REMOVING ILLUMINATION ARTIFACTS FROM FACE IMAGES USING THE NUISANCE ATTRIBUTE PROJECTION

Vitomir Štruc, Boštjan Vesnicer, France Mihelič, Nikola Pavešić

Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia

ABSTRACT

Illumination induced appearance changes represent one of the open challenges in automated face recognition systems still significantly influencing their performance. Several techniques have been presented in the literature to cope with this problem; however, a universal solution remains to be found. In this paper we present a novel normalization scheme based on the nuisance attribute projection (NAP), which tries to remove the effects of illumination by projecting away multiple dimensions of a low dimensional illumination subspace. The technique is assessed in face recognition experiments performed on the extended YaleB and XM2VTS databases. Comparative results with state-of-the-art techniques show the competitiveness of the proposed technique.

Index Terms— Image processing, biometrics, face recognition, pattern recognition, nuisance attribute projection.

1. INTRODUCTION

Current face recognition technology has evolved to a point where its performance allows for the deployment in a wide variety of applications. These applications typically ensure (or impose) controlled conditions for the acquisition of the facial images and, hence, try to minimize the intra-class variability in the appearance of the (facial) images. In these, controlled conditions state-of-the-art face recognition systems are capable of achieving high performance levels. However, the majority of the existing recognition techniques featured in these systems deteriorate in their performance when applied on facial data captured in uncontrolled environments [1].

In these settings the captured data exhibits appearance variations caused by pose-, expression- and illumination-changes, which pose challenging problems even to the most advanced face recognition techniques. Among the listed influential factors, illumination is often the most critical. In fact, it was empirically shown that the illumination induced variability in facial images is often larger than the variability induced to different images by the subject's identity [2], [3].

In this paper we present a novel normalization technique for illumination invariant face recognition based on the nuisance attribute projection (NAP), which removes the illumination induced artifacts from the facial images by projecting away multiple dimensions of a low dimensional illumination subspace. As will be shown in the paper, the technique exhibits several desirable properties and ensures recognition rates comparable to those achieved with state-of-the-art normalization techniques from the literature.

2. THE PROPOSED TECHNIQUE

This section commences by presenting the basic theory underlying the majority of photometric normalization techniques and proceeds by introducing the novel NAP-based normalization scheme.

2.1. The Luminance-Reflectance Model

The luminance-reflectance model (LR-model) represents a mathematical model explaining the process of image formation [4]. It states that a digital image $I(x, y)$ can be modeled as a product of the so-called luminance and reflectance functions or equally as the sum of these two functions in the logarithm domain:

$$\ln[I(x, y)] = \ln[R(x, y)] + \ln[L(x, y)]. \tag{1}$$

Here, the reflectance $\ln[R(x, y)]$ relates to the characteristics of the objects in the image and is dependant on the reflectivity (or albedo) of the objects' surfaces [5], while the luminance $\ln[L(x, y)]$ relates to the illumination source and is dependant on the amount of illumination falling on the observed scene.

Clearly, the reflectance is insensitive to the external illumination conditions and, thus, represents an illumination invariant representation of the given image. However, it is impossible to compute the reflectance from (1) unless some assumptions are made with respect to the characteristics of the LR-model. The most common assumption made are that the luminance varies slowly with the spatial position and that edges in the image also corresponds to edges in the reflectance [1], [5]. Based on these assumptions the luminance is estimated, for example, as a low-pass filtered version of the input image, while the reflectance is subsequently computed by manipulating the expression in (1).

Various techniques were already presented in the literature exploiting the LR-model. However, most of these techniques remove all frequency components of the facial image up to a selected cut-off frequency, regardless whether these components are affected by illumination changes or not. As a result of this procedure, only the high-frequency components

are retained in the images, while the potentially valuable (discriminative) information contained in the low-frequency part of the images is removed. This property of the photometric normalization techniques is reflected in the histograms of the normalized facial images, which usually exhibit a distinctive peak around one value. This peak indicates that most of the variability needed for modeling the intra-class variations intrinsic to a subject has been removed. Hence, on small datasets these techniques work well, while their scalability is at least questionable. Furthermore, as they make no distinction between illumination-induced and subject-specific low-frequency information, they commonly degrade the recognition performance on images captured in controlled conditions.

To tackle the above issues we present in this paper a novel technique, which uses the nuisance attribute projection (NAP) to remove illumination induced appearance changes from the facial images. As the NAP projection directions are learned from a specially selected generic dataset and encode solely low-frequency attributes related to illumination variations, the technique is expected to perform well on images captured in controlled and uncontrolled conditions alike.

2.2. The Nuisance Attribute Projection

The nuisance attribute projection (NAP) is a powerful technique regularly used in the field of speaker recognition for compensation of channel effects, which are the central problem in automated speaker recognition systems [6]. In one of its most commonly exploited variants the technique projects away multiple dimensions of a specifically designed subspace with the goal of reducing the channel induced variability. These dimensions (or better said the variability along these dimensions) are commonly assigned to some confusing (unwanted) attribute of the data, hence, the name nuisance attribute projection [6].

Consider a set of n images (in vector form) of size $N = a \times b$ pixels arranged into a $n \times N$ column data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$. The NAP technique tries to remove any (additive) distortion in the images as follows:

$$\mathbf{X}' = \mathbf{P}(\mathbf{X} - \mathbf{M}), \qquad (2)$$

where $\mathbf{X}'$ denotes the new data matrix with compensated 'channel' effects, $\mathbf{M}$ denotes a matrix of the same size as $\mathbf{X}$ containing in each of its columns the global mean of the images in $\mathbf{X}$ and $\mathbf{P}$ stands for the $n \times n$ projection matrix:

$$\mathbf{P} = \mathbf{I} - \sum_{i=1}^{d'} \mathbf{w}_i \mathbf{w}_i^T. \qquad (3)$$

Here, $\mathbf{I}$ denotes the $n \times n$ identity matrix, $\mathbf{w}_i$ represents the i-th NAP direction, d' stands for the number of NAP directions.

The obvious question arising when looking at Eq. (3) is how to determine the d' NAP directions needed for the compensation scheme. To answer this question let us assume that we have n_{C_j} sample images from the j-th class, where $C_1, C_2, ..., C_r$ denote the class labels of the images in $\mathbf{X}$.

For each of these images we can write:

$$\mathbf{x}_{C_j,k} = \hat{\mathbf{x}}_{C_j,k} + \mathbf{c}_k. \qquad (4)$$

Here, the notation for our images was slightly changed when compared to the notation used to define the data matrix $\mathbf{X}$ to account for the class membership information and number of images in each of the classes. Thus, the symbols used in Eq. (4) denote: C_j - the class label of the image, k - index of the image in the j-th class, $\hat{\mathbf{x}}_{C_j,k}$ - the channel-free part of $\mathbf{x}_{C_j,k}$, $\mathbf{c}_k$ - the vector encoding the channel effects for the k-th image of the j-th class.

Let us further assume that the channel effect vector $\mathbf{c}_k$ represents a random variable drawn from the standardized normal distribution $\mathcal{N}(0, 1)$. Then the class-conditional sample mean is defined as follows:

$$\boldsymbol{\mu}_{C_j} = \frac{1}{n_{C_j}}(\sum_{k=1}^{n_{C_j}} \hat{\mathbf{x}}_{C_j,k} + \sum_{k=1}^{n_{C_j}} \mathbf{c}_k) = \frac{1}{n_{C_j}} \sum_{k=1}^{n_{C_j}} \hat{\mathbf{x}}_{C_j,k}. \qquad (5)$$

The above expression suggests that the mean value of each of the r classes $\boldsymbol{\mu}_{C_j}$ ($j = 1, 2, ..., r$) represents a channel-free estimate of an image from the j-th class (if $n_{C_j} \gg 1$). Thus, by removing the corresponding class means from each image in the data matrix $\mathbf{X}$ we arrive at a new data matrix containing only information about the channel effects in the original data. To remove these effects from the input data, we estimate the NAP directions $\mathbf{w}_i$ ($i = 1, 2, ..., d'$) as principal axes of the scatter matrix $\boldsymbol{\Sigma}_w$:

$$\boldsymbol{\Sigma}_w = \sum_{j=1}^{r} \sum_{k=1}^{n_{C_j}} (\mathbf{x}_{C_j,k} - \boldsymbol{\mu}_{C_j})(\mathbf{x}_{C_j,k} - \boldsymbol{\mu}_{C_j})^T, \qquad (6)$$

where the axes are computed as the leading eigenvectors of the following eigenproblem:

$$\boldsymbol{\Sigma}_w \mathbf{w}_i = \lambda_i \mathbf{w}_i, \quad i = 1, 2, ..., d' \leq n - r. \qquad (7)$$

2.3. Illumination Invariance using NAP

To remove the illumination induced artifacts from the facial images using the presented NAP compensation scheme we first compare the additive form of the image model given by Eq. (1) and the model defined in Eq. (4). It is easy to see that given appropriate data in $\mathbf{X}$ we can estimate the NAP directions corresponding to the illumination induced variability in the logarithm domain. Any input image $\mathbf{x}$ (in vector form) can easily be normalized with respect to illumination artifacts by projecting away a number of directions in the NAP subspace. The normalization procedure can be written as:

$$\mathbf{x}' = \mathbf{P}(\mathbf{x} - \boldsymbol{\mu}), \qquad (8)$$

where $\boldsymbol{\mu}$ represents the global mean of the images in $\mathbf{X}$. Some examples of normalized images using the presented NAP scheme are shown in Fig. 1.

To effectively remove the effects of illumination, the data matrix $\mathbf{X}$ must be constructed in such a way as to include the
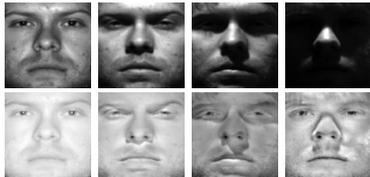
Fig. 1. Sample images from the YaleB database processed with the NAP compensation scheme with d'=40.

highest available number of images captured in different illumination conditions. In the experiments presented in Section 3 of the paper, we used part of the extended YaleB database for this purpose. The data in the database was captured under 64 different illumination conditions and can, therefore, be used as a generic database for estimating the NAP directions. The computed directions can of course be used with arbitrary frontal face images without the need for adapting them to unseen data.

3. EXPERIMENTS AND RESULTS

To assess the effectiveness of the presented NAP normalization scheme two popular face databases were chosen, namely, the extended YaleB and the XM2VTS database [7], [8].

The first, the extended YaleB database, contains 2432 frontal face images of 38 subjects captured under 64 illumination conditions. For the experiments the database is partitioned into two disjoint parts. The first part, featuring images of the first ten subjects of the database (labeled as B01 through B10) is exploited as the test image set. The second part, featuring images of the remaining subjects (labeled as B11 through B39) is employed as the development set and used to estimate the NAP directions. The reason for such a partitioning lies in the fact that the first ten subjects of the database comprise the original YaleB database and, hence, the results obtained on these images are comparable with other results published in the literature. Images of the first ten subjects are then further divided into five subsets in accordance with the extremity in illumination (the reader is referred to [7] for a detail description of these subsets). The first of these subsets (featuring 7 *controlled* images per subject) is employed to estimate the feature space and the rest is used for testing the performance of the normalization techniques in face identification experiments.

The second, the XM2VTS database, features 2360 facial images of 295 subjects with the images being captured under controlled conditions. For the experiments the database is divided into client and impostor groups and these groups are further partitioned into evaluation and test sets as defined by the first configuration of the experimental protocol associated with the database [8]. As the XM2VTS database was recorded in controlled conditions, it is employed as a reference database for measuring the deterioration in recognition performance when using photometric normalization techniques. It has to be noted that while the YaleB database is ex-
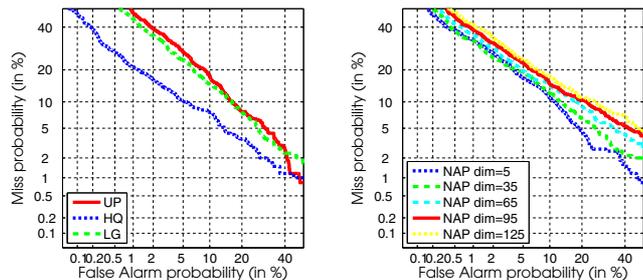

Fig. 2. DET curves (XM2VTS database): for three baseline systems (left), for varying number of NAP directions (right)

ploited to perform identification experiments, the XM2VTS database is used for verification experiments. For both datasets principal component analysis (PCA) is employed as the feature extraction technique and the nearest neighbor (to the mean) (1NN) classifier in conjunction with the cosine similarity (Cos) measure is adopted for the recognition task.

In the first series of our experiments we establish the baseline performance of our system (PCA+1NN+Cos) without any modification of the input images (UP) as well as with some basic pre-processing techniques, such as the logarithm transform (LG) and histogram equalization (HQ). The rank one recognition rates (in %) for the YaleB database and the DET curves for the XM2VTS database are shown in Table 1 and Fig. 2 (left), respectively. In Table 1 the four test image subsets of the YaleB database are labeled as S2 through S5 and the numbers in the brackets next to the labels denote the number of identification tests carried out for each subset.

Table 1. Rank one recognition rates (in %) for three baseline recognition systems.

Subset No.	S2 (120)	S3 (120)	S4 (140)	S5 (190)
UP	99.2	65.0	35.0	13.7
LG	100	78.3	43.6	42.1
HQ	98.3	65.8	35.0	42.6

From the presented results we can find that the recognition rates on the YaleB database systematically deteriorate with the increasing extremity in illumination and that the tested pre-processing techniques only partially compensate for the deterioration. Similar results are observed for the XM2VTS database, where the pre-processing techniques result in small improvements in the verification performance.

Now that we have a basic estimates of the baseline performance of our system we turn our attention to the assessment of the proposed NAP-based normalization scheme. First, we build the NAP subspace using all images of the subjects labeled as B11 through B39 in the extended YaleB database. Then, we evaluate the performance of our recognition system using varying numbers of NAP directions d'. To this end, we increase the number of NAP directions from $d' =5$ to $d' =125$ (with a step of 30) and observe the rank one recognition rate

for each of the four test image subsets of the YaleB database and generate DET curves using the evaluation client and impostor sets of the XM2VTS database. The results of this assessment are presented in Table 2 and Fig. 2 (right).

Table 2. Impact of varying number of NAP directions on the rank one recognition rates (in %) on the YaleB database.

Subset No.	S2 (120)	S3 (120)	S4 (140)	S5 (190)
NAP dim=5	100	100	95.0	91.6
NAP dim=35	100	100	99.3	96.8
NAP dim=65	100	100	97.1	95.8
NAP dim=95	100	100	99.3	98.4
NAP dim=125	100	100	100	98.4

Note that removing only 5 NAP directions from the images improves the recognition rates on all subsets of the YaleB database to over 90% and leads to a small improvement in the verification performance on the XM2VTS database when compared to the unprocessed and logarithm transformed images. In general, removing more than 5 NAP directions results in better recognition rates on images affected by illumination changes, while it degrades the performance on images captured in controlled conditions.

Last but not least we provide some comparative results obtained with other popular illumination normalization techniques on the YaleB and XM2VTS databases. Particularly, we provide recognition rates for the following normalization techniques: the single scale retinex algorithm (SR) [9], the logarithmic total variance technique (LTV) [11] and a wavelet-denoising-based method (WD) [10]. All the listed techniques including our NAP scheme are implemented without any pre- or post-processing (e.g., histogram equalization), which are known to further boost the recognition performance. The NAP technique is implemented with 35 NAP directions. The results of the comparison are presented in Table 3 and 4. The results suggest that the NAP technique achieves competitive results on both databases when compared to state-of-the-art normalization techniques. Note that

Table 3. Comparison with state-of-the-art techniques from the literature on the YaleB database.

Subset No.	S2 (120)	S3 (120)	S4 (140)	S5 (190)
NAP	100	100	99.3	96.8
SR	100	99.2	82.9	81.1
LTV	100	100	99.3	99.5
WD	100	100	98.6	99.5

the results for the XM2VTS are provided for the test sets (and not for the evaluation sets as in Fig. 2) and with the decision threshold that ensured equal values of the false acceptance (FAR) and false rejection rates (FRR) on the evaluation set. We also provide a single performance measure in form of the half total error rate (HTER = FAR+FRR).

Table 4. Comparison with state-of-the-art techniques from the literature on the XM2VTS database.

Technique	UP	NAP	SR	LTV	WD
FAR (%)	11.6	11.7	18.6	21.0	21.9
FRR (%)	10.0	8.5	17.5	15.5	18.8
HTER (%)	10.8	10.1	18.1	18.3	20.3

4. CONCLUSION

In this paper we have presented a novel nuisance attribute projection (NAP) based image normalization technique for tackling the problem of illumination invariant face recognition. The technique was tested in a series of face recognition experiments performed on the extended YaleB and XM2VTS databases. In all experiments the technique achieved state-of-the-art results.

5. REFERENCES

[1] V. Štruc, N. Pavešić:, "Photometric normalization techniques for illumination invariance," in: Y. Zhang (Ed.): Advances in Face Image Analysis: Techniques and Technologies, IGI Global, *accepted chapter*.

[2] Y. Adini, Y. Moses and S. Ullman:, "Face recognition : The problem of compensating for changes in illumination direction," *TPAMI*, vol. 19, no. 7, pp. 721-732, 1997.

[3] G. Heusch, F. Cardinaux, and S. Marcel:, "Lighting normalization algorithms for face verification", Tech. Rep. 05-03, IDIAP, March 2005.

[4] E.H. Land and J.J. McCann:, "Lightness and Retinex Theory," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 1-11, 1971.

[5] J. Short, J. Kittler, and K. Messer:, "A comparison of photometric normalisation algorithms for face verification," in: *Proc. of the AFGR*. IEEE, 2004, pp. 254-259.

[6] A. Solomonoff, C. Quillen and W.M. Campbell:, "Channel Compensation for SVM Speaker Recognition," in: *Proc. of Odyssey*, 2004, pp. 5762.

[7] A.G. Georghiades, P.N. Belhumeur, and D.J. Kriegman:, "From few to many: illumination cone models for face recognition under variable lighting and pose," *TPAMI*, vol. 23, no. 6, pp. 643-660, 2001.

[8] K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre:, "XM2VTSDB: the extended M2VTS database," in *Proc. of AVBPA*, 1999, pp. 72-77.

[9] D.J. Jobson, Z. Rahman, and G.A. Woodell:, "Properties and performance of a Center/Surround Retinex, *Trans. on Image Processing*, vol. 6, no. 3, pp. 451-462, 1997.

[10] T. Zhang, B. Fang, Y. Yuan, Y.Y. Tang, Z. Shang, D. Li and F. Lang:, "Multiscale facial structure representation for face recognition under varying illumination," *Pattern Recognition*, vol. 42, no. 2, pp. 252-258, 2009.

[11] T. Chen, W. Yin, X.S. Zhou, D. Comaniciu and T.S. Huang:, "Total variation models for variable lighting face recognition", *TPAMI*, vol. 28, no. 9, pp. 1519-1524, 2006.