

FACE RECOGNITION USING ORTHO-DIFFUSION BASES

Sravan Gudivada and Adrian G. Bors

Dept. of Computer Science, University of York, York YO10 5GH, UK

E-mail: adrian.bors@york.ac.uk

ABSTRACT

This paper proposes a new approach for face recognition by representing inter-face variation using orthogonal decompositions with embedded diffusion. The modified Gram-Schmidt with pivoting the columns orthogonal decomposition, called also QR algorithm, is applied recursively to the covariance matrix of a set of images forming the training set. At each recursion a set of orthonormal bases functions are extracted for a specific scale. A diffusion step is embedded at each scale in the QR decomposition. The algorithm models the main variations of face features from the training set by preserving only the most significant bases while eliminating noise and non-essential features. Each face is represented by a weighted sum of such representative bases functions, called ortho-diffusion faces.

Index Terms— Diffusion wavelets, Face recognition, Gram-Schmidt orthogonal decomposition, Eigenfaces.

1. INTRODUCTION

A challenge in complex data analysis is when we have to recover a low dimensional intrinsic manifold which manifests itself through a very complex representation in the observable space. The use of kernels on undirected graphs was shown to lead to good results in machine learning tasks. Various approaches have been adopted for modeling data representations using diffusion processes on graphs [1, 2, 3]. Diffusion maps [2, 3, 4] achieve dimensionality reduction by reorganizing data according to the parametrization of its underlying geometry on orthogonal sub-spaces. When the diffusion is propagated, it integrates the local data structure to reveal relational properties of the data set at different scales [3, 4]. Maggioni and Mahadevan proposed in [5] a new multi-scale orthogonal decomposition on graphs into sets of bases functions and diffusion wavelets.

One of the classical face recognition methods was the eigenfaces method proposed by Turk and Pentland [6]. Each face is represented as a linear combination of the eigenvectors resulting from eigendecomposing the covariance matrix of a training set of faces. Extensions of this method include the Fisherfaces [7], the kernel PCA method [8] and the Laplacianfaces [9]. Orthogonal Laplacianfaces approach was proposed in [10] while a multilinear discriminant analysis was

employed in [11] for face recognition. Nevertheless, human identification depends on incorporating multi-modal human features such as the 3-D appearance and speech characteristics as well [12].

In this paper we propose to use orthonormal diffusion wavelet methodology for face recognition. The modified Gram-Schmidt algorithm with pivoting the columns QR , is used recursively for extracting a set of orthogonal basis functions, each representing an ortho-diffusion face. At each recursion, we consider data diffusion on the graph leading to the application of the QR algorithm in the following step on a dilated scale of the data. Ortho-diffusion bases which do not represent significant information are removed from further processing at each level. Each human face is defined by its projections onto the ortho-diffusion face space. In the testing stage, faces are classified according to the nearest neighborhood to the training data defined in the ortho-diffusion face space. The proposed ortho-diffusion faces method is described in Section 2. The face recognition method using ortho-diffusion faces is detailed in Section 3. In Section 4 we provide the experimental results while the conclusions are given in Section 5.

2. ORTHONORMAL DECOMPOSITIONS USING QR ALGORITHM

Let us consider a training set of M face images $\{\mathbf{I}_i, i = 1, \dots, M\}$, of size $m \times n$, and consider each of them as a vector of mn pixel entries. In the following we model the face variation within the given training set by calculating the deviation of each face from the mean face as in [6]:

$$\bar{\mathbf{I}} = \frac{1}{M} \sum_{i=1}^M \mathbf{I}_i, \quad (1)$$

where \mathbf{I}_i represents the mean face

$$\mathbf{S}_i = \mathbf{I}_i - \bar{\mathbf{I}}. \quad (2)$$

Each \mathbf{S}_i forms a column in a matrix \mathbf{A} of size $M \times mn$. The spread of the face variation within the training set is calculated by means of the covariance matrix \mathbf{C} :

$$\mathbf{C} = \mathbf{A}^T \mathbf{A} \quad (3)$$

where matrix \mathbf{C} is the diagonal covariance matrix of the training set, of size $M \times mn$, with each column corresponding to

the squared difference of pixels' value in a training set image from the mean face image. We apply the diffusion wavelet analysis onto \mathbf{C} representing the the face variation within the given data set.

In the following we describe how to use the orthogonal decompositions of diffusion wavelets in order to extract the ortho-diffusion faces from the matrix \mathbf{C} . A diffusion wavelet tree is produced by the recursive orthogonal decomposition of the data representation matrix \mathbf{C} into a set of diffusion scaling functions and their orthogonal wavelet functions at each scale j , [5]. The scaling functions at scale j span the subspace \mathcal{V}_j , while the wavelets span the space \mathcal{W}_j , representing the orthogonal complement of \mathcal{V}_j into the \mathcal{V}_{j+1} domain:

$$\mathcal{V}_{j+1} = \mathcal{V}_j \oplus^\perp \mathcal{W}_j \quad (4)$$

where the diffusion bases are characterized by $\mathcal{V}_{j+1} \subseteq \mathcal{V}_j$. There are three steps to construct the ortho-diffusion face at each scale: orthogonalization, operator compression and dilation. The diffusion scaling functions are smoothly bumped functions with some oscillations of scale 2^j , while the orthogonal wavelets are localized oscillatory functions representing the high frequency residuals at the same scale. In the following we look for modeling significant image features which are mostly represented by extended bases functions. Meanwhile, the wavelet functions represent random texture and noise information and are not of interest for further analysis. A diffusion wavelet tree is achieved by using dyadic powers \mathbf{C}^{2^j} , corresponding to dilations, which are used to create smoother and wider bumps functions employed for data analysis on the given graph structure.

The diffusion wavelet method represents a recursive application of the modified Gram-Schmidt with pivoting the columns (\mathcal{QR} algorithm) [5]. The \mathcal{QR} algorithm decomposes a given matrix into an orthogonal matrix \mathbf{Q} whose columns are orthonormal bases functions and a triangular matrix \mathbf{R} . In the following we use the notation $[\Phi_b]_{\Phi_a}$ for a matrix representing the base Φ_b with respect to Φ_a . We denote the triangular matrix by $[\mathbf{C}]_{\Phi_a}^{\Phi_b}$ whose column space is represented using bases Φ_a at scale a , while the row space is represented using bases Φ_b at scale b . The matrix \mathbf{C} is represented initially at the scale $j = 0$ on the basis set Φ_0 as $[\Phi_0]_{\Phi_0}$ and let us consider its columns as the set of functions $\Phi_0 = \{[\Phi_0]_{\Phi_0} \delta_k\}_k$ on the given graph, where δ_k is a set of Dirac functions. The \mathcal{QR} procedure decomposes $[\mathbf{C}]_{\Phi_0}^{\Phi_0}$ at the first level $j = 0$ as:

$$[\mathbf{C}]_{\Phi_0}^{\Phi_0} = [\mathbf{C}]_{\Phi_0}^{\Phi_1} [\Phi_1]_{\Phi_0} \quad (5)$$

\mathcal{QR} decompositions produces a linear transformation represented through the triangular matrix $[\mathbf{C}]_{\Phi_0}^{\Phi_1}$, which is the transformation of the matrix \mathbf{C} , whose columns are represented with respect to the base Φ_1 while the rows are represented with respect to Φ_0 , and the orthonormal matrix $[\Phi_1]_{\Phi_0}$ representing the base $[\Phi_1]$ with respect to $[\Phi_0]$.

We consider that the \mathcal{QR} decomposition is applied recursively, where at each scale j , the base Φ_{j+1} replaces Φ_j ,

starting with $j = 0$, as in equation (5). The orthonormal bases functions spanning Φ_j and representing columns in $[\Phi_j]_{\Phi_j}$ represent characteristic features of the given face image set. The \mathcal{QR} decomposition is followed by a data reduction step. The columns, at a scale $j > 0$, whose norms are smaller than a given precision ϵ , are eliminated due to their low contribution to the data representation. Henceforth, we consider for further processing only the columns that carry the important data representation on the given diffused graph:

$$\tilde{\Phi}_j = \{ \|\Phi_{j+1}\}_{\Phi_j} \delta_k \mid \|\Phi_{j+1}\}_{\Phi_j} \delta_k > \epsilon \}_k, \quad (6)$$

where $\|\cdot\|$ represents the norm of the column extracted by the Dirac function δ_k . When removing a column from $[\Phi_{j+1}]_{\Phi_j}$, we remove its corresponding row from $[\mathbf{C}^{2^{j+1}}]_{\Phi_j}^{\Phi_{j+1}}$, as well. For enforcing the sparseness of the triangular matrix $[\mathbf{C}^{2^{j+1}}]_{\Phi_j}^{\Phi_{j+1}}$ we neglect all entries which are smaller than a threshold θ and consider them as zero henceforth. Column pivoting is employed as well when the given matrix is nearly rank deficient.

The decomposition and data reduction stages described above are followed by data representation dilations on the graph defined as $[\mathbf{C}^{2^j}]_{\Phi_j}^{\Phi_{j+1}}$ for the scale j . At the scale $j + 1$, we square the operator to obtain the dilation $[\mathbf{C}^{2^{j+1}}]_{\Phi_{j+1}}^{\Phi_{j+1}}$ by using:

$$\begin{aligned} [\mathbf{C}^{2^{j+1}}]_{\Phi_{j+1}}^{\Phi_{j+1}} &= ([\mathbf{C}^{2^j}]_{\Phi_j}^{\Phi_{j+1}} [\Phi_{j+1}]_{\Phi_j})^2 \\ &= [\mathbf{C}^{2^j}]_{\Phi_j}^{\Phi_{j+1}} ([\mathbf{C}^{2^j}]_{\Phi_j}^{\Phi_{j+1}})^\tau. \end{aligned} \quad (7)$$

This corresponds to implementing a diffusion on the data representation on the scale 2^j and reprojecting the given manifold data to the scale 2^{j+1} .

The extended base $[\Phi_{j+1}]_{\Phi_0}$ is calculated at each recursion $j + 1$ with respect to the initial base Φ_0 as in the following:

$$[\Phi_{j+1}]_{\Phi_0} = [\Phi_{j+1}]_{\Phi_j} [\Phi_j]_{\Phi_0}. \quad (8)$$

At each scale $j + 1$, the number of basis functions decreases and the matrix $[\Phi_{j+1}]_{\Phi_j}$ becomes smaller. Each basis function, representing a column of $[\Phi_{j+1}]_{\Phi_j}$, can be used to represent image features in a low resolution image, while the extended base function, representing a column of $[\Phi_{j+1}]_{\Phi_0}$, maps these features to the size of the original image. At the scale $j + 1$, the representation of $\mathbf{C}^{2^{j+1}}$ is compressed based on the amount of the remaining underlying data and the desired precision ϵ .

3. FACE RECOGNITION USING ORTHO-DIFFUSION FACES

Each face image contributes more or less to each column vector in the extended base representation $[\Phi_j]_{\Phi_0}$ and this can be mapped into a set of ortho-diffusion faces. The ortho-diffusion faces are obtained by using a similar approach to the diffusion maps from [2, 3, 4], but using orthonormal extended bases functions as produced by the methodology described above, instead of the eigenvectors of the Laplacian

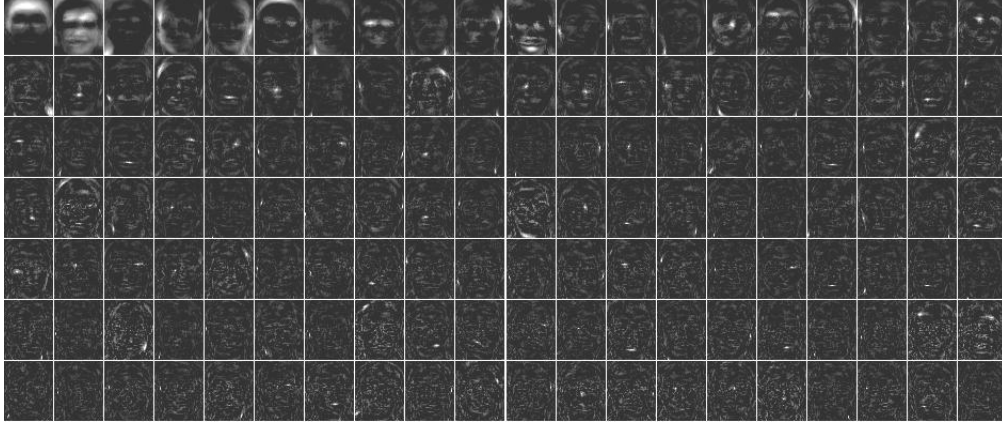


Fig. 1. Representation of 140 out of a total 198 of ortho-diffusion faces used for the ORL database face classification, where we add the value of 50 to the graylevel value of each pixel for better visualization.

matrix. The diffusion distance metric calculated based on the extended basis functions at the level j between two image pixels at locations i and k is given by:

$$D_{\Phi_j}(i, k) = \sqrt{(\Phi_j(i) - \Phi_j(k))^\tau (\Phi_j(i) - \Phi_j(k))} \quad (9)$$

where we consider all the extended bases functions defined on the graph between the two given pixels. The ortho-diffusion face is obtained by calculating the feature representation at a certain image location by summing up the diffusion distances from all the other image pixels:

$$\Delta_{i,j} = \sum_{k=1}^{mn} D_{\Phi_j}(i, k). \quad (10)$$

After calculating these distances for $i = \{1, \dots, mn\}$ we can have an approximate reconstruction of a ghost face.

A set of column vectors from $\mathbf{Y} = \{[\Phi_j]_{\Phi_0} \delta_k\}_k$ for $k = 1, \dots, N$ can be considered as a set of features that jointly characterizes the variation of face images in the given training set. We calculate the weight vector Ω_i of the face \mathbf{I}_i , by projecting the training face difference from the average face, denoted as \mathbf{S}_i from (2), onto the given ortho-diffusion face space:

$$\Omega_i = \mathbf{Y}^\tau \mathbf{S}_i \quad (11)$$

where $i = 1, \dots, M$ characterize all face images from the training set. The weights Ω_i define the coordinates for each image \mathbf{I}_i in the ortho-diffusion feature space of N feature vectors. We can approximate a face by using the ortho-diffusion faces as:

$$\hat{\mathbf{I}}_k = \bar{\mathbf{I}} + \mathbf{Y} \Omega_k \quad (12)$$

Where, $\hat{\mathbf{I}}_k$ is the reconstructed face and Ω_k represent its weights in the given ortho-diffusion face space.

For a given image face \mathbf{I}_t , which is not in the training set, we can calculate its corresponding weights in the given ortho-diffusion space. Face recognition is performed according to the minimum Euclidean distance in the space defined by the

ortho-normal faces, representing a nearest neighbor classification in the ortho-diffusion face space:

$$\arg \min_{k=1}^M [(\Omega_t - \Omega_k)^\tau (\Omega_t - \Omega_k)] \quad (13)$$

where we assume known the person identity for all M training faces.

4. EXPERIMENTAL RESULTS

In the following we apply the ortho-diffusion wavelets for face recognition on the ORL and Yale face databases. We consider the precision for eliminating non-essential bases functions in the \mathcal{QR} algorithm as $\epsilon = 10^{-6}$, while the sparseness threshold for each element in the triangular matrix is chosen as $\theta = 2.2 \times 10^{-16}$. The ORL database¹ contains 40 subjects, with 10 images for each subject. The images were taken at different times, varying the lighting, facial expressions and with small variations in the face orientation. The images are resized to 56×46 from their original size of 112×96 in order to reduce the computational complexity. Fig. 2 displays all the images from the ORL database. We consider a training set of five faces for each person and this results into an initial data representation matrix of size 2576×200 . After applying the proposed ortho-diffusion decomposition methodology, each extended basis function is represented as an ortho-diffusion face. Fig. 1 displays the first 140 ortho-diffusion faces, from the total of 198 which have been found as significant by the proposed ortho-diffusion methodology in order to represent the given set of training faces. It can be observed from Fig. 1, that the ortho-diffusion faces represent various face features characteristic to the training set. The weights representing the mapping of each face in this ortho-diffusion face space are calculated for each face using equation (11).

¹The ORL database was obtained from <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

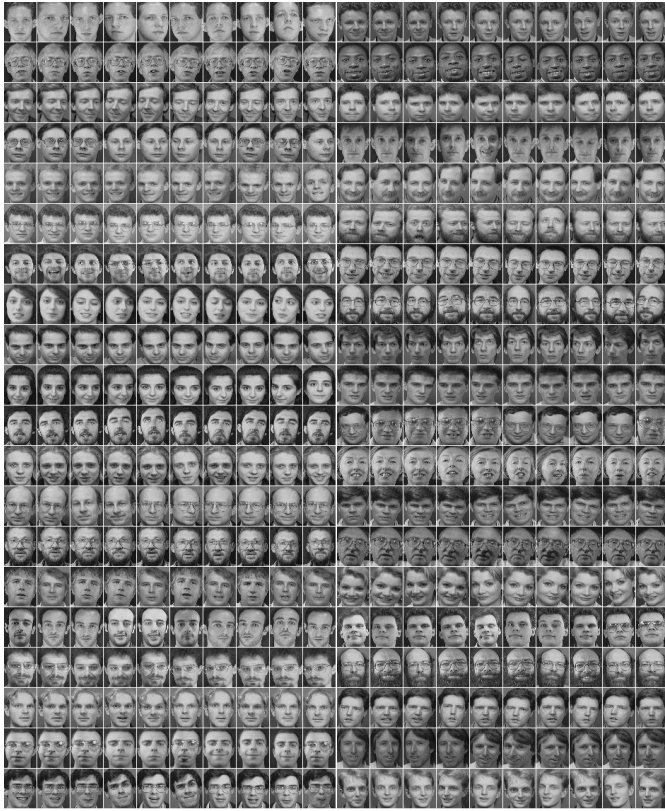


Fig. 2. The face image data from the ORL database.

We reconstruct the whole database of 400 faces by using 198 ortho-diffusion faces according to equation (12). The reconstructed face set is shown in Fig. 3. and we can observe that each face is reconstructed up to a certain precision. A human observer can identify correctly the individual represented in each reconstructed face. When considering the training set of five faces for each human subject, we classify the other five based on nearest neighbor classification in the ortho-diffusion face space. Only 18 faces have not been classified correctly, when considering for training the first five faces for each subject in the ORL database and these are shown within black contours in Fig. 3. It can be observed that these faces display significant changes with respect to those representing the same human subjects from the training set.

In the following we vary the size of the training set and evaluate the classification results on the remaining faces. The face classification results are displayed in Fig. 4. Under each test result we specify the number of ortho-diffusion faces used for defining the ortho-diffusion face space. We can observe that when increasing the size of training set we can improve the recognition results up to the rate of 96.25% classification accuracy when using 80% of database for training and the rest of 20% for testing.

The Yale database² contains 165 images of 15 subjects,

²Yale database was taken from

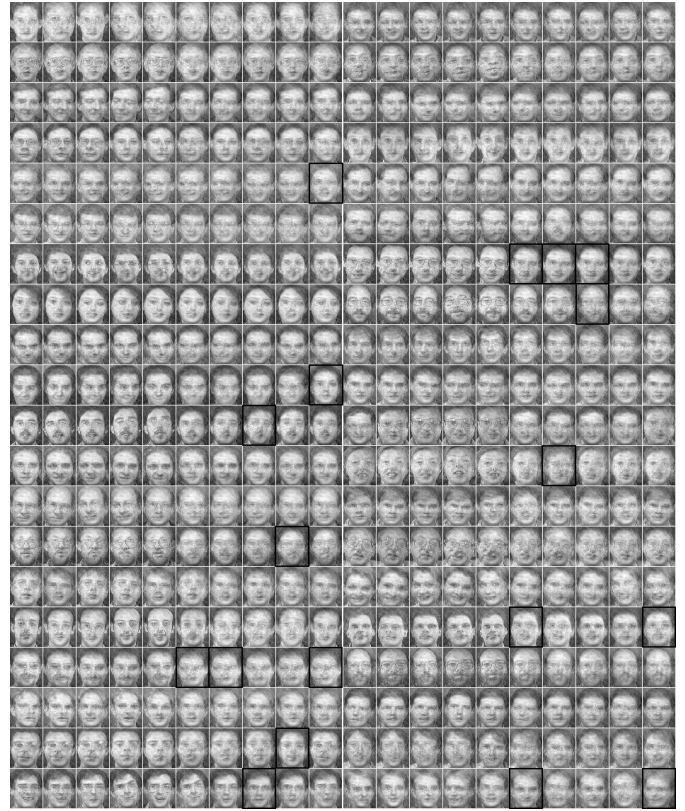


Fig. 3. Reconstructed faces from the ORL database.

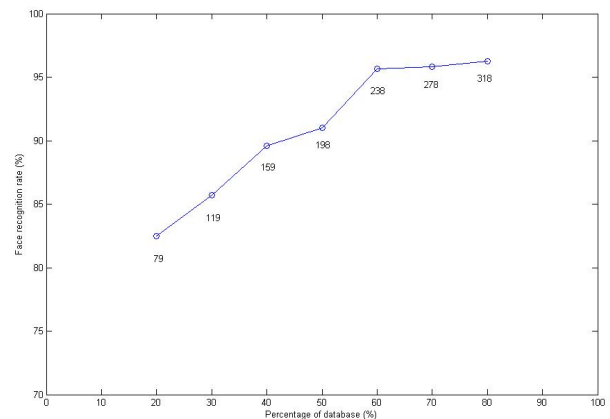


Fig. 4. Face recognition rate when varying the training set size as a percentage of the entire ORL database. The number of ortho-diffusion faces is indicated for each experiment.

which are shown in Fig. 5(a). There are 11 images per subject shot when varying facial expressions, with or without spectacles and within a wide variation of lighting. We can observe that the 4th and 7th sample face images for each individual have significant lighting variation. We consider five faces for each subject for the training stage. We evaluate the covariance

<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

matrix of the training set and then apply the ortho-diffusion face methodology from Section 3. The reconstructed faces from the Yale database using the ortho-diffusion faces are shown in Fig. 5(b). Faces which are not recognized in this case are represented within black contours in Fig. 5(b). We can observe that 6 out of the 8 wrongly classified faces are characterized by significant illumination changes from those characteristic to the training set.

In the following we consider 20 different sets of training faces or the ORL database and 15 different training sets for the Yale database, which are chosen based on different combinations of 50% faces for training and the rest for testing. The average of the classification rates for the proposed methodology, Eigenfaces [6], Fisherfaces [7], Laplacianfaces [9] and Orthogonal Laplacian faces [10] are provided in Table 1. The experiments have been conducted under the same conditions and when resizing all faces to 56×46 pixels.

Table 1. Comparison of various face recognition methods for ORL and Yale databases.

Method	ORL Database (%)	Yale Database (%)
Diffusion Wavelet	93.95	87.56
Eigenfaces	93.62	86.93
Fisherfaces	92.02	94.84
Laplacianfaces	91.70	94.22
Ortho-Laplacianfaces	93.60	86.93

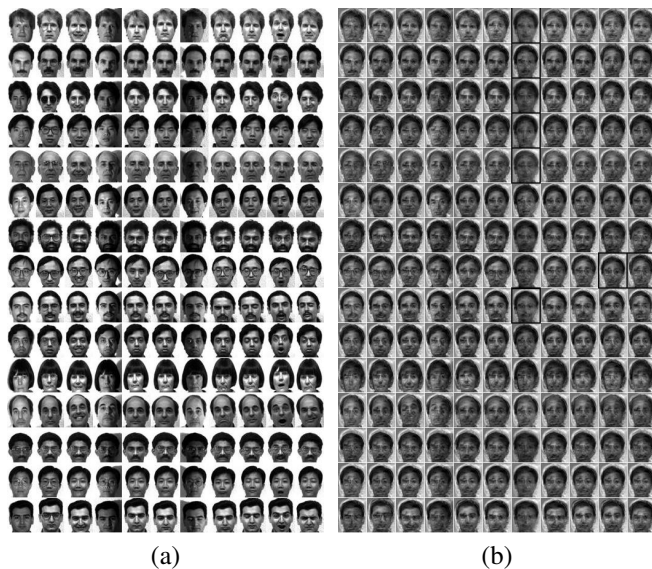


Fig. 5. Images from the Yale face database. (a) Original data set. (b) Reconstructed faces from the Yale database. Faces which are not recognized when considering a training set of 75 faces are shown within black contours.

5. CONCLUSION

In this paper we describe a new face recognition approach based on graph orthonormal decompositions. The proposed ortho-diffusion face decomposition method is based on the modified Gram-Schmidt algorithm with pivoting the columns. The inter-face variation matrix for a given training set is recursively processed using orthonormalization, reduction and dilation stages. A set of representative orthonormal extended basis functions are extracted from the given data set. Each of these extended bases functions represent an ortho-diffusion face and face classification takes place using the nearest neighbor rule in the ortho-diffusion face space.

6. REFERENCES

- [1] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [2] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.
- [3] A. Singer and R. R. Coifman, "Non-linear independent component analysis with diffusion maps," *Applied Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 226–239, 2008.
- [4] S. Wartak and A. G. Bors, "Optical flow estimation using diffusion distances," in *Proc. of International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 189–192.
- [5] M. Magioni and S. Mahadevan, "Fast direct policy evaluation using multiscale analysis of markov diffusion processes," in *Proc. of the 23rd International Conference on Machine Learning*, 2005, pp. 601–608.
- [6] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [7] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [8] C. Liu, "Gabor-based kernel pca with fractional power polynomial models for face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572–581, 2004.
- [9] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [10] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans on Image Processing*, vol. 15, no. 11, pp. 3608–3614, 2006.
- [11] S. Yan, D. Yu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans on Image Processing*, vol. 16, no. 1, pp. 212–220, 2007.
- [12] V. Chatzis, A. G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication," *IEEE Trans on Systems, Man and Cybernetics - Part A: Systems and Humans*, vol. 29, no. 6, pp. 674–681, 1999.