ILLUMINATION INVARIANT FACIAL RECOGNITION USING A PIECEWISE-CONSTANT LIGHTING MODEL

Niall McLaughlin, Ji Ming, Danny Crookes

Institute of Electronics, Communications and Information Technology
Queen's University Belfast, Belfast BT3 9DT, United Kingdom

ABSTRACT

In this paper we demonstrate a simple and novel illumination model that can be used for illumination invariant facial recognition. This model requires no prior knowledge of the illumination conditions and can be used when there is only a single training image per-person. The proposed illumination model separates the effects of illumination over a small area of the face into two components; an additive component modelling the mean illumination and a multiplicative component, modelling the variance within the facial area. Illumination invariant facial recognition is performed in a piecewise manner, by splitting the face image into blocks, then normalizing the illumination within each block based on the new lighting model. The assumptions underlying this novel lighting model have been verified on the YaleB face database. We show that magnitude 2D Fourier features can be used as robust facial descriptors within the new lighting model. Using only a single training image per-person, our new method achieves high (in most cases 100%) identification accuracy on the YaleB, extended YaleB and CMU-PIE face databases.

***Index Terms*—** facial recognition, illumination invariance, lighting model, limited training data

1. INTRODUCTION

The problem of facial recognition under realistic lighting has recently received much research attention and several major approaches have been explored including: illumination modelling, photometric normalisation and illumination invariant representations. In this paper we study facial identification given unknown realistic lighting, with a single training image per-person.

The standard theory of facial illumination is the Retinex model [1]. In the Retinex model, lighting is assumed to be represented by the low-frequency spatial components of the face image. This motivates a number of recognition approaches aimed at recovering an illumination invariant facial representation by removing the low-frequency components from a given face image. In self quotient imaging (SQI) [2] the illumination information is approximated by a smoothed version of the face, which is subtracted from the logarithm of the facial image intensity, yielding an 'invariant' facial representation. The performance of SQI can be improved by using non-local means de-noising to better estimate the facial illumination [3]. Similar filtering can be performed in the frequency domain, using filters based on the Fourier [4] or DCT [5] transforms. Both spatial and frequency domain filtering encounter problems with 'haloing' at sharp edges caused by cast shadows which can be somewhat alleviated by using multi-scale filters to better approximate the illumination information [6]. Retinex theory has also been used derive feature representations which are invariant under illumination change, such as Gradientfaces [7] or local ternary patterns [8].

Our new approach extends the Retinex illumination model to express illumination as a piecewise-constant function over the face image. We empirically demonstrate, using the YaleB face database, that over a small area of the face, illumination can be modeled as two constant components. An additive component representing the shift in pixel mean with illumination and a multiplicative component representing the pixel variance. Using this simple illumination model, both illumination components can by normalized for each small area of the face, allowing facial recognition to be performed despite large variations in illumination.

This work has been funded by Intel Ireland.

2. ILLUMINATION MODEL

In the Retinex model [1], the intensity of every facial image pixel $I(x, y)$ is dependent on the intrinsic reflectivity (or albedo) of the face $R(x, y)$ and the lighting conditions $L(x, y)$. We can express this model as

$$I(x, y) = R(x, y)L(x, y) \tag{1}$$

In order to perform illumination invariant facial recognition, we must recover the invariant reflectance information $R(x, y)$ given only the pixel intensity values $I(x, y)$.

It has been observed that in realistic facial images the illumination information varies slowly over the face image i.e., the illumination information $L(x, y)$ is present at low spatial frequencies while the intrinsic reflectance information $R(x, y)$ is present at higher spatial frequencies. Using this observation we can assume that over a small area of the face image, illumination $L(x, y)$ is approximately constant. We propose an illumination model which separates the effects of illumination over a small area of the face image into two components. The first component is additive, modeling the change of the mean pixel intensity in the area, which varies with illumination. The second component is multiplicative, modeling the change of the variance of the pixel intensity in the area. Let ϕ represent a small area over a given face image. The observed pixel intensity $I(x, y)$ within ϕ can thus be expressed as:

$$I(x, y) \approx k_\phi R(x, y) + c_\phi \quad \forall (x, y) \in \phi \tag{2}$$

where $R(x, y)$ corresponds to the intrinsic reflectance information to be retrieved for recognition, c_ϕ represents an additive bias, and k_ϕ represents a multiplicative factor. Together c_ϕ and k_ϕ model the variation of lighting conditions under which $R(x, y), \forall (x, y) \in \phi$, is observed. We assume that for a small area ϕ, both c_ϕ and k_ϕ are approximately constant. Previous approaches based on removal of low frequency coefficients, or on homomorphic filtering, are attempts at removing the constants c_ϕ and k_ϕ. In this paper, we propose a novel

method for illumination invariant facial recognition based on the illumination model (2). This is achieved by dividing each facial image into small sub-images such that each sub-image can be modeled by (2). Then, we use a cosine similarity function combined with mean removal to simultaneously remove c_ϕ and k_ϕ.

3. ILLUMINATION INVARIANT FACE RECOGNITION

3.1. Illumination Invariant Similarity Measure

We compare realistically illuminated training and test face images. Both face images are aligned and divided into small blocks. For each block ϕ, we can remove the constant additive bias c_ϕ by subtracting the local mean from the block (more details of the mean removal will be discussed later). Then, corresponding blocks may be compared by using a cosine similarity measure that is invariant to the constant multiplier k_ϕ. (Note: it possible to normalize the multiplier k_ϕ using the block variance, allowing a different similarity measure e.g. Euclidean distance to be used; this will be discussed later). The cosine similarity $C(\mathbf{a}, \mathbf{b})$ between two vectors $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_N)$, can be expressed as

$$C(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}.\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|} \tag{3}$$

Cosine similarity calculates the cosine of the angle between two vectors. It can be shown that the cosine similarity $C(k\mathbf{a}, l\mathbf{b})$ between two new vectors, $k\mathbf{a}$ and $l\mathbf{b}$, which are $\mathbf{a}$ and $\mathbf{b}$ each subject to constant scalar multipliers k and l respectively, equals $C(\mathbf{a}, \mathbf{b})$

$$\begin{aligned} C(k\mathbf{a}, l\mathbf{b}) &= \frac{k\mathbf{a}.l\mathbf{b}}{\|k\mathbf{a}\|\|l\mathbf{b}\|} \\ &= \frac{kla_1b_1 + \cdots + kla_Nb_N}{\sqrt{(ka_1)^2 + \cdots + (ka_N)^2}\sqrt{(lb_1)^2 + \cdots + (lb_N)^2}} \\ &= \frac{kl(a_1b_1 + \cdots + a_Nb_N)}{kl\sqrt{(a_1{}^2 + \cdots + a_N{}^2)}\sqrt{(b_1{}^2 + \cdots + b_N{}^2)}} \\ &= \frac{\mathbf{a}.\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|} \end{aligned} \tag{4}$$

Here $\mathbf{a}$ and $\mathbf{b}$ can be two image blocks and k and l correspond to their respective multiplicative illumination factor. Eq. (4) then offers an illumination-invariant similarity measure between the two image blocks. Let $X = (X_1, X_2, \ldots, X_M)$ represent a training face image and $Y = (Y_1, Y_2, \ldots, Y_M)$ represent a test face image, both divided into M aligned small blocks where X_m and Y_m denote the mth block in X and Y. It can be shown that the cosine similarity between the two full images X and Y can be approximated by the sum of the cosine similarities between their constituent blocks [9]. In this paper, we use the expression:

$$C(X, Y) \simeq \sum_{m=1}^{M} C(X_m - \bar{X}_m, Y_m - \bar{Y}_m) \tag{5}$$

where $\bar{X}_m$ and $\bar{Y}_m$ represent the mean pixel intensity of blocks X_m and Y_m respectively. This illumination invariant similarity measure can be viewed as finding the Pearson correlation coefficient between the local facial areas X_m and Y_m.

3.2. Feature Representation

So far we have assumed that the feature vector representing each facial block is formed by concatenating the pixel intensity values of the block. However, the block-by-block comparison of two facial images rests on the assumption that both facial images are very well aligned. Small changes in an individual's facial expression and head pose mean that even well aligned facial images captured under controlled conditions will not correspond exactly. Facial blocks can instead be represented using their magnitude 2D Fourier representation. By taking the magnitude, we omit the phase information, which allows us to take advantage of the shift invariance of the magnitude Fourier representation. This means the system should be more robust to small misalignment errors and small facial expression changes.

Apply 2D Fourier transform to the illumination model (2)

$$\mathcal{I}(u, v) \simeq k_\phi \mathcal{R}(u, v) + c_\phi \tag{6}$$

where $\mathcal{I}(u, v)$ and $\mathcal{R}(u, v)$ are the Fourier transforms of $I(x, y)$ and $R(x, y)$, respectively. We ignore the 0^{th} Fourier coefficient $\mathcal{I}(0, 0)$, which is equivalent to subtracting the block mean in (5). Thus we obtain an illumination model for the magnitude Fourier features

$$||\mathcal{I}(u, v)|| \simeq k_\phi ||\mathcal{R}(u, v)|| \quad (u, v) \neq (0, 0) \tag{7}$$

Using the magnitude Fourier image representation, (5) can be rewritten as

$$C(X, Y) \simeq \sum_{m=1}^{M} C(||\mathcal{X}_m||, ||\mathcal{Y}_m||) \tag{8}$$

where $||\mathcal{X}_m||$ is the magnitude Fourier representation of image block X_m with mean removed (likewise for $||\mathcal{Y}_m||$). Next we show the use of (8) for facial recognition with varying lighting conditions.

4. EXPERIMENTS

We now test the performance of our proposed illumination invariant facial recognition system on faces with realistic lighting. We perform experiments using the YaleB, extended YaleB and CMU-PIE facial databases. The YaleB database contains the frontal face images of 10 persons, each captured under 64 illumination conditions. The extended YaleB database adds an additional 28 persons, captured under the same conditions as the YaleB database, increasing the total number of persons to 38. During testing we divide the face images from each person into 5 subsets dependent on illumination angle: subset 1 $(0^\circ - 12^\circ)$, subset 2 $(13^\circ - 25^\circ)$, subset 3 $(26^\circ - 50^\circ)$, subset 4 $(51^\circ - 77^\circ)$, and subset 5 $(78^\circ - 90^\circ)$. Example images from each subset of YaleB 6are shown in Fig. 3. The CMU-PIE database consists of 68 persons illuminated from 21 directions.

4.1. Experimental Test of Lighting Model

In this experiment we test the assumptions underlying our lighting model (2) , i.e. that lighting can be modeled as a constant change in both the mean and variance of pixel intensities in a small facial area. For the first part of this experiment we use Euclidean distance, rather than cosine similarity to compare blocks, demonstrating the generality of our lighting model. This experiment was conducted on the YaleB face database using non-overlapping blocks, represented by pixel intensity feature vectors. The block size was varied linearly from 2×2 pixels to the whole image size of 168×192 pixels, in order to find the range of block sizes for which our assumption of constant illumination holds. A single facial image per-person from subset 1 was used for training and all other images were used for testing. First, baseline identification accuracies using Euclidean distance and no illumination compensation were found. The experiment was then repeated using three modified feature representations:

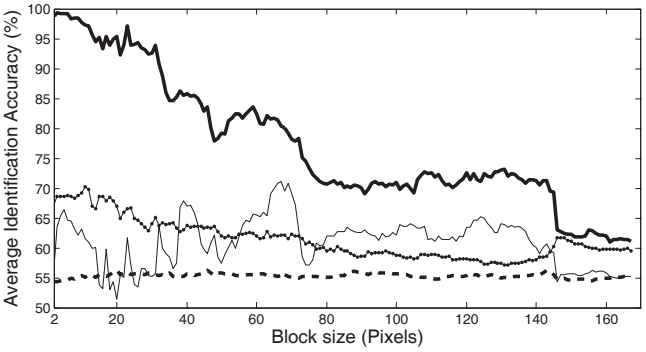

Fig. 1. Average identification accuracy (%) over all subsets of YaleB, using Euclidean distance, plotted as a function of block size. Thick solid line - block mean and variance normalized. Dotted line - block mean normalized. Thin solid line - block variance normalized. Dashed line - no illumination compensation.

block mean c_ϕ normalized, then with block variance k_ϕ normalized so the values in each block had unit range, and finally with both block mean and variance normalized. The results of this experiment are shown in Fig. 1. We can see from Fig. 1 that normalising the block mean improves identification accuracy over the baseline. Normalising only the block variance generally has an unpredictable effect on accuracy and for certain block sizes performs worse than the baseline. The best identification accuracy is consistently achieved when both the mean and variance of each block have been normalized. Due to the trade-off between the illumination invariance and discriminative ability of each block size, accuracy remains highest while the block size remains relatively small. Maximum average identification accuracy occurs at a block size of 3×3 pixels, which is larger than the minimum block size of 2×2 pixels, reflecting the invariance/discrimination trade-off. As block size is increased, our assumption of constant illumination within each block becomes invalid, so identification accuracy tends to decrease. The results of this experiment support our hypothesis that over a small facial area, illumination can be modeled by two constant components: a change in both the mean and variance of the pixel values.

Having verified the effectiveness our illumination model using euclidean distance and pixel intensity features, we next performed a similar experiment on the YaleB database using cosine similarity and magnitude 2D Fourier features. We use cosine similarity rather than normalized Euclidean distance for the remainder of these experiments as when both measures were compared in further testing (results not shown due to space constraints), cosine similarity was found to give slightly better accuracy. Block size was varied systematically from 2 pixels to 50 pixels in each dimension. Any blocks overlapping the image edge were padded with zeros. A single evenly illuminated image from subset 1 was used as training for each person. Testing was performed using all the remaining images from all subsets. All faces images were 168×192 pixels and were not resized. As a preprocessing step, each face image was convolved with a 3×3 Gaussian kernel to reduce noise in the shadow areas, as this has been found to increase identification accuracy.

Fig. 2 shows mean identification accuracy over all 5 illumination subsets as a function of block size. We again observe a trade-off between illumination invariance and discriminative ability as block size is varied. Maximum identification accuracy occurs when block size is 6×7 pixels. This optimal block size differs slightly from the previous experiment, demonstrating that the trade-off between discriminative ability and illumination invariance is dependent on the feature representation. It should be noted that while block size remains small, average identification accuracy remains consistently high i.e. the variance in average identification accuracy for block sizes close to 6×7 pixels is less than 1%. This shows that our system is robust to changing block size, and the high identification accuracy is not the result of over-fitting parameters to this particular database. However, we also note that using smaller blocks may make our system more sensitive to misalignment errors, and we return to this problem in the next section. In Table 1 we compare the optimal results produced by our system using 6×7 pixel blocks to representative results from the literature which also used a single face image per person for training.

Table 1. *Percentage accuracy of our facial identification system compared to example systems in the literature, trained with a single facial image and tested under the same conditions on YaleB.*

Subset	**Our System**	Gradientfaces [7]	Adaptive Weiner [10]
1	**100**	100	100
2	**100**	100	100
3	**100**	99.76	100
4	**100**	96.23	94.29
5	**99.57**	99.47	98.42

4.2. The Effect of Features on Identification Accuracy

We hypothesised in Section 3.2 that using magnitude Fourier features would lead to improved accuracy compared to pixel intensity features, as magnitude Fourier features omit the phase information from each block, making them more robust to small misalignment errors. We test this hypothesis by comparing the identification accuracy when representing each block using magnitude 2D Fourier features, pixel intensity features, and 2D discrete cosine transform (DCT) features. The DCT is a Fourier derived transform that represents a real-valued signal as the linear sum of a series of real cosine basis functions, and thus does not have the ability to discard the phase of a signal. We therefore expect the accuracy when using 2D DCT features to decrease compared to magnitude 2D Fourier features. This experiment was performed using the YaleB face database with a single face image per-person for training and all the other images from all subsets used for testing. We used cosine similarity to compare blocks and the optimal block size of 6×7 pixels. When using pixel intensity features we subtracted the block mean from each block, and when using DCT features the 0^{th} DCT coefficient was discarded as it corresponds to the block mean. The results

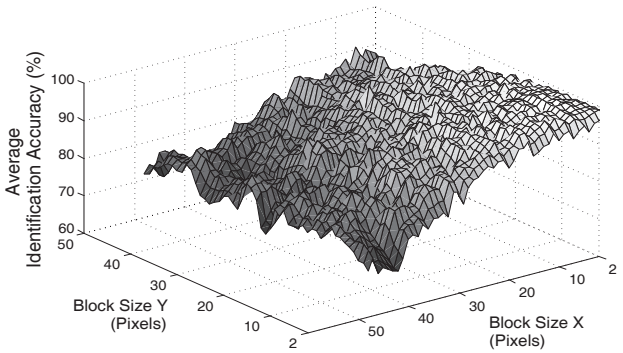

Fig. 2. Average identification accuracy (%) over all subsets of YaleB, using Cosine similarity. Results shown as a function of block size.

of this experiment are presented in Table 2. As expected magnitude Fourier features produce better accuracy than 2D DCT features or pixel intensity values, which supports our hypothesis that magnitude Fourier features show increased robustness due to the omission of phase information.

Table 2. *Percentage accuracy of our facial identification system when tested using 2D magnitude Fourier, 2D DCT and pixel intensity feature types on YaleB database.*

Subset	**2D mag. Fourier**	2D DCT	Pixel Intensity
1	**100**	100	100
2	**100**	100	100
3	**100**	99.00	99.00
4	**100**	93.33	93.33
5	**99.57**	93.04	93.04

4.3. Enhancing Identification Accuracy

The identification accuracy of our system can be enhanced by a number of simple modifications to our original algorithm. As the identification accuracy of our system on YaleB is almost 100%, we now move to the extended YaleB database, which was captured under the same conditions as YaleB but includes 38 persons. Baseline identification accuracy on the extended YaleB database using our above described system is shown in Table 3 under 'Original System'.

The first modification is to extract features using overlapping blocks, as opposed to our original system which used non-overlapping blocks. Blocks were overlapped by 50% in each dimension, using the same 6×7 block size as previous experiments. Using overlapping blocks has the effect of reducing discontinuities that may occur near edges such cast shadows. The results, in Table 3, show that overlapping blocks improves accuracy compared to our original system. Identification accuracy can be further enhanced by preprocessing each facial image using a band-pass filter. This filtering approach is similar to self quotient imaging (SQI) [2], in assuming that low spatial frequencies represent illumination information, while high spatial frequencies represent noise. Band-pass filtering was implemented using 3×3 and 9×9 Gaussian kernels. After this preprocessing step, the band-pass filtered face image was processed in the usual manner by our originally described system. The band-pass filtering results are shown in Table 3. By combining feature extraction using overlapping blocks with a band-pass filter preprocessing step, identification accuracy can be improved again. Results for the combined system are shown in Table 3 and are comparable with the best existing systems [8]. Additionally, when tested on the original YaleB database our enhanced system achieves 100% identification accuracy on all lighting subsets. Finally, we tested our enhanced system on the CMU-PIE database which contains 68 individuals illuminated from 21 directions. Using a single training image per-person and the same parameters as the YaleB database, 100% identification accuracy was achieved.

5. CONCLUSION

In this paper we have demonstrated a simple and novel illumination model for facial recognition under realistic illumination conditions. We model the effects of illumination over a small facial area using a piecewise-constant function with separate multiplicative and additive components. We have shown that magnitude 2D Fourier features can be used as robust facial descriptors. Our tests have shown that the novel illumination model achieves very high identification accuracy on the YaleB, extended YaleB and CMU-PIE facial databases.

Table 3. *Percentage accuracy of our facial identification system tested on the extended YaleB database, with overlapping block feature extraction and band-pass filter preprocessing of image.*

Subset	Original System	Overlap	Bandpass	**Bandpass + Overlap**
1	99.62	100	100	**100**
2	100	100	100	**100**
3	97.89	98.42	99.74	**100**
4	91.29	93.97	99.78	**100**
5	88.80	91.80	96.42	**97.11**

6. REFERENCES

[1] Shiguang Shan, Wen Gao, Bo Cao, and Debin Zhao, "Lightness and retinex theory," *Journal of Optical Society of America*, vol. 61, no. 1, pp. 1–11, 1971.

[2] H. Wang, S.Z. Li, and Y. Wang, "Face recognition under varying lighting conditions using self quotient image," *IEEE Conf. Automatic Face and Gesture Recognition*, pp. 819–824, 2004.

[3] Vitomir Štruc and Nikola Pavešić, "Illumination invariant face recognition by non-local smoothing," *Conf. Biometric ID management and multimodal communication*, pp. 1–8, 2009.

[4] Sang-Il Choi and Gu-Min Jeong, "Shadow compensation using fourier analysis with application to face recognition," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 23–26, 2011.

[5] Virendra P. Vishwakarma, Sujata Pandey, and M. N. Gupta, "An illumination invariant accurate face recognition with down scaling of dct coefficients," *Journal of Computing and Information Technology*, pp. 53–67, 2010.

[6] D.J. Jobson, Z. Rahman, and G.A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[7] Taiping Zhang, Yuan Yan Tang, Bin Fang, Zhaowei Shang, and Xiaoyu Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2599–2606, 2009.

[8] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[9] N. McLaughlin J. Ming and D. Crookes, "Robust bimodal person identification using face and speech with limited training data and corruption of both modalities," *Interspeech*, 2011.

[10] Lin Jiang, Bin Fang, Tai-Ping Zhang, Yuan-Yan Tang, and Dong-Hui Li, "Face recognition under varying illumination using adaptive filtering," *ICWAPR 2009*, pp. 38–42, 2009.

Fig. 3. Example images form each of the five illumination subsets of the YaleB database for one subject. The neutral condition, subset 1 is shown on the left, with illumination angle increasing to the right.