PATCH-WISE LOW-DIMENSIONAL PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS FOR FACE RECOGNITION

Vitomir Štruc, Nikola Pavešić[*]

Faculty of Electrical Engineering UL
Tržaška cesta 25, 1000 Ljubljana, Slovenia

Jerneja Žganec-Gros, Boštjan Vesnicer

Alpineon Ltd., Ulica Iga Grudna 15
1000 Ljubljana, Slovenia

ABSTRACT

The paper introduces a novel approach to face recognition based on the recently proposed low-dimensional probabilistic linear discriminant analysis (LD-PLDA). The proposed approach is specifically designed for complex recognition tasks, where highly nonlinear face variations are typically encountered. Such data variations are commonly induced by changes in the external illumination conditions, viewpoint changes or expression variations and represent quite a challenge even for state-of-the-art techniques, such as LD-PLDA. To overcome this problem, we propose here a patch-wise form of the LD-PLDA technique (i.e., PLD-PLDA), which relies on local image patches rather than the entire image to make inferences about the identity of the input images. The basic idea here is to decompose the complex face recognition problem into simpler problems, for which the linear nature of the LD-PLDA technique may be better suited. By doing so, several similarity scores are derived from one facial image, which are combined at the final stage using a simple sum-rule fusion scheme to arrive at a single score that can be employed for identity inference. We evaluate the proposed technique on experiment 4 of the Face Recognition Grand Challenge (FRGCv2) database with highly promising results.

Index Terms— Biometrics, face recognition, pattern recognition, probabilistic linear discriminant analysis

1. INTRODUCTION

Face recognition technology is becoming increasingly important in various application domains ranging from forensic investigations, where large amounts of surveillance data need to be inspected for known subjects, to ambient intelligence applications, where a small group of people needs to be recognized consistently from the recorded video footage. Regardless of the application domain, face recognition technology has to meet strict performance criteria even if the input data is of questionable quality and exhibits unseen characteristics. Such conditions are typically encountered in uncontrolled conditions, where illumination, pose, expression changes and alike induce highly nonlinear data variations and make it difficult to rely on linear techniques, which by today are still very present the field of face recognition (see, for example, [1], [2], [3], [4], [5]).

In this paper we introduce a novel technique that is specifically designed to tackle face recognition tasks, where such nonlinear data variations are present. The novel technique builds upon a variant of the recently proposed probabilistic linear discriminant analysis (PLDA) [3], [4], called low-dimensional (or simplified) probabilistic linear discriminant analysis (LD-PLDA) [6], [7]. Instead of applying the LD-PLDA technique on entire face images, we propose here to use a patch-wise approach and subject spatially local image patches to the LD-PLDA technique. By doing so, the proposed approach, named *patch-wise low-dimensional probabilistic linear discriminant analysis*, tackles the complex classification problem by constructing a number of simpler classification problems for which the linear assumption may be more appropriate. The results of the locally applied LD-PLDA technique is combined in the end using a simple sum-rule fusion scheme. The proposed approach is assessed in face recognition experiments on the Face Recognition Grand Challenge (FRGC) database [8], where highly encouraging results are obtained.

The rest of the paper is structured as follows. In Section 2 we survey previous work and elaborate on the relation to the technique introduced here. We propose the novel patch-wise LD-PLDA technique in Section 3 and evaluate it in face verification experiments in Section 4. We conclude the paper with some final comments in Section 5.

2. RELATION TO PRIOR WORK

Early research on face recognition was dominated by so-called appearance-based methods, such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2] and alike (e.g., [9], [10]). These methods represent facial images in the form of subspace projection coefficients and conduct recognition based on distance measurements

[*]The work presented in this paper was supported in parts by the national research program P2-0250(C) Metrology and Biometric Systems, the post-doctoral project BAMBI (ARRS ID Z2-4214) and by the European Union, European Regional Fund, in scope of the framework of the Operational Programme for Strengthening Regional Development Potentials for the Period 2007-2013, contract No. 3211-10-000467 (KC Class).

between probe and target subspace coefficient vectors.

It soon became evident that the linear nature of the appearance-based methods is only suitable for face recognition in constrained environments, but is too restrictive for complex face recognition tasks typically encountered in uncontrolled conditions. Thus, non-linear extensions of appearance based methods, better suited for deployment in such conditions, started appearing in the literature. Two different strategies emerged for extending the linear techniques. The first focuses on kernel methods and is exemplified best by the non-linear (kernel) extensions of PCA or LDA, called kernel principal component analysis (KPCA) [11] and kernel Fisher analysis (KFA) [12], while the second, which is of relevance to this paper, tries to extend the linear techniques to a non-linear form with piecewise (or locally) linear models. Examples of techniques adopting this strategy are locally linear embedding (LLE) [13], [14], locally linear regression (LLR) [15] or locally linear discriminant analysis (LLDA) [16] among others.

The technique proposed in this paper builds upon the ideas of the locally linear techniques and applies them to a probabilistic framework. Specifically, the proposed patch-wise LD-PLDA technique relies on a special form of PLDA, which was originally introduced for the problem of speaker verification [7], [17], [18] and was only recently applied to the field of face recognition [6]. It has to be noted that probabilistic techniques, such as the ones proposed in [3], [4], [19], [20] or [21] have demonstrated state-of-the-art performance for complex face recognition tasks over the years. However, despite their success they have received far less attention from the scientific community than the appearance based methods described above.

3. PATCH-WISE LOW-DIMENSIONAL PLDA

3.1. On PLDA and its Variants

Probabilistic Linear Discriminant Analysis (PLDA), first introduced to the field of face recognition in [3] and [4], represents a probabilistic extension of Linear Discriminant Analysis (LDA) [2]. The technique decomposes a given feature vector into an identity-specific part that describes the class-membership of the feature vector and a so-called channel-specific part that accounts for the sample (or intra-class) variability of the feature vector. Identity inference is performed using the estimated identity-specific part of the feature vector, while the channel-specific part is used to describe non-identity-related variability. While the original formulation of PLDA [4] is relatively complex, recent research (related mainly to the field of speaker recognition, e.g., [7], [17], [18], [22]) has shown that under certain assumptions equally good and in some cases even superior recognition results can be achieved with a simpler PLDA model. However, this simpler model is applicable only if the feature vectors used are of a *sufficiently low dimension*. We will refer to this form of PLDA as *low-dimensional PLDA* and denote it with LD-PLDA in the remainder of the paper.

The LD-PLDA technique can formally be described as follows: let $\{\eta_r : r = 1,..,R\}$ denote a collection of feature vectors extracted from a set of biometric samples (i.e., face images) of a particular individual. Then LD-PLDA decomposes each feature vector into the following form:

$$\eta_r = m + \Phi\beta + \varepsilon_r, \tag{1}$$

where m denotes a global offset, representing the average feature vector, Φ provides the basis for the identity-specific subspace, β represents a latent identity vector having a standard normal distribution, and ε_r denotes a sample-dependant residual term, which is assumed to be normally distributed with a mean of zero and a full covariance matrix Σ. It has to be noted at this point that the parameters of the LD-PLDA model $\{m, \Phi, \Sigma\}$ are not determined analytically as with LDA. Instead, they are learned from some development data via an EM algorithm [17]. Once the LD-PLDA model parameters are known, inferences about the identity of a given feature vector η_r can be made based on the hidden identity variable β.

For the verification problem, for example, we need to determine which of the two following hypotheses is more likely:

- $\mathcal{H}_s$: the two given feature vectors η_1 and η_2 share the same identity variable β, or
- $\mathcal{H}_d$: the two given feature vectors η_1 and η_2 were generated by two different identity variables β_1 and β_2.

The hypotheses test can be performed based on the following log-likelihood ratio:

$$s_{lr} = \log \frac{p(\eta_1, \eta_2|\mathcal{H}_s)}{p(\eta_1|\mathcal{H}_d)p(\eta_2|\mathcal{H}_d)}, \tag{2}$$

which can be computed in closed form for the LD-PLDA model if the global offset m is removed from all low-dimensional feature vectors used in the calculations, i.e.:

$$s_{lr} = \eta_1^T Q \eta_1 + \eta_2^T Q \eta_2 + 2\eta_1^T P \eta_2. \tag{3}$$

Here, Q and P are defined as

$$Q = \Sigma_{tot}^{-1} - (\Sigma_{tot} - \Sigma_{ac}\Sigma_{tot}^{-1}\Sigma_{ac}) \text{ and} \tag{4}$$

$$P = \Sigma_{tot}^{-1}\Sigma_{ac}(\Sigma_{tot} - \Sigma_{ac}\Sigma_{tot}^{-1}\Sigma_{ac}), \tag{5}$$

where $\Sigma_{tot} = \Phi\Phi^T + \Sigma$ and $\Sigma_{ac} = \Phi\Phi^T$ [18]. As we can see from the above derivations, the identity variable is never computed explicitly.

3.2. The Patch-wise Approach

It was shown in [6] that the LD-PLDA technique represents a viable solution for face recognition, even though it was originally developed in the field of speaker verification. Furthermore, it was demonstrated that the LD-PLDA model overcomes some shortcomings of the original PLDA technique

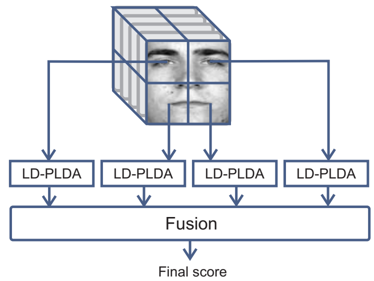

Fig. 1. Illustration of the proposed patch-wise approach.

and often results in superior recognition performance. To ensure a sufficiently low dimensionality of the feature vectors the facial images were projected onto a subspace computed with the Fisherface approach [2] and further processed the subspace feature vectors using within-class covariance normalization (WCCN) [23].

Here, we build upon the same setup, but take it a step further. Face image variations, induced by illumination-, pose-, or expression-changes among others, typically result in complex and highly nonlinear data distributions [24], which make it difficult to effectively discriminate between different identities using linear models. To address this problem, we propose in this section a patch-wise version of LD-PLDA (denoted as PLD-PLDA hereafter), where the LD-PLDA technique is applied to local image patches, instead of the entire image. By doing so, our complex face recognition problem is decomposed into N simpler problems (where N is the number of image patches used), for which linear models may be more suitable. A hypotheses test is conducted for each patch, resulting in a number of similarity scores (i.e., log-likelihood ratios - see Eq. 3) that are combined in the last step using a selected fusion scheme. Note that in this paper we use a simple (equally weighted) sum-rule fusion scheme to combine the individual similarity scores. The adopted fusion scheme serves only as a proof of concept and may be replaced by a more elaborate fusion scheme in the future. The entire procedure is illustrated in Fig. 1, while the training and verification stages presented in more detail in Algorithms 1 and 2.

TRAINING ALGORITHM

Data: Training or development images
Result: LD-PLDA model parameters for all N patches

for *i-th image patch of all training images (i=1:N)* **do**
  $\rightarrow$ compute low-dimensional feature vectors $\eta^{(i)}$ for all training images using LDA and WCCN;
  $\rightarrow$ estimate LD-PLDA model parameters $\{m^{(i)}, \Phi^{(i)}, \Sigma^{(i)}\}$ for the i-th patch;
  $\rightarrow$ save results;
end

Algorithm 1: PLD-PLDA training

VERIFICATION

Data: Pair of images X_1 and X_2
Result: PLD-PLDA similarity score

$\rightarrow$ set output score s_{fin} to zero;
for *i-th image patch of both images (i=1:N)* **do**
  $\rightarrow$ compute low-dimensional feature vectors $\eta_1^{(i)}$ and $\eta_2^{(i)}$ from input data using LDA and WCCN;
  $\rightarrow$ load the $i-th$ LD-PLDA model parameters $\{m^{(i)}, \Phi^{(i)}, \Sigma^{(i)}\}$;
  $\rightarrow$ compute similarity score (Eq. (3));
  $\rightarrow$ add computed similarity score to s_{fin};
end
$\rightarrow$ verify identity based on s_{fin};

Algorithm 2: Verification procedure

4. EXPERIMENTS

For our experiments we make use of the second version of the Face Recognition Grand Challenge (FRGCv2) dataset [8], which is a large scale dataset of facial images featuring more than 40 000 images of 466 subjects. We select the most challenging of the experiments defined by the experimental protocol of the database for our assessments, namely, experiment 4, which defines a target (or gallery) set of 16026 facial images, captured in controlled conditions, and a probe set of 8014 facial images, captured in uncontrolled conditions. Experiment 4 also defines a development/training set of 12776 (controlled and uncontrolled) images that can be used for training (in our case for learning the PLD-PLDA model parameters). Prior to the experiments we subject all images to a preprocessing procedure that crops the facial region based on annotated eye-center coordinates, geometrically normalizes the facial region, rescales it to a size of 128×128 pixels.

As suggested by the experimental protocol of the FRGCv2 database, we present our results in the form of Receiver Characteristic Curves (ROC) and verification rates at specific false acceptance rates. Specifically, we provide results in terms of the verification rate at the false accept rate of 0.1% (VER@0.1%FAR), the verification rate at the false accept rate of 1% (VER@1%FAR) and the value of the equal error rate (EER). When generating ROC curves we use a subset of all similarity scores computed in the scope of Experiment 4. This subset results in performance curves denoted as ROC III curves by the experimental protocol of the FRGCv2 dataset.

We use fixed hyper-parameters for the Fisherface and WCCV techniques for all experiments regardless of the image (or patch) dimensionality. Hence, we use 600 PCA eigenvectors for dimensionality reduction and the LDA subspace dimensionality of 200 within the Fisherface technique.

In our first series of experiments we evaluate the feasibility of the low-dimensional PLDA technique for face recognition. Since the technique was originally developed for the

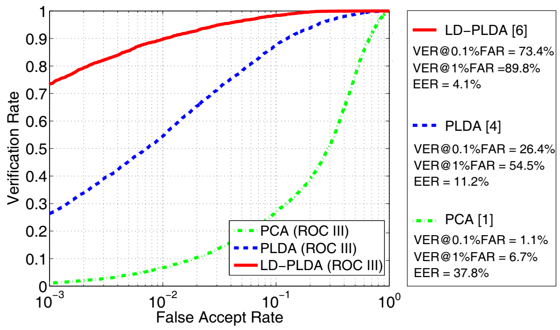

Fig. 2. Results of the first series of experiments (ROC III)

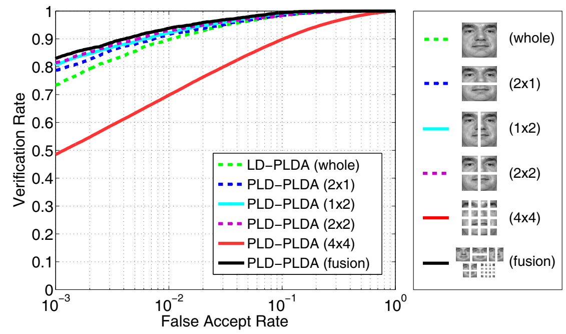

Fig. 3. Results of the second series of experiments (ROC III)

task of speaker verification, we need to established that it is suitable for face recognition as well. To this end, we implement the LD-PLDA [6] and compare it to the original PLDA technique from [3], [4][1] as well as our own implementation of the baseline PCA technique [1], where we use a subspace dimensionality of 600. While the probabilistic methods conduct verification based on Eq. (3), we use the cosine similarity measure with the PCA technique. The results of this series of recognition experiments are presented in Fig. 2 in the form of ROC curves and values of several performance metric. Note that the PCA and PLDA techniques perform significantly worse than the LD-PLDA technique, which suggests that LD-PLDA is not only suitable for face recognition, but also results in performance gains over PCA and PLDA. Thus, we use LD-PLDA for our following comparative assessments.

In our second series of recognition experiments we assess the performance of the proposed PLD-PLDA technique and compare it to the LD-PLDA technique. To evaluate the impact of the patch size on the performance of PLD-PLDA we implement the technique with different sizes of the image patches. Specifically, we implement the PLD-PLDA technique by partitioning the images into: 2×1 patches, 1×2 patches, 2×2 patches, and 4×4 patches. An illustration of the partitioning schemes used is also shown on the right hand side of Fig. 3. Note that in addition to the presented experiments we also conduct an experiment, where we combine the similarity scores of all versions of PLD-PLDA implemented in this series of experiments using a equally-weighted sum-fusion rule. This experiment, which is denoted as "PLD-PLDA (fusion)" in Fig. 3 and Table 1, assesses whether patches of different sizes carry complementary information that can be exploited for recognition. From the results presented in Fig. 3 and Table 1 we can see that applying the LD-PLDA technique to smaller image patches has indeed a positive effect on the recognition performance. This is most likely the consequence of a better fit of the linear model and the fact that smaller image patches result in an improved ratio between the amount of available training data and the data's dimensionality. This, in turn, results in better estimation of the LD-PLDA model parameters and consequently in better recognition performance. However, as shown by the results of the PLD-PLDA technique implemented with 16 image patches, there is a limit on how small the image patches can be. Once the patches become too small, the technique exhibits a drop in performance.

[1] We use the code of the authors of the PLDA technique for our experiments: `http://web4.cs.ucl.ac.uk/research/vis/pvl/images/stories/FaceRecognition/pldacode.zip`

Table 1. Performance metrics of the experiments (ROC III)

Method	VER@0.1%FAR	VER@1%FAR	EER
LD-PLDA	0.734	0.898	0.378
PLD-PLDA (2x1)	0.787	0.916	0.039
PLD-PLDA (1x2)	0.807	0.922	0.036
PLD-PLDA (2x2)	0.814	0.929	0.035
PLD-PLDA (4x4)	0.485	0.698	0.101
PLD-PLDA (fusion)	0.828	0.939	0.031

Table 2. VER@0.1%FAR for the state-of-the-art (ROC III)

Method (ROC III)	VER	Method (ROC III)	VER
PLD-PLDA (fusion)	0.828	MLPQ+FS [25]	0.795
GaborKFA [26]	0.760	RRGb+L3 [27]	0.790
LBP [28]	0.735	MBC-F [29]	0.839
Gabor [28]	0.735	rLDA+Color [30]	0.782
Gabor+LBP [28]	0.836	BBE baseline (NIST)	0.120

Last but not least we compare the performance of the best performing technique from the previous series of experiments, i.e., PLD-PLDA (fusion), to the performance of state-of-the-art techniques from the literature (see Table 2). Note that the proposed technique is among the top performers. Since we have not used any photometric normalization, color information, local descriptors and alike, we believe that the performance of our technique could be further improved.

5. CONCLUSION

We have presented a novel approach to face recognition, which applies the LD-PLDA technique to local image patches instead of the entire image. We have shown that the proposed approach ensures state-of-the-art results on the FRGC dataset.

6. REFERENCES

[1] M. Turk and A. S. Pentland, "Eigenfaces for Recognition," *J. of Cog. Neur.*, vol. 3, no. 1, pp. 71–86, 1991.

[2] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE TPAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[3] S.J.D. Prince and J.H. Elder, "Probabilistic Linear Discriminant Analysis for Inferences About Identity," in *IEEE ICCV*, 2007, pp. 1–8.

[4] P. Li, Y. Fu, U. Mohammed, J. Elder, and S.J.D. Prince, "Probabilistic models for inference about identity," *IEEE TPAMI*, vol. 34, no. 1, pp. 144–157, 2012.

[5] N. Poh, C.H. Chan, J. Kittler, S. Marcel, C. McCool, E.A. Rua, J.L.A. Castro, M. Villegas, R. Paredes, V. Štruc, N. Pavešić, A.A. Salah, H. Fang, and N. Costen, "An Evaluation of Video-to-video Face Verification," *IEEE TIFS*, vol. 5, no. 4, pp. 781–801, 2010.

[6] B. Vesnicer, , J. Žganec Gros, N. Pavešić, and V Štruc, "Face Recognition using Simplified Probabilistic Linear Discriminant Analysis," *International Journal of Advanced Robotic Systems*, vol. 9, pp. 1–10, 2012.

[7] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE TASLP*, vol. 19, no. 4, pp. 788–798, 2011.

[8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *CVPR*, pp. 947–954, 2005.

[9] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE TNN*, vol. 13, no. 6, pp. 1450–1464, 2000.

[10] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face Recognition Using LDA-Based Algorithms," *IEEE TNN*, vol. 14, no. 1, pp. 195–200, 2003.

[11] B. Scholkopf, A. Smola, and K.R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," in *Technical Report No. 44*, 1996, p. 18 pages.

[12] M.H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," in *AFGR*, 2002, pp. 215–220.

[13] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[14] N. Mekuz, C. Bauchhage, and J.K. Tsotsos, "Face recognition with weighted localy linear embedding," in *CCCRV*, 2005, pp. 290–296.

[15] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally Linear Regression for Pose-Invariant Face Recognition," *IEEE TIP*, vol. 16, no. 7, pp. 1716–1725, 2007.

[16] T.K. Kim and J. Kittler, "Locally Linear Discriminant Analysis for Multimodally Distibuted Classes for Face Recognition with a Single Model Image," *IEEE TPAMI*, vol. 27, no. 3, pp. 1–10, 2005.

[17] N. Brummer and E. de Villiers, "The speaker partitioning problem," in *Odyssey*, 2010.

[18] D. Garcia-Romero and C. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *Interspeech 2011*, 2011.

[19] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition," *Pattern Recognition*, vol. 33, pp. 1771–1782, 2000.

[20] S. Ioffe, "Probabilistic Linear Discriminant Analysis," in *ECCV*, 2006, pp. 531–542.

[21] S.J.D. Prince, J. Warrell, J.H. Elder, and F.M. Felisberti, "Tied Factor Analysis for Face Recognition Across Large Pose Differences," *IEEE TPAMI*, vol. 30, no. 6, pp. 970–984, 2008.

[22] M. Sennoussaoui, P. Kenny, N. Brummerand E. de Villiers, and P. Dumouchel, "Mixture of PLDA Models in I-Vector Space for Gender-Independent Speaker Recognition," in *Interspeech*, 2011.

[23] A. Hatch and A. Stolcke, "Generalized linear kernels for one-versus-all classification: application to speaker recognition," in *ICASSP*, 2006.

[24] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Boosting linear discriminant analysis for face recognition," in *IEEE ICIP*, 2003, pp. I–657–660.

[25] Z. Lei and S.Z. Li, "Fast multi-scale local phase quantization histogram for face recognition," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1761–1767, 2012.

[26] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE TPAMI*, vol. 28, no. 5, pp. 725–737, 2006.

[27] C. Liu, "Effective use of color information for large scale face verification," *Neurocomputing*, vol. 101, pp. 43–51, 2013.

[28] X. Tan nad B. Triggs, "Fusing Gabor and LBP Feature Sets for Kernel-based Face Recognition," in *AMFG '07*, 2007, pp. 235–249.

[29] M. Yang, L. Zhang, S.C.K. Shui, and D. Zhang, "Monogetic Binary Coding: An efficient Local Feature Extraction Approach to Face Recogntion," *IEEE TIFS*, vol. 7, no. 6, pp. 1738–1751, 2012.

[30] B.Y.L. Li, W. Liu, S. An, A. Krishna, and T. Xu, "Face recognition using various scales of discriminant color space transform," *Neurocomputing*, vol. 94, pp. 68–76, 2012.