sample data is to verify the soundness of the proposed method and the second category of experiments is to provide standard results.

3.1. Experiment on Designed Data. Suppose we have a tensor $\mathcal{A}$ of size $4 \times 4 \times 11$ with the first dimension representing content, the second style, and the third feature dimensions. Entries of the tensor are defined as

$$\mathcal{Y}(1,:,:) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix};$$
Table 1: Recognition accuracies and approximation errors of the enhanced asymmetric model on the designed data sample.

| Iteration number | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Accuracies       | 0.25  | 1.00  | 1.00  | 1.00  | 1.00  | 1.00  | 1.00  | 1.00  | 0.75  |
| Error            | 100000| 5.20  | 5.62  | 5.78  | 6.17  | 3.33  | 3.66  | 5.18  | 7.16  |

Table 2: Recognition accuracies and approximation errors of the enhanced asymmetric model on illumination-robust face recognition experiment.

| Iteration number | 1     | 2     | 3     | 4     | 5     |
|------------------|-------|-------|-------|-------|-------|
| Accuracies       | 0.13  | 0.56  | 0.92  | 0.92  | 0.74  |
| Error (divided by \(1.0e+12\)) | 0.38  | 0.0031| 0.0023| 0.0024| 0.0025|

\[
Y(2,:,:) = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0
\end{pmatrix};
\]

\[
Y(3,:,:) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1
\end{pmatrix};
\]  

(9)

With the proposed method, we get the experimental results in Table 1. The tables show recognition accuracies and approximation errors in each iteration. From the table, we can see that the proposed method can reach 100 percent recognition accuracy on the designed dataset.

3.2. Experiment on Extended Yale B Dataset. The extended Yale Face Database B contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions and they provide a subset of cropped images. In our experiment we use the cropped images. With this dataset, we design two sets of experiments: illumination-robust face recognition and identity-robust illumination recognition.

In illumination-robust face recognition, we pick those illuminations with id ranging from 1 to 39 as training and illuminations with id 50 for test. The experimental results for this experiment setting are shown in Table 2. From the table, we can see that after 3 iterations the recognition accuracy reaches the maximum 0.92 and then starts decreasing dramatically.

In identity-robust illumination recognition, we pick those persons with id ranging from 1 to 38 as training and person with id 39 as test. The experimental results for this experiment setting are shown in Table 3. From the table, we can see that although with similar amount of training data, illumination recognition gives better results than identity recognition. This might be due to sufficient training data (38) for the identity space or illumination to be an easier target. From Tables 2 and 3, we can say that the enhanced asymmetric model gives high recognition accuracies. Recognition accuracies increase quite fast with approximation error decreasing and with several steps, we can achieve the optimum recognition accuracies for our method.

We further show a set of experiments exploring the effect of training sample numbers. In illumination-robust identity recognition application, we pick the same test data as the previous experiment but we pick the illumination with identity ranging from 1 to 5 as training in experiment 1 (Exp1), illuminations with identity ranging from 1 to 9 as training in experiment 2 (Exp2), and illuminations with identity ranging from 1 to 13 as training in experiment 3 (Exp3). Experimental results are shown in Figure 2.

From the figure, we can see that, in the term of errors, the training sample numbers differences have an obvious effect in the first and the second iterations but tend to be subtle in the following iterations. In fact, with more training...
sample numbers, the optimization tends to be unstable. The best accuracy 92.31% occurs in all experiment settings. In the third iteration, experiment 1 and experiment 2 reach the maximum while experiment 3 reaches the maximum in iteration 6.

The proposed algorithm is further compared with a nearest neighbor method to show the importance of separating style and content. We design the nearest neighbor method applied in identity recognition problem: first, varying numbers of training data are selected (i.e., 5, 9, and 13); then, for each test data, it is compared with all styles (illumination variations in this case) from all contents (face identity in this case); finally, the label of the test data is set as the label of the training data with the minimum distance. The recognition accuracy of the nearest neighbor method is shown in Table 4. From the table, we can see that the raw nearest neighbor method performs very badly on the identity recognition problem here. What is more, when the number of training data increases, the recognition accuracy drops. Note that each training data case has a different style compared to other training data in the dataset. We can see from the recognition accuracies that more training data from variant styles confuses the method. Adding different styles of training data without separating them makes the recognition accuracies worse.

| Iteration number | 1         | 2          | 3          | 4          |
|------------------|-----------|------------|------------|------------|
| Accuracies       | 0.88      | 1.0000     | 0.9219     | 0.3906     |
| Error (divided by 1.0e + 11) | 5.5517 | 0.1521     | 0.1679     | 0.2006     |

**Table 3: Recognition accuracies and approximation errors of the enhanced asymmetric model on identity-robust illumination recognition experiment.**

**Table 4: Recognition accuracies of the nearest neighbor method on identity recognition experiment.**

| Training case number | 5 | 9 | 13 |
|----------------------|---|---|---|
| Accuracies           | 0.0526 | 0.0526 | 0.0263 |

**4. Conclusions and Future Works**

In this paper, we present an enhanced asymmetric bilinear model and apply it to illumination-robust face recognition problem and identity-robust illumination recognition problems. By initializing the content and style probability with nearest neighbor method and adding update for content matrix, we achieve quite good recognition accuracies. In the future, it would be interesting to explore the application of the proposed method on models containing more than two factors. Also, we would like to explore applications of asymmetric models on applications other than face recognition.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

Authors acknowledge the support of the Fundamental Research Funds for the Central Universities HCUOX0214A, start-up funds for introduced talent at China University of Petroleum 201410583, and the Spanish Projects TIN2009-14501-C02-02 and TIN2012-39051.

**References**

[1] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684–698, 2005.

[2] A. Shashua, "On photometric issues in 3D visual recognition from a single 2d image," International Journal of Computer Vision, vol. 21, no. 1-2, pp. 99–122, 1997.

[3] L. Zhao and Y.-H. Yang, "Theoretical analysis of illumination in PCA-based vision systems," Pattern Recognition, vol. 32, no. 4, pp. 547–564, 1999.

[4] P. W. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 995–999, June 1994.

[5] S. K. Nayar and H. Murase, "Dimensionality of illumination in appearance matching," in Proceedings of the 13th IEEE International Conference on Robotics and Automation, pp. 1326–1332, April 1996.

[6] X. Gao and C. Tian, "Multi-view face recognition based on tensor subspace analysis and view manifold modeling," Neurocomputing, vol. 72, no. 16-18, pp. 3742–3750, 2009.

[7] A. Elgammal and C. S. Lee, "Homeomorphic manifold analysis (HMA): generalized separation of style and content on manifolds," Image and Vision Computing, vol. 31, no. 4, pp. 291–310, 2013.

[8] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," Neural Computation, vol. 12, no. 6, pp. 1247–1283, 2000.

[9] B. Jun, J. Lee, and D. Kim, "A novel illumination-robust face recognition using statistical and non-statistical method," Pattern Recognition Letters, vol. 32, no. 2, pp. 329–336, 2011.

[10] D. Shin, H.-S. Lee, and D. Kim, "Illumination-robust face recognition using ridge regressive bilinear models," Pattern Recognition Letters, vol. 29, no. 1, pp. 49–58, 2008.

[11] S. Ioffe, "Probabilistic linear discriminant analysis," in Computer Vision—ECCV 2006, vol. 3954 of Lecture Notes in Computer Science, pp. 531–542, Springer, Berlin, Germany, 2006.

[12] M. Kan, S. Shan, D. Xu, and X. Chen, "Side-information based linear discriminant analysis for face recognition," in Proceedings of the 22nd British Machine Vision Conference (BMVC’11), 125.0, p. 125.1, September 2011.

[13] C. Bauckhage, "Robust tensor classifiers for color object recognition," in Image Analysis and Recognition, vol. 4633 of Lecture Notes in Computer Science, 2007.
[14] K. Jia and S. Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition," in Proceedings of the 10th IEEE International Conference on Computer Vision, pp. 1683–1690, October 2005.

[15] S. Rana, W. Liu, M. Lazarescu, and S. Venkatesh, "A unified tensor framework for face recognition," Pattern Recognition, vol. 42, no. 11, pp. 2850–2862, 2009.

[16] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’12), pp. 2518–2525, June 2012.

[17] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’14), pp. 1891–1898, Columbus, Ohio, USA, June 2014.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: closing the gap to human-level performance in face verification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’14), pp. 1701–1708, IEEE, Columbus, Ohio, USA, June 2014.

[19] S. Yan, C. Liu, S. Z. Li, H. Zhang, H.-Y. Shum, and Q. Cheng, "Face alignment using texture-constrained active shape models," Image and Vision Computing, vol. 21, no. 1, pp. 69–75, 2003.

[20] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of Mis-alignment in face recognition: problem and a novel Mis-alignment learning solution," in Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 314–320, May 2004.

[21] L. Zhang, H. Ai, S. Xin, C. Huang, S. Tsukiji, and S. Lao, "Robust face alignment based on local texture classifiers," in Proceedings of the IEEE International Conference on Image Processing (ICIP ’05), vol. 2, pp. 354–357, September 2005.

[22] L. Liang, R. Xiao, F. Wen, and J. Sun, "Face alignment via component-based discriminative search," in Computer Vision—ECCV 2008, vol. 5303 of Lecture Notes in Computer Science, pp. 72–85, Springer, Berlin, Germany, 2008.

[23] H. Wu, X. Liu, and G. Doretto, "Face alignment via boosted ranking model," in Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’08), pp. 1–8, Anchorage, Alaska, USA, June 2008.

[24] J. M. Saragih, S. Lucey, and J. F. Cohn, "Face alignment through subspace constrained mean-shifts," in Proceedings of the 12th International Conference on Computer Vision (ICCV ’09), pp. 1034–1041, October 2009.

[25] T. Berg and P. N. Belhumeur, "Tom-vs-Pete classifiers and identity-preserving alignment for face verification," in Proceedings of the 23rd British Machine Vision Conference (BMVC ’12), pp. 1–11, September 2012.

[26] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: robust alignment and illumination by sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 2, pp. 372–386, 2012.

[27] X. Shen, Z. Lin, J. Brandt, and Y. Wu, "Detecting and aligning faces by image retrieval," in Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’13), pp. 3460–3467, June 2013.

[28] X. Xiong and F. de la Torre, "Supervised descent method and its applications to face alignment," in Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’13), pp. 532–539, June 2013.

[29] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," International Journal of Computer Vision, vol. 107, no. 2, pp. 177–190, 2014.

[30] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’14), pp. 1867–1874, Columbus, Ohio, USA, June 2014.
