ANALYZING FACIAL IMAGES USING EMPIRICAL MODE DECOMPOSITION FOR ILLUMINATION ARTIFACT REMOVAL AND IMPROVED FACE RECOGNITION

Ramamurthy Bhagavatula and Marios Savvides
Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh PA 15213
rbhagava@andrew.cmu.edu, msavvid@ri.cmu.edu

ABSTRACT

A popular modality of biometrics, facial recognition is effective when used in controlled environments as in those situations where factors such as camera position, facial expression, and illumination effects are either completely or partially controlled in a beneficial way. Regulation of such factors has an immediate effect on the performance of facial recognition algorithms, in particular illumination effects which can not be controlled by even the most cooperative of users. In this paper we describe a method to address illumination effects in the biometric modality of face recognition using the signal processing analysis tool of Empirical Mode Decomposition (EMD) to decompose images into their intrinsic mode function that correspond to the dominant illumination factors. Using these illumination modes we reconstruct the facial image without these illumination distortion components to synthesize a more illumination neutral facial image. We then perform verification experiments using algorithms such as Principal Component Analysis (PCA), Fisher Linear Discriminant Analysis (FLDA), and Advanced Correlation Filters (ACF's) to demonstrate the fundamental effectiveness of EMD as an illumination compensation method. Results are reported on the Carnegie Mellon University Pose-Illumination-Expression (CMU PIE) database.

Keywords: Image Processing, Image Classification, Security.

1. INTRODUCTION

Due to its availability and relative uniqueness, facial recognition is becoming one of the most popular modalities of biometrics. Though evaluations such as the Face Recognition Vendor Test (FRVT) [14] and the Face Recognition Grand Challenge (FRGC) [15] indicate strong progress in the overall area of face recognition, there are still exist many obstacles to widespread use.

When considering all potential areas of deployment for facial recognition systems, the issue of handling illumination-variations becomes a priority. It has been shown experimentally and theoretically that illumination variations can cause a significant degradation in performance of facial recognition systems [1]. Many potential solutions to this problem have been proposed varying in complexity and effectiveness [3, 4, 7]. These solutions are improvements to facial recognition under illumination-variant conditions, but are not ideal. The complexity and assumptions of idealities in many of these methods often limit their overall applicability.

We will show the power of Empirical Mode Decomposition (EMD) in addressing illumination effects in facial recognition. Using EMD to decompose two-dimensional facial images into their fundamental source signals, we can isolate the effects of illumination to one or more of these source signals. By reconstructing the image without these illumination artifact source signals, we can reduce the overall effect of illumination variation. As EMD is very algorithmic, implementation is simple while still being effective. Recognition results to demonstrate the improvement of EMD processing are reported using Principal Component Analysis (PCA), Fisher Linear Discriminant Analysis (FLDA), and Correlation Filters (CF's) on the Carnegie Mellon University Pose-Illumination-Expression (CMU PIE) database.

2. EMPIRICAL MODE DECOMPOSITION

EMD was pioneered [9] as a signal processing technique for adaptive representation of nonstationary signals as sums of zero-mean AM and FM components. Employed in multiple applications not directly related to facial recognition [6, 10], EMD has been to shown to be an effective tool for analysis and manipulation of signals. However, EMD's definition as an algorithm as opposed to theory lends itself to varying implementations ranging in complexity and accuracy.

A general overview of EMD and its implementation is presented in [6], but we will briefly summarize EMD here. EMD aims to capture information about local trends in the signal data by measuring and quantizing oscillations. Such oscillations can be quantized by a local high frequency or local detail and correspondingly a local low frequency or local trend. The source signal being composed of these local details and trends can be iteratively reduced to characteristic signals. The following algorithm defines this procedure and outlines most EMD implementations. Given a source signal $x(t)$:

1. Identify all local extrema of x(t) (minima and maxima)
2. Interpolate between all minima (resp. maxima) to yield an envelope $e_{min}(t)$ (resp. $e_{max}(t)$)
3. Compute the mean envelope $m(t) = (e_{min}(t) + e_{max}(t))/2$
4. Extract the detail $d(t) = x(t) - m(t)$
5. Check if $|m(t)| < \varepsilon$. If not repeat steps 1-4 with d(t) as the input signal x(t). If so, d(t) is an intrinsic mode function (IMF)
6. Calculate residual $r(t) = x(t) - d(t)$
7. Go back to step 1 with r(t) as the input signal x(t)
8. Repeat until input signal no longer has any extrema

An IMF satisfies two conditions [9]. First, the number of extrema and the number of zero-crossings must be equal or at most differ by one. Second, at any point, the mean values of the envelopes defined by the local maxima and the the local minima respectively must equal zero. The first four steps shown above are collectively referred to as the "sifting process" and are the most computationally expensive and error prone portion of the algorithm due to identification of extrema and subsequent interpolation. The primary power of EMD is that once these IMF's have been found, we can easily go back and forth from them to the original data. By simply summing all the IMF's together we will recover the original data [6, 10] accommodating for minor variations due to the interpolation present in the algorithm. EMD also allows us to selectively reconstruct the data, ignoring the IMF's whose contributions to the data are undesirable. For our application, such contributions are those of illumination effects. If we can use EMD to decompose our original facial images into their IMF's, there is a strong likelihood that the effects of illumination will be isolated to one or more IMF's. Selective reconstruction of facial images using IMF's that do not contain illumination effects will enable us to reconstruct the fundamental nature of the data without the unwanted effects of illumination variation.

3. RECOGNITION ALGORITHMS

PCA [16] is applied to a collection of facial images to compute the principal directions of variation in the high-dimensional facial space called Eigenfaces. Derivation of this basis requires solving the following generalized eigenvalue problem where X is a matrix containing the vectorized training facial images along its columns.

$$\boldsymbol{XX^T v} = \boldsymbol{Cv} = \lambda \boldsymbol{v} \qquad (1)$$

where the covariance matrix $\boldsymbol{C}$ is symmetric and positive semi-definite. The eigenvectors corresponding to the largest eigenvectors computed in Eq. (1) form an optimal orthogonal basis in a minimum mean squared error sense.

FLDA [8] finds the optimal projection vectors $\boldsymbol{w}$ such that the projected samples have a small within-class scatter ($\boldsymbol{w^T S_W w}$), and large between-class scatter ($\boldsymbol{w^T S_B w}$). This is done by maximizing the ratio of determinant of the projected between-class scatter matrix $\boldsymbol{S_B}$ to the determinant of the within-class scatter matrix $\boldsymbol{S_W}$. However, the dimensionality of the data is larger than the number of samples causing the within-class scatter matrix $\boldsymbol{S_W}$ to not be full rank leading to a zero determinant. To avoid a singular matrix $\boldsymbol{S_W}$, PCA is applied to the data to reduce the dimensionality and maintain a full-rank $\boldsymbol{S_W}$ and then multi-class FLDA is implemented the reduced-dimensional space [3]. Maximizing this ratio leads to the following generalized eigenvalue problem:

$$\boldsymbol{S_B w} = \lambda \boldsymbol{S_W w} \qquad (2)$$

Once LDA is performed, we can cascade the two projections into one transformation termed Fisherfaces [3] for convenience. Testing is performed by projecting training images into the Fisherface subspace and a simple nearest neighbor classifier is used to label the test image based on the residue.

Advanced Correlation Filters (CF's) [17] are template-based classifiers derived from spatial-frequency analysis that when correlated with an image result in a correlation plane with pre-desired response. The correlation plane C measures the correlation between the filter and the image at all possible shifts. We use a standard measure called Peak to Correlation Energy (PCE) [17] to quantify the degree of correlation present. The Minimum Average Correlation Energy (MACE) Filter [11] minimizes correlation plane energy in C while constraining peak values at the origin to pre-specified values. To create the MACE filter we analyze the spectral power density of the training data $\boldsymbol{X}$ (containing the Fourier transformed images vectorized along each column) and placed on the diagonal of the matrix $\boldsymbol{D}$. Our goal is to minimize energy E which is defined as:

$$\boldsymbol{E} = \boldsymbol{h^+ Dh} \qquad (3)$$

whose constrained minimization yields the filter h_{MACE}:

$$\boldsymbol{h_{MACE}} = \boldsymbol{D^{-1} X}\left(\boldsymbol{X^+ D^{-1} X}\right)^{-1} \boldsymbol{u} \qquad (4)$$

where $\boldsymbol{u}$ is the constrained peak values (vector of ones).

The Unconstrained MACE (UMACE) Filter [13] removes the constraint on the correlation peak value allowing for more solutions to the minimization problem. We also try to maximize the average value of the peaks. The closed form solution to the filter $\boldsymbol{h_{UMACE}}$:

$$\boldsymbol{h_{UMACE}} = \boldsymbol{D^{-1} m} \qquad (5)$$

where $\boldsymbol{m}$ is a vector containing the average Fourier transform training image.

We consider generalizations of the MACE and UMACE filters called the Optimal Tradeoff Synthetic Discriminant Function (OTSDF) filter and the Unconstrained OTSDF (UOTSDF) filter respectively [18]. These generalized filters offer sharp correlation peaks and noise tolerance. Given a desired proportion of peak sharpness to noise tolerance α, $\boldsymbol{h}_{OTSDF}$ and $\boldsymbol{h}_{UOTSDF}$ follow Eqs. (4) and (5) respectively except $\boldsymbol{D}$ is replaced by $\boldsymbol{T}$ defined as:

$$\boldsymbol{T} = \alpha \boldsymbol{D} + \left(\sqrt{1-\alpha^2}\right)\boldsymbol{P} \qquad \textbf{(6)}$$

where $\boldsymbol{P}$ is the Gaussian white noise matrix (identity matrix).

4. EMD PREPROCESSING

While there are extensions of the basic one-dimensional algorithm [9] to two-dimensional data [5], they are unnecessary in our experiments. Illumination effects are primarily due to one primary source of light which creates the majority of shadows and specular effects such as in Fig. 2. As such we can treat these effects as linear in a one-dimensional sense, although not necessarily along the typical axes. Treating each row or column of a facial image as separable we can string two-dimensional facial images into one-dimensional vectors. Application of EMD to these vectors yields a set of vector IMF's which are then reshaped into matrix IMF's as shown in Fig. 1.

The stopping conditions set in the EMD algorithm determine the exact number of IMF's but for our experiments and data we found that we obtain thirteen IMF's. Regardless of the exact number of IMF's, the last two IMF's contains the majority of the illumination effects. Due to nature of the EMD algorithm, as the order of the IMF increases the relative mean of the data approaches zero [6]. As such, by applying EMD to facial images that are subject to illumination effects we can partition the effects into two types, shadowing and specular reflections. Since shadowing darkens regions of an image, it creates low-valued regions while specular reflections create relatively high-valued regions. These are effectively the largest magnitude extrema in the images, but also most slowly changing. In other words they represent the lower spatial frequency contents of the image. EMD isolates these frequencies in the last few IMF's.

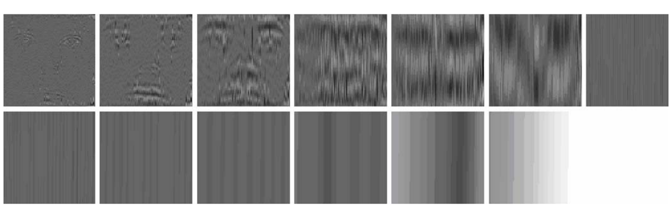

Figure 1: Resulting IMF's from Person 1, Image 1. Ordered from left to right, top to bottom in increasing order

With this in mind, we look at the last two IMF's and determine which one introduced the shadowing artifacts to the data. This is easily done by comparing the means of the two IMF's and choosing the smaller one. Once we have determined which IMF is responsible for the effects of shadowing, we reconstruct the image without that IMF. The resulting facial image now contains significantly less shadowing effects and allows the fundamental nature of the facial image to come through more.

(a) (b) (c) (d)

Figure 2: Examples of facial image reconstruction excluding unwanted IMF. (a) Original image with right-side specular reflection. (b) Reconstructed image minus specular reflection. (c) Original image with left-side cast shadow. (d) Reconstructed image minus left-side cast shadow effects

5. EXPERIMENTAL RESULTS

We used the PIE database [15] which focuses on illumination variation with minimal pose and expression variation. The subjects are captured under lighting conditions creating shadows of varying orientations and degrees. The data consists of 100×100 pixel facial images of 65 different people of both genders. Each person has 22 images yielding a total of 1430 images.

Table 1: Average Equal Error Rate (EER) for PCA and FLDA and EMD processed images.

# of Training Images	PCA	EMD-PCA	FLDA	EMD-FLDA
2	0.2362	0.0474	0.1907	0.2422
3	0.1557	0.0332	0.1553	0.2022
4	0.1560	0.0241	0.1548	0.1523
5	0.1214	0.0136	0.1153	0.1256
6	0.1775	0.0159	0.1457	0.1417
7	0.1327	0.0155	0.1235	0.0945
8	0.1419	0.0090	0.0942	0.0877
9	0.1305	0.0069	0.1149	0.0840
10	0.1683	0.0083	0.1065	0.0883
11	0.1816	0.0075	0.1165	0.1275

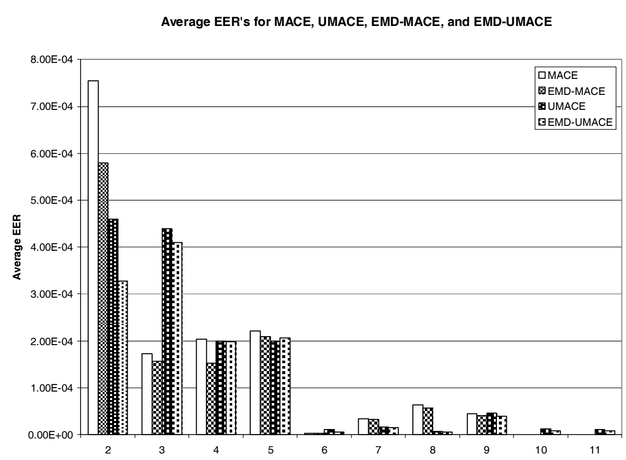

Figure 3: Average EER's for MACE and UMACE

Applying our EMD preprocessing to the entire database removed the significant illumination variation from the facial images. Training sets vary in size and composition by random selection over multiple experiments for each of the three recognition algorithms. Each experiment involved training the recognition algorithm using the specified training set and then recording verification results. For PCA and FLDA ten experiments were run each while for CF's only five due to computational intensity. Performance is quantified by average Equal Error Rate (EER) over all experiments.

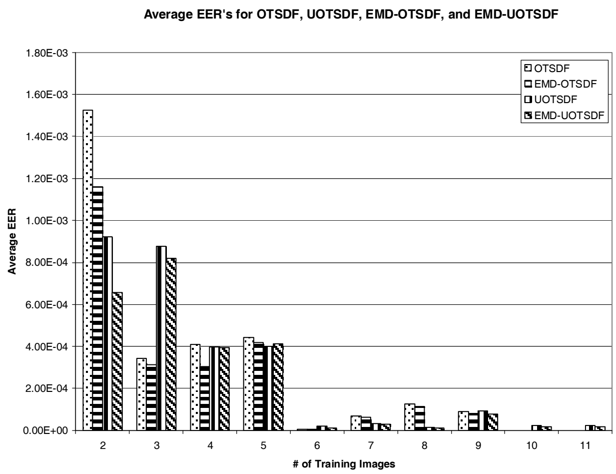

Figure 4: Average EER's for OTSDF and UOTSDF

6. CONCLUSIONS AND FUTURE WORK

Experimental results demonstrate that EMD preprocessing is an effective approach in normalizing a facial image in both space and frequency especially when using very small sample size training sets. Results from CF's, OTSDF and UOTSDF in particular, show that EMD preprocessing does not introduce undesired noise to the images. Even in its most simple form (one-dimensional, standard algorithm) EMD preprocessing achieves significant normalization of facial images subject to illumination variation with no a priori information. The simple implementation and effectiveness of EMD preprocessing indicates its usefulness as a preprocessing step in facial recognition algorithm. Expanding on the work presented here, we plan to improve results through multiple modifications such as the use of a true two-dimensional EMD algorithm [5] (that can possible be more suited when dealing from illumination artifacts arising from more than one illumination source). Different interpolation techniques and intelligent boundary conditions should increase the effectiveness of EMD preprocessing as illumination-variation normalization tool.

REFERENCES

[1] Adini, Y., Moses, Y., & Ullman, S., "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 721-732, July 1997.

[2] Basri, R., & Jacobs, R., "Lambertian Reflectance and Linear Subspaces," ICCV 2001, Vol. 2, p383-390, 2001.

[3] Belhumeur, P. N., Hespanha, J. P., & D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, 1997.

[4] Chen, T., Hsu, Y. J., Liu, X., & Zhang, W., "Principle Component Analysis and its Variants for Biometrics," presented at Image Processing. 2002. Proceedings. 2002 International Conference on, 2002.

[5] Damerval, C., Meignen, S., & Perrier, V., "A fast algorithm for bidimensional EMD," Signal Processing Letters, IEEE, vol.12, no.10pp. 701- 704, Oct. 2005.

[6] Flandrin, P., Goncalves, P., Rilling, G., "On Empirical Mode Decomposition and its Algorithms," Signal Processing Letters, IEEE, 2003.

[7] Gross, R., Matthews, I., & Baker, S., "Eigen Light-Fields and Face Recognition Across Pose", Proc. of FG02, 2002.

[8] Hart, P. E., Duda, R.O., & Stork, D.G., "*Pattern Classification*" (2nd Edition), 2001.

[9] Huang, N.E., Shen, Z., Long, S.R., Wu, M.L., Shih, H.H., Zheng, Q., Yen, N.C., Tung C.C., & Liu, H.H., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. Roy. Soc. London A, Vol. 454, pp. 903-995, 1998.

[10] Linderhed, A., "Compression by image empirical mode decomposition," Image Processing, 2005. ICIP 2005. IEEE International Conference on , vol.1, no. pp. I- 553-6, Sept. 2005.

[11] Mahalanobis, A., Vijaya Kumar, B.V.K., & Casasent, D., "Minimum average correlation energy filters," Applied Optics, vol. 26:3633-3640, 1987.

[12] Mahalanobis, A., Vijaya Kumar, B.V.K., Song, S., Sims, S.R.F., & Epperson, J.F., "Unconstrained correlation filters," Applied Optics, vol. 33:3751-3759, 1994.

[13] Phillips, P.J., Grother, P., Micheals, R., Blackburn, D.M., Tabassi, E., & Bone, M., "Face recognition vendor test 2002," Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on, pp. 44-, Oct. 2003.

[14] Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Jin Chang, Hoffman, K., Marques, J., Jaesik Min, & Worek, W., "Overview of the face recognition grand challenge," CVPR 2005. IEEE Computer Society Conference on , vol.1, no.pp. 947- 954 vol. 1, June 2005.

[15] Sim, T., Baker, S., & Bsat, M., "The CMU Pose, Illumination, and Expression Database," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, pp. 1615-1618, Dec. 2003.

[16] Turk, M. A., & Pentland, A. P., "Face Recognition Using Eigenfaces," presented at Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91, IEEE Computer Society Conference, 1991.

[17] Vijaya Kumar, B.V.K., "Tutorial Survey of Composite Filter designs for Optical Correlators," Applied Optics, vol. 31, pp. 4773-4801, 1992.

[18] Vijaya Kumar, B.V. K., Carlson, D., & Mahalanobis, A., "Optimal tradeoff synthetic discriminant function (OTSDF) filters for arbitrary devices," Opt. Lett., vol. 19, pp. 1556–1558, 1994.

[19] Zhao, W., & Chellappa, R., "Robust Face Recognition Using Symmetric Shape-from-Shading," technical report, Center for Automation Research, Univ. of Maryland, 1999.