Hyperspectral Face Recognition using 3D-DCT and Partial Least Squares

Muhammad Uzair
uzair@csse.uwa.edu.au

Arif Mahmood
arif.mahmood@uwa.edu.au

Ajmal Mian
ajmal.mian@uwa.edu.au

Computer Science and Software
Engineering
The University of Western Australia
35 Stirling Highway, Crawley, WA,
Australia.

Abstract

Hyperspectral imaging offers new opportunities for inter-person facial discrimination. However, compact and discriminative feature extraction from high dimensional hyperspectral image cubes is a challenging task. We propose a spatio-spectral feature extraction method based on the 3D Discrete Cosine Transform (3D-DCT). The 3D-DCT optimally compacts information in the low frequency coefficients. Therefore, we represent each hyperspectral facial cube by a small number of low frequency DCT coefficients and formulate Partial Least Square (PLS) regression for accurate classification. The proposed algorithm is evaluated on three standard hyperspectral face databases. Experimental results show that the proposed algorithm outperforms five current state of the art hyperspectral face recognition algorithms by a significant margin.

1 Introduction

Despite vast research, face recognition is still a challenging problem because of the fact that intra-person dissimilarities often exceed the inter-person dissimilarities in the presence of view point and illumination variations. Most of the current research is based on the facial appearance captured by the 2D grayscale or RGB face images usually acquired in the visible spectrum [9, 22]. The inherent dimensionality of the 2D images is often very low limiting the discriminative capability of face recognition algorithms based on such data. Moreover, the spatial features extracted from 2D images, such as PCA, LBP or Gabor features, may perform poorly in the presence of view point or illumination variations resulting in low inter-person discrimination.

In order to increase the inherent dimensionality of the facial data, many researchers have considered the use of hyperspectral imaging [5, 11, 12, 14, 18]. Hyperspectral imaging can increase facial discrimination by capturing more biometric measurements such as the spectral response of faces. As the number of subjects to be discriminated increases, the performance of 2D image based algorithms decreases due to the reduction in the inter-subject distance in the face space. In such cases, hyperspectral imaging may be useful for increasing the accuracy because of its high dimensionality. A hyperspectral image is a data cube with two spatial and one spectral dimension. It is captured by a hyperspectral camera which
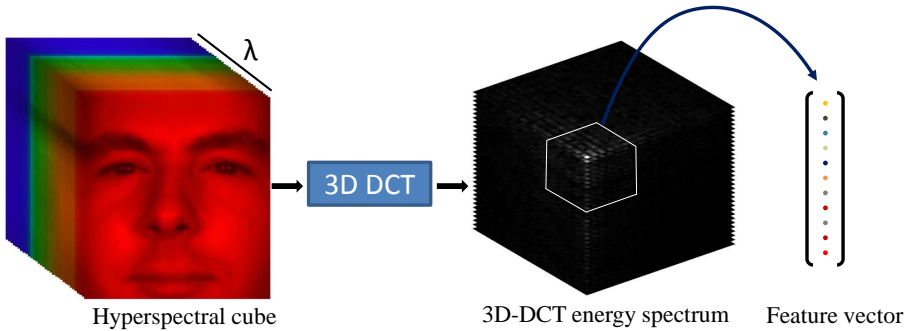
Figure 1: 3D-DCT based feature extraction. A hyperspectral cube (each band is rendered as RGB) and the corresponding 3D-DCT energy spectrum. It can be seen that only a few coefficients around the origin contain most of the energy.

operates in multiple narrow bands and densely samples the radiance measure in space and wavelength, producing precise optical spectra at every pixel. In addition to facial appearance, spectral measurements in multiple wavelengths also sense the subsurface tissue features which may be significantly different for each person. Pan et al. [11] used these subsurface features as spectral discriminants for recognition over a large range of poses and expressions. Hyperspectral imaging also offers liveliness test by distinguishing the spectral signature of a real human face from a synthetic mask or a mere photograph.

Although the high dimensionality of hyperspectral data presents a desirable feature for increasing the uniqueness of faces, at the same time it poses new challenges such as interband misalignments, low signal to noise ratios in certain bands and high dimensionality. Inter-band misalignment is due to subject movements during hyperspectral data acquisition. Low signal to noise ratio is usually observed in bands near the blue wavelength (420nm) due to low photon energy in these wavelengths. Therefore, these bands may contain relatively high noise. Moreover, high data dimensionality makes discriminative feature extraction a difficult task.

Due to the high dimensionality of hyperspectral data, discriminative feature extraction for face recognition is more challenging than 2D images. Most researchers have simply sub sampled the hyperspectral data [11, 12, 18] for dimensionality reduction and feature extraction while others have used simple PCA [5]. In contrast to the existing dimensionality reduction and feature extraction techniques, we propose the three dimensional Discrete Cosine Transform (3D-DCT) for feature extraction (Fig. 1). Exploiting the fact that hyperspectral data is usually highly correlated in spatial and spectral dimensions, a transform such as DCT is expected to perform information compaction in a few coefficients by providing maximal decorrelation. DCT transform being an approximation of the KL-Transformation optimally compacts the signal information in a given number of transform coefficients. Moreover, compared to other transforms, such as the Fourier transform, the transformed coefficients are real and thus require less data to process.

For the purpose of classifying the 3D-DCT features, we propose two types of classifiers, namely the Sparse Representation (SRC) [21] and Partial Least Square (PLS) regression. In our experiments, we observed that PLS regression performed better than SRC. This is because PLS basis projects the feature vectors into a latent space in which feature vectors corresponding to the same subject are closer than the feature vectors corresponding to dif-

ferent subjects. We perform experiments on three standard hyperspectral face databases including the PolyU Hyperspectral [5, 13], CMU Hyperspectral [4] and UWA Hyperspectral databases. The results are compared with five existing hyperspectral face recognition algorithms. The proposed algorithm has consistently outperformed the existing techniques on all three databases with significant margins.

1.1 Related Work

Hyperspectral face recognition is a relatively less explored research area. Existing work on hyperspectral face recognition can be divided into two broad categories. The first category of algorithms use only spectral features [11, 14] while the second category of algorithms use both spatial and spectral features.

Initial work on hyperspectral face recognition was done by Pan *et al*. [11] based on the spectral features using hyperspectral images in the near-infrared range (700-1000nm). They manually sampled the 31 dimensional (31 band) spectral signatures from hand picked locations on the hair, forehead, cheeks, lips and chin of each subject. The spectral signatures were compared using Mahalanobis distance to perform face recognition. The spectral features showed good recognition accuracy on a proprietary database of 200 subjects. However recognition rate dropped when the time lapse between probe and gallery image acquisition increased. Moreover, spectral features showed poor recognition accuracy on public hyperspectral face databases [17]. Robila [14] also used spectral features of different face regions in hyperspectral images of 120 bands (400-900nm) but compared them using spectral angle measurement. However, his experiments involved a small database of only eight subjects.

Pan *et al*. [12] extracted spatiospectral features, called the 2D spectral-face, by recursively sampling each subsequent pixel from the next band. This feature extraction method was simple; however, the sampling process of pixels to generate the 2D spectral-face was ad-hoc and about 96% of the pixels were discarded. Di *et al*. [5] used 2D PCA to extract low dimensional features from hyperspectral cubes. The 2D PCA based features were then compared using Euclidean distance to perform face recognition. However, the 2D PCA coefficients are sensitive to noise in the bands. Shen and Zheng [18] extracted spatiospectral features from each hyperspectral cube using 3D-Gabor Wavelets. However, their set of 52 3D-Gabor Wavelets transformed each hyperspectral cube to 52 feature cubes. Therefore, their method increased the dimensionality of features making feature selection more difficult. An ad-hoc sub-sampling scheme was applied on the feature cubes for feature selection.

2 Proposed Algorithm

The proposed algorithm has two steps, 3D-DCT based features extraction and Partial Least Squares (PLS) regression based classification. Both steps are separately discussed in the following sections.

2.1 3D-DCT based Feature Extraction

The Discrete Cosine Transform (DCT) [2] expresses a discrete signal, such as a 2D image or a hyperspectral cube, as a linear combination of mutually uncorrelated cosine basis functions [8]. DCT generates a compact energy spectrum of the signal where the low-frequency coefficients encode most of the signal information. A compact signal representation can be obtained by selecting only the low-frequency coefficients as features. Two dimensional DCT has previously been used for feature extraction from spatial data in applications such

as image compression [20], face recognition [2] and image retrieval [6]. Similarly, the 3D-DCT has been used to simultaneously exploit the spatiotemporal information contained in the correlated video frames in applications such as video coding [4, 15] and visual tracking [10]. However, we use 3D-DCT, for the first time, for information compaction and features extraction from hyperspectral data. The closest work in this regard is that of Abousleman *et al*. [1]. They used block based 3D-DCT for compression of hyperspectral images obtained by satellite sensors for remote sensing. In contrast, we use global 3D-DCT and extract features from the complete hyperspectral image cubes for face recognition. Our choice of global 3D-DCT for feature extraction is motivated from the fact that it can effectively exploit the high inter-band correlation and simultaneously model the spatiospectral information.

The 2D-DCT of a 2D image $h(x,y)_{N_1 \times N_2}$, and the 3D-DCT of a hyperspectral cube $H(x,y,\lambda)_{N_1 \times N_2 \times N_3}$ are given by

$$C(u,v) = \Omega_1(u)\Omega_2(v) \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} h(x,y)\{cos[\frac{\pi(2x+1)u}{2N_1}]cos[\frac{\pi(2y+1)v}{2N_2}]\} \qquad (1)$$

$$F(u,v,w) = \Omega_1(u)\Omega_2(v)\Omega_3(w) \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} \sum_{\lambda=0}^{N_3-1} H(x,y,\lambda)$$
$$\{cos[\frac{\pi(2x+1)u}{2N_1}]cos[\frac{\pi(2y+1)v}{2N_2}]cos[\frac{\pi(2\lambda+1)w}{2N_3}]\} \qquad (2)$$

where $u = \{0,1,...,N_1-1\}$, $v = \{0,1,...,N_2-1\}$, $w = \{0,1,...,N_3-1\}$ and $\Omega_i(u)$ is defined as

$$\Omega_i(u) = \begin{cases} \sqrt{\frac{1}{N_i}} & \text{if } u=0; \\ \sqrt{\frac{2}{N_i}} & \text{otherwise} \end{cases} \qquad (3)$$

The low frequency coefficients near the origin of $F(u,v,w)$ represent most of the energy of the hyperspectral cube, therefore, the high-frequency coefficients can be discarded. In order to construct our proposed feature vector, we sample a frequency sub-cube $\Gamma(u,v,w)$ of dimensions $(\alpha \times \beta \times \gamma)$ by retaining only the low-frequency elements around the origin of $F(u,v,w)$ i.e., $\{(u,v,w)|u \leq \alpha, v \leq \beta, w \leq \gamma\}$. The sub-cube $\Gamma(u,v,w)$ is vectorized and normalized to unit magnitude to obtain the final feature vector $f \in \mathcal{R}^d$, where $d = (\alpha\beta\gamma)$, which is used for classification. Figure 1 shows the process of our proposed 3D-DCT based feature extraction.

2.2   Classification by PLS Regression

Let c be the number of subject classes contained in the gallery and $G_f = \{f_{jk}\}_{j=1,k=1}^{n_k,c} \in \mathcal{R}^{d \times g}$ be the final gallery representation such that $n_k \geq 1$ are the number of hyperspectral cubes in each class, $g = \sum_{k=1}^c n_k$, and f_{jk} is a 3D-DCT based feature vector representing the jth hyperspectral cube in the kth class. We classify the features f_{jk} using Partial Least Squares (PLS) regression [16]. PLS models relations between sets of observed variables by means of latent variables. In its general form, PLS creates orthogonal score vectors by maximizing the covariance between different variable sets. Let the gallery G_f denote a set of predictor variables. We represent the response variables $Z = \{z_{jk}\}_{j=1,k=1}^{n_k,c}$ as a set of indicator vectors, where each indicator vector $z_{jk} \in \mathcal{R}^c$ shows the membership of kth class. z_{jk} is defined as a binary vector having 1 at the kth index and zeros otherwise.

PLS decomposes matrices G_f and Z into the form

$$G_f = AP' + E, \tag{4}$$

$$Z = BQ' + F, \tag{5}$$

where A and B are the matrices containing the extracted latent vectors, the matrices P and Q represent loadings, and the matrices E and F are the residuals. Based on the nonlinear iterative partial least squares (NIPALS) algorithm, PLS finds weight vectors p and q such that

$$\max_{|p|=|q|=1} [cov(G_f p, Zq)]^2 = [cov(a,b)]^2 \tag{6}$$

where a and b are the column vectors of A and B respectively and $cov(a,b)$ is the sample covariance. The regression coefficients between the two sets of variables G_f and Z can be estimated by PLS regression formulation [16]

$$W = G_f' B (A' G_f G_f' B)^{-1} A' Z. \tag{7}$$

Using W, we can predict labels of the query feature vector f_t:

$$\hat{z}_t = f_t' W, \tag{8}$$

where $\hat{z}_t \in \mathcal{R}^c$ is an indicator variable, ideally containing 1 at only one location (indicating the class membership) and 0 at all other locations. However due to noise in the data and approximation errors in the regression process, this is seldom the case. Often $\hat{z}_t$ contains some non-zero value at each location. In our implementation, the location of the maximum of $\hat{z}_t$ is considered as the predicted label for f_t.

3 Experiments and Results

The performance of the proposed algorithm using 3D-DCT based features and PLS based classification is compared with five existing hyperspectral face recognition algorithms on three hyperspectral face databases. In our experiments, the proposed algorithm has outperformed all of the existing algorithms by a significant margin.

3.1 Hyperspectral Face Databases

The first hyperspectral face database used is the Hong Kong Polytechnic University Hyperspectral Face Database(PolyU-HSFD) [5, 13] (Fig. 2). This database consists of hyperspectral face image cubes acquired by an indoor imaging system using a CRI's VariSpec Liquid Crystal Tuneable Filter. Each cube contains 33 bands acquired in the spectral range of 400-720nm with a step size of 10nm. The database was constructed over a long period of time and significant appearance variations of the subjects such as changes of hair style and skin condition are presented. Signal to noise ratio (SNR) in bands near the blue wavelength is very low, and the database contains inter-band alignment errors due to subject movements. The database contains 48 subjects with 13 females and 35 males. Each of the first 25 subjects has four to seven cubes while the remaining 23 subjects only have one cube per subject. Following the experimental protocol of [5, 18], we use the first 25 subjects in our experiments. Our final experimental data of PolyU consists of 113 hyperspectral image cubes. The first six and the last three bands have very low SNR, therefore these bands are discarded as
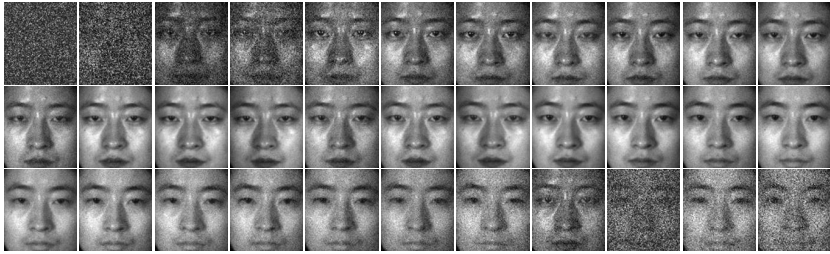
Figure 2: A hyperspectral face cube from the PolyU-HSFD dataset

recommended by [5, 18]. For each subject, two cubes are randomly selected for the gallery and the remaining 63 cubes are used as probes.

The second database is CMU Hyperspectral Face Database (CMU-HSFD)[4] (Fig. 3) which consists of hyperspectral face image cubes acquired with a prototype CMU developed spectro-polarimetric camera. Each hyperspectral face cube contains 65 bands covering the spectral range of 450-1100nm with a step size of 10nm. For illumination, three identical lamps with 600W halogen bulbs were used. The database contains 48 subjects, each subject has 4 to 20 cubes acquired in different sessions and different lighting combinations. The individual bands in the start of the cube have low SNR and most of the subjects perform eye blinking or small face movements during image acquisition. In our experiments we have used only the cubes acquired with all lights turned on. Our experimental data consists of 147 hyperspectral cubes of 48 subjects. Each subject has 1 to 5 cubes. The gallery is constructed by randomly selecting one cube for each of the 48 subjects while the remaining 99 cubes are used as probes.

The third database is the locally acquired UWA Hyperspectral Face Database (UWA-HSFD) (Fig. 4) which consists of hyperspectral face images acquired by an indoor imaging system using a CRI's VariSpec LCTF filter integrated with a Photon focus camera. Each hyperspectral face cube consists of 33 bands acquired in the spectral range of 400-720nm with a step size of 10nm. During acquisition, we adaptively selected the camera exposure time according to signal strength in different wavelengths. Therefore, noise levels in this database are relatively lower than the other two databases. Alignment errors are present between individual bands due to subjects movements and eye blinking during image acquisition. The UWA-HSFD database consists of 120 cubes of 70 subjects, each having one or two cubes. The gallery is constructed by randomly selecting one cube for each of the 70 subjects. The remaining 50 cubes are used as probes. The UWA-HSFD dataset will soon be made publicly available.

Manually clicked eye coordinates in each band are used to crop a normalized face image

Figure 3: A hyperspectral face cube from the CMU-HSFD (only odd numbered bands)

Figure 4: A hyperspectral face cube from locally acquired UWA-HSFD

of size 46×38, 40×40 and 30×30 pixels for PolyU, CMU and UWA HSFD respectively. For all the databases we performed ten-fold cross validation experiments by randomly selecting different gallery/probe combinations in each fold.

3.2 Compared Hyperspectral Face Recognition Algorithms

The five existing hyperspectral face recognition algorithms used for comparisons include Spectral Signature Matching [11], Spectral Angle Measurement [14], Spectral Eigenface [12], 2D PCA [5] 3D Gabor Wavelets [18]. We use our own implementations of all of these algorithms. The parameters of these algorithms are set as follows. For spectral signature matching algorithm we use 5 adjacent square regions of size 17×17 pixels arranged in a cross pattern to represent hair, forehead and cheeks. For lips we use square regions of size 9× 9 pixels [11]. For Spectral Eigenface we preserve 99% energy by retaining 48 PCA basis vectors for all three databases. For 2D PCA [5], 99% energy is preserved by retaining 27, 30 and 38 PCA basis vector for PolyU, CMU and UWA databases respectively. For the 3D Gabor method, we use 52 Gabor wavelets for feature generation as recommended by [18].

3.3 Results

In our experiments, we first filter the individual bands with a circular $(8,1)$ neighborhood LBP [19] filter to normalize the illumination variations. Figure 5 shows sub-cubes of size $(10 \times 10 \times 10)$ sampled from the full 3D-DCT energy spectrums of four different subjects from the UWA database. It can be seen that the energy patterns are very discriminative.

In order to choose the optimal number of 3D-DCT coefficients to construct the feature vector, we tested different values for the parameters α, β and γ. Figure 6 shows the relationship of the accuracy and the sub-cube size $\Gamma(u,v,w)$ to construct the feature vector. The best results are obtained by setting α, β and γ to 12 for PolyU, 10 for CMU and 10 for UWA database. We fixed the number of PLS basis to 45, 47, and 50 for PolyU, CMU and UWA databases respectively to obtain the best results (Fig. 6). We also perform experiments to
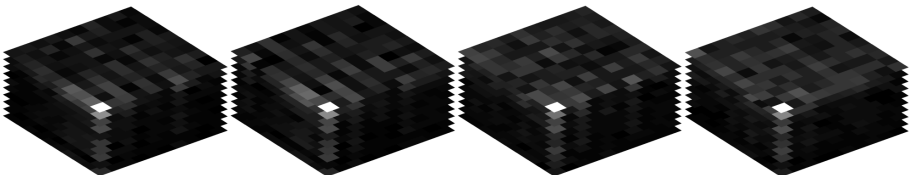

Figure 5: Sub-cubes of size $10 \times 10 \times 10$ sampled around the origin of the 3D-DCT energy spectrum of four different subjects of the UWA Database.
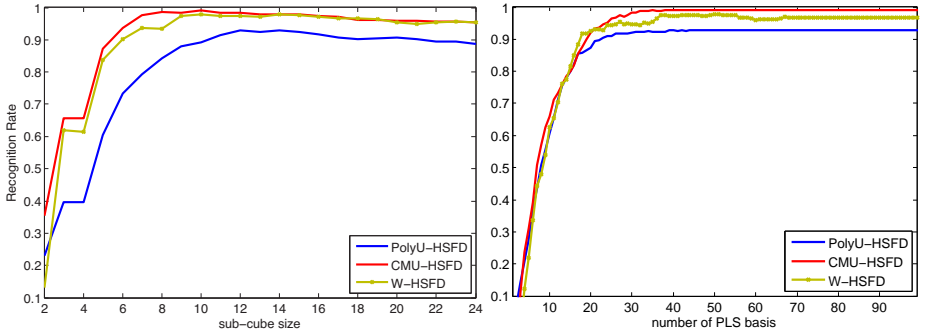
Figure 6: Proposed algorithm: Recognition rate vs. sub-cube size (left) Recognition rate vs. the number of PLS basis (right).

classify the 3D-DCT based features using Sparse Representation based Classification technique (SRC) [21]. For SRC the value of the regularization parameter λ is set to 4.7×10^{-4} for all three databases in our experiments. In addition, we also analyse the performance of using 2 dimensional Discrete Cosine Transform features for hyperspectral face recognition. The 2D-DCT is applied on individual bands according to (1) to obtain 2D-DCT energy spectrum of the individual bands. The low frequency coefficients from individual 2D-DCT energy spectrums are sampled and concatenated to form the final feature vector. SRC and PLS regression is used to classify the 2D-DCT features. The parameter setting is kept the same as that of 3D-DCT based feature extraction and classification.

Table 1 shows that the proposed algorithm has out-performed five existing hyperspectral face recognition algorithms on all three databases. We observe that PLS regression performed better than SRC. It is because PLS basis projects the feature vectors into a latent space in which feature vectors corresponding to the same subject are closer than the feature vectors corresponding to different subjects. Note that larger accuracy has been previously reported for spectral signature matching by Pan et al. [11] using hyperspectral face data in the NIR range (700-1000nm). In this spectral range, electromagnetic radiations has a larger penetration depth in the human skin as compared to electromagnetic radiations in the visible range. The NIR spectral measurements may sense the subsurface tissue features which are significantly different for each person resulting in higher accuracy. However, the database

Table 1: Average recognition rates and standard deviations (%) for ten fold experiments on three databases.

Algorithm	PolyU Database	CMU Database	UWA Database
Hyperspectral			
Spectral Signature [11]	24.63±3.87	38.18±1.89	40.52±1.08
Spectral Angle [14]	25.49±4.36	38.16±1.89	37.95±4.15
Spectral Eigenface [12]	70.30±3.61	84.54±3.78	91.51±3.07
2D PCA [8]	71.11±3.16	72.10±5.41	83.85±2.42
3D Gabor Wavelets [13]	90.19±2.09	91.67±2.86	91.50±3.07
Proposed			
2D-DCT + SRC	75.86±2.92	97.44±1.24	97.00±1.29
2D-DCT + PLS	91.43±2.10	97.78±1.28	97.25±1.87
3D-DCT + SRC	87.02±1.72	98.10±0.69	98.00±1.84
3D-DCT + PLS	**93.00±2.27**	**99.00±0.85**	**98.00±1.39**

used by Pan *et al.* [□] is not publicly available. Although, the spectral range of CMU-HSFD is 450-1100nm and includes the NIR range but the spectral features still does not give high accuracy on this database. Ryer reported similar recognition rate for spectral signature features using the CMU database in his PhD thesis [□].

4 Conclusion

In this paper we presented a 3D-DCT based feature extraction method for the representation of hyperspectral face images. We represented hyperspectral faces using compact feature vectors generated by sampling the low-frequency coefficients of the 3D-DCT energy spectrum. To perform face recognition the proposed features were classified using PLS regression. Experiments were performed on three hyperspectral datasets and the results of the proposed algorithm were compared with five existing hyperspectral face recognition algorithms. Experimental results showed that the proposed algorithm consistently outperforms the existing methods by significant margin.

5 Acknowledgements

References

[1] G.P. Abousleman, M.W. Marcellin, and B.R. Hunt. Compression of hyperspectral imagery using the 3-D DCT and hybrid DPCM/DCT. *IEEE Transactions on Geoscience and Remote Sensing*, 33(1):26–34, 1995.

[2] N. Ahmed, T. Natarajan, and K.R. Rao. Discrete Cosine Transform. *IEEE Transactions on Computers*, 1974.

[3] Nikola Bozinovic and Janusz Konrad. Motion analysis in 3D DCT domain and its application to video coding. *Signal Processing: Image Communication*, 2005.

[4] L. Denes, P. Metes, and Y. Liu. Hyperspectral face database. Technical Report CMU-RI-TR-02-25, 2002.

[5] Wei Di, Lei Zhang, David Zhang, and Quan Pan. Studies on hyperspectral face recognition in visible spectrum with feature band selection. *IEEE Trans. SMC-A*, 2010.

[6] Guocan Feng and Jianmin Jiang. JPEG compressed image retrieval via statistical features. *Pattern Recognition*, 36(4):977–985, 2003.

[7] Ziad M. Hafed and Martin D. Levine. Face recognition using the discrete cosine transform. *Int. Journal of Computer Vision*, 43(3):167–188, 2001.

[8] S. A. Khayam. The Discrete Cosine Transform (DCT): Theory and application. Technical Report Michigan State University, 2003.

[9] Stan Z. Li and Anil K. Jain, editors. *Handbook of Face Recognition, 2nd Edition.* Springer, 2011. ISBN 978-0-85729-931-4.

[10] Xi Li, Anthony Dick, Chunhua Shen, Anton van den Hengel, and Hanzi Wang. Incremental learning of 3d-dct compact representations for robust visual tracking. *IEEE TPAMI*, 35(4):863–881, 2013.

[11] Zhihong Pan, Glenn Healey, Manish Prasad, and Bruce Tromberg. Face recog. in hyperspectral images. *IEEE TPAMI*, 2003.

[12] Zhihong Pan, Glenn Healey, and Bruce Tromberg. Comparison of spectral-only and spectral/spatial face recog. for personal identity verification. *EURASIP JASP*, 2009.

[13] PolyU-HSFD. www4.comp.polyu.edu.hk/~biometrics/.

[14] Stefan A. Robila. Toward hyperspectral face recognition. In *Procceddings of SPIE ISTE Imaging*, 2008.

[15] John A. Roese, W. Pratt, and Guner S. Robinson. Interframe cosine transform image coding. *IEEE Transactions on Communications*, 25(11):1329–1339, 1977.

[16] R. Rosipal and N. Kramer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection Techniques, LNCS*, 2006.

[17] D. Ryer. *Quest Hierarchy for Hyperspectral Face Recognition*. PhD thesis, Air Force Institute of Tech., 2012.

[18] L. Shen and S. Zheng. Hyperspectral face recognition using 3d gabor wavelets. In *ICPR*, 2012.

[19] T. Ahonen, A. Hadid and M. Pietikainen. Face recognition with local binary patterns. In *ECCV*, 2004.

[20] Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, pages 30–44, 1991.

[21] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 2009.

[22] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 2003.