

On the Use of Monogenic Scale Space for Efficient Face Representation and Recognition

M. Sharmila Kumari¹ and B.H. Shekar²

¹ Department of Computer Science and Engineering, P. A. College of Engineering, Mangalore, Karnataka, India

{sharmilabp,bhshekar}@gmail.com

² Department of Computer Science, Mangalore University, Karnataka, India

Abstract. In this paper, we present a novel monogenic scale space based Principal Component Analysis (PCA) method by integrating the Reisz transform of face images at different scales and the PCA method for face representation and recognition. The Reisz transform captures desirable facial features characterized by local phase information and local energy at different scales in order to cope with the variations due to illumination and facial pose changes. The PCA method is then employed to reduce the dimension of the feature vectors and hence for efficient face representation and recognition. The feasibility of the proposed monogenic scale space based method integrated with PCA has been successfully tested on many standard face databases such as AT&T and YALE face databases. The recognition accuracy of the proposed approach is compared with the other well known face recognition approaches namely the PCA method, the kernel PCA method and the Gabor wavelet-based RCM method and it is found that the proposed approach exhibit better recognition accuracy when compared to these well known methods.

Keywords: log-Gabor Transform, Riesz transform, Monogenic scale space, Principal component analysis, Face recognition.

1 Introduction

In these days, the biometrics is the well addressed research area in the domain of computer vision and we have seen plethora of algorithms on face based biometrics because of its wide acceptability in several applications ranging from access control, identity authentication, and visual surveillance to human-computer interaction/communication. Devising an efficient face recognition algorithm is quite challenging because of many inherent problems such as inter-class similarity and intra-class variability, occlusion, variations in illuminations, pose changes etc. In order to address these problems, efforts are made by many researchers that result in different classes of algorithms. The local descriptor based algorithms gain much importance because of their robustness to noise and occlusion-the common problems encountered in real recognition environment. The transform based algorithms are most useful in face recognition as they are capable of withstanding illumination problems.

The local descriptors [8], [14], [22] are commonly employed in a number of real-world applications such as object recognition [5], [14] and image retrieval [16] as they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint. Mikolaczyk and Schmid [15] presented a comparative study of several local descriptors including steerable filters [7], differential invariants [10], moment invariants [22], Scale Invariant Feature Transform [13], and cross-correlation of different types of interest points [8], [16]. Their experiments showed that the ranking of accuracy for the different algorithms was relatively insensitive to the method employed to find interest points in the image but was dependent on the representation used to model the image patch around the interest point. In the SIFT algorithm, each keypoint is represented by its neighborhood, described as a set of orientation histograms computed from the gradient image. The SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change (in a narrow range). The most common implementation uses 16 histograms of 8 bins (8 orientations), which gives a 128 dimensional descriptor. Ke and Sukthankar [9] proposed PCA-SIFT descriptor which is also based on the gradient image, the main difference with SIFT being the further compression using PCA. Recently, the Speeded Up Robust Features (SURF) descriptor [1] has appeared as an alternative to SIFT. Its main advantage is its fastest computation, while keeping a high descriptive power. It is partially inspired by SIFT, but instead of using the gradient image, it computes first order Haar wavelet responses. The Local Energy based Shape Histogram (LESH) has been specifically designed for face recognition applications. Its goal is to encode the underlying shape present in the image. Basically, the descriptor is a concatenation of histograms obtained by accumulating local energy at several filter orientations.

On the other hand, we have seen appearance based approaches which are capable of withstanding noise and illumination and simple in terms of implementation. In these approaches, data transformation is a fundamental step and the goal is to obtain highly discriminative lower-dimensional data from high-dimensional data. Principal component analysis (PCA) and linear discriminate analysis (LDA) are the widely used techniques in the face recognition domain, which encode high-dimensional face images as lower dimensional eigenfaces [20] and fisherfaces [2] respectively. PCA is a linear method that ensures that the data transformed are uncorrelated and preserve maximally the second order statistics of the original data, and hence is insensitive to the dependencies of multiple features in the patterns. To overcome this problem, kernel PCA [19] is proposed as a non-linear extension of PCA that computes the principal components in a high-dimensional feature space. On the similar line, kernel FLD is proposed as a non-linear extension to linear discriminant analysis.

We have also seen the frequency domain based approaches where high frequency components are used as facial features because of their robustness to illumination changes. The local phase information is used in some techniques as it is proved to be sufficient to completely reconstruct a signal within a scale factor [17]. As the Gabor wavelet representation captures salient visual properties

such as spatial localization, orientation selectivity, Lades et al [11] applied Gabor wavelets for face recognition using dynamic link architecture framework. Wiskott et al. [23] extended this basic frame work to devise Gabor wavelet based elastic bunch graph matching method to label and recognize faces. Liu [12] proposed Gabor-based kernel PCA method by integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition. Porikli and Tuzel [18] and Tuzel et al.[12] proposed a new descriptor framework called region covariance matrices (RCM) for object detection and tracking. The RCMs can be categorized as a matrix-form feature. However, direct application of RCM to face recognition has not produced satisfactory results and hence Yanwei et al., [24] introduced Gabor-based region covariance matrices as face descriptors. The results of Yanwei et al. [24] are highly encouraging when compared to any other Gabor wavelet based techniques. Although Gabor wavelet based face recognition models possess very high recognition accuracy irrespective of noise, occlusion and illumination problems, massive computing and space requirements are the major bottleneck in these approaches. In this context, we have proposed a new frame work for face recognition based on monogenic scale space approach. The monogenic signal requires fewer convolutions when compared to Gabor wavelet based model and hence require much lesser computing time. The principal component analysis is employed on monogenic scale space based images to obtain compact representation of face images.

The remaining part of the paper is organized as follows. In section 2, we discussed monogenic scale space and its application in the context of face recognition is analyzed. The PCA based face representation in monogenic scale space is given in section 3. The experimental results and comparative study is given in section 4. Conclusion is reached in section 5.

2 Monogenic Scale Space: A Review

The monogenic signal [3], [4] is based on the Riesz transform which is used instead of the Hilbert transform. The monogenic signal analysis [3], [4] is a framework to interpret images in terms of the local phase, local orientation and local energy. The monogenic signal is an effective tool to analyze 2-D signals in a rotation invariant manner. The signal is built upon the first order Riesz transform. The spatial representation of the Riesz kernel in 2D space is:

$$(R_x(\mathbf{x}), R_y(\mathbf{x})) = \left(\frac{x}{2\pi |\mathbf{x}|^3}, \frac{y}{2\pi |\mathbf{x}|^3} \right), \mathbf{x} = (x, y) \in \mathbb{R}^2 \quad (1)$$

and its transfer function in the Fourier domain is:

$$(F_u(\mathbf{u}), F_v(\mathbf{u})) = \left(-i \frac{u}{|\mathbf{u}|}, -i \frac{v}{|\mathbf{u}|} \right), \mathbf{u} = (u, v) \in \mathbb{R}^2 \quad (2)$$

For any image, say $I(x)$, the monogenic signal is defined as the combination of I and its Riesz transform:

$$i.e. I_m(\mathbf{x}) = (I(\mathbf{x}), R_x \{I\}(\mathbf{x}), R_y \{I\}(\mathbf{x}))$$



Fig. 1. Original image; log-Gabor image; Energy image and Orientation image

$$= (I, R_x * I, R_y * I) \tag{3}$$

where $*$ stands for the convolution operation. And hence the local orientation is calculated as [3, 4]:

$$\theta = \arctan \frac{R_y \{ \mathbf{I} \}}{R_x \{ \mathbf{I} \}}, \theta \in [0, \pi) \tag{4}$$

The local phase is defined as:

$$\phi = \arctan 2 \left(\sqrt{R_x^2 \{ \mathbf{I} \} + R_y^2 \{ \mathbf{I} \}}, \mathbf{I} \right), \phi \in [0, \pi) \tag{5}$$

The local energy is defined as:

$$E = \left(\sqrt{R^2 \{ \mathbf{I} \} + R_x^2 \{ \mathbf{I} \} + R_y^2 \{ \mathbf{I} \}}, \mathbf{I} \right) \tag{6}$$

where $R^2 \{ \mathbf{I} \} = I * F^{-1}(G(w))$. Here $G(w)$ is the log-Gabor filter in the Fourier domain. Since log-Gabor filters are band-pass filters, usually multi-scale monogenic representation is required to fully describe a signal. In Fig 1, we have given the convolved images in monogenic scale space filter showing the log-Gabor transformed image with its energy and orientation images. One can see that the local structure is well captured in monogenic components.

3 Face Representation Using PCA in Monogenic Scale Space

Let there be N number of training images. Let $A_i, i=1...N$, be an image of size $m \times n$. Let each image $A_i, i=1...N$, has p number of local phase and local energy captured images. These $2p$ number of images are concatenated to form $2pm \times n$ dimensional feature vector. Hence, we are having $2p(m \times n) \times N$ sized training matrix called U . Let \bar{A} be the average image of all the images: $2p \times N$. Let U be made as mean centred. i.e.,

$$U = [U_{1,1} - \bar{A}, U_{1,2} - \bar{A}, \dots, U_{1,K_N} - \bar{A}] \tag{7}$$

where $U_{1,1}$ represents the first training image obtained by both local phase and local energy at different scales, $U_{2,1}$ represents the second training image obtained by both local phase and local energy at different scales and so on. The training matrix \mathbf{U} given in Eq. (7) contains $2pmn$ -dimensional feature vector which are mean centered. Unlike appearance based models where intensity values are used as feature vectors, here we have considered local energy and local phase information associated with each image as a feature vector. However, a large number of training samples are often needed to get reliable and robust estimation about the characteristics of data distribution which is often called as *curse of dimensionality*. The eigenvectors, e_i and the corresponding eigenvalues λ_i of \mathbf{U} are determined by solving the well-known eigen-structure decomposition problem:

$$\lambda_i e_i = \mathbf{U} e_i \quad (8)$$

Though all the eigenvectors are needed for accurate recognition, only a small number, q is generally sufficient for capturing the primary characteristics of the feature descriptors. The q eigenvectors, corresponding to the q largest eigenvalues, constitute the monogenic eigenspace. Thus monogenic eigenspace analysis can drastically reduce the dimension $2pmn$ to the monogenic eigenspace dimension q while keeping several of the most effective features that summarize the original information.

3.1 Feature Extraction

The optimal projection axes: $\lambda_1, \lambda_2, \dots, \lambda_q$ are used for feature extraction. Given an image A , compute the local phase and local energy images, S of the image A and project the local phase and local energy images onto the optimal projection axes λ_s that results in a feature matrix. That is,

$$F = S\lambda_j, \forall j = 1..q \quad (9)$$

The above process of projection of training images onto monogenic eigenspace is to be repeated for all the training images to create a knowledge base.

3.2 Face Recognition

Let I be an image given for recognition. Let I' be the local phase and local energy captured image which is projected onto the q number of optimal projection axes λ_s that results in test image feature matrix say, T computed by $T = I' \lambda$. Given two images, say γ_{i_1} and γ_{i_2} of any two face(s), represented by feature vectors: $r = [r_1, r_2, \dots, r_q]$ and $s = [s_1, s_2, \dots, s_q]$, the $dist(r, s)$ is defined as

$$dist(r, s) = \sum_{j=1}^q \|r_j - s_j\| \quad (10)$$

where $\|a - b\|_2$ denotes the Euclidean distance between the two vectors a and b . For classifying a given test image, the nearest neighbour is identified among the training samples and the corresponding class label of nearest neighbour training sample is attached to the image under testing.

4 Experimental Results

This section presents the results of the experiments conducted to corroborate the success of the proposed model. We have conducted experimentation on AT&T and YALE face datasets. We have specifically chosen this dataset as these are used by many researchers as a benchmark dataset to verify the validity of their proposed face recognition models. All experiments are performed on a P-IV 2.99GHz Windows machine with 8GB of RAM.

Experimentation on AT&T face dataset: The AT&T face dataset contains images from 40 individuals, each providing 10 different images of size 112x92. In our experiment, we have considered alternate five samples per class for training and the remaining samples for testing. Similarly, we have conducted experiments considering 160 faces as training faces of the AT&T database choosing remaining 6 faces from each person as test faces and the recognition performance has been obtained considering the remaining faces as test faces. The recognition accuracy of the proposed approach with varying dimension of feature vectors is given in Table 1.

Table 1. Recognition accuracy of the proposed approach and other techniques on AT&T Face database

Algorithm	No. of Training samples	Dimension of feature vector				
		30	35	40	45	50
Eigenface	200	88	87.5	86.5	86.5	86.5
	160	83.33	80.41	80.41	81.25	81.25
Kernel Eigenface	200	88	88	88	88	88
	160	87.5	87.5	87.5	87.5	87.5
Proposed approach	200	94.5	95	96	96	96.5
	160	95	95	95	95.41	95.41
Gabor	200	86.50 – Here the dimension of the feature vector is : 1600				
Wavelet	160	84.00 – Here the dimension of the feature vector is : 1600				

Table 2. Recognition accuracy of the proposed approach and other techniques on YALE Face database

Algorithm	No. of Training samples	Dimension of feature vector				
		30	35	40	45	50
Eigenface	75	86.66	86.66	86.66	86.66	86.66
	90	82.66	82.66	84	84	82.66
Kernel Eigenface	75	85.55	85.55	85.55	85.55	85.55
	90	82.66	82.66	82.66	82.66	82.66
Proposed approach	75	91.11	92.22	92.22	92.22	92.22
	90	89.33	89.33	90.66	90.66	90.66
Gabor	75	85.89 – Here the dimension of the feature vector is : 1600				
Wavelet	90	84.30 – Here the dimension of the feature vector is : 1600				

Experimentation on YALE face dataset: The YALE face dataset contains 165 images of 15 subjects that include variation in both facial expression and lighting. The training set comprised of five and six images randomly chosen for each person with remaining number of face images for each person. The recognition accuracy of the proposed approach with varying dimension of feature vectors is given in Table 2. In Tables 1 and 2, we have also provided the recognition accuracy of the standard eigenface approach, kernel PCA and Gabor wavelet based RCM techniques. It shall be observed from Table 1 and 2 that the proposed approach possess best recognition accuracy when compared to the existing approaches.

5 Conclusion

We have presented a new face recognition algorithm within the frame work of monogenic scale space that captures the local energy and local phase information at different scales. For the compact representation of the phase and energy contained images, principal component analysis is employed and there by crated the knowledge base. Experimental results on the standard benchmark face databases and comparative study with the well known face recognition approaches reveal the superiority of the proposed approach for face recognition problems and its suitability in real environment.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
3. Felsberg, M., Sommer, G.: The monogenic scale-space: A unifying approach to phase-based image processing in scalespace. *Journal of Mathematical Imaging and Vision* 21(1), 5–26 (2004)
4. Felsberg, M., Sommer, G.: The monogenic signal. *IEEE Trans. SP* 49, 3136–3144 (2001)
5. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *Proceedings of Computer Vision and Pattern Recognition* (June 2003)
6. Forstner, W., Moonen, B.: A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University (1999)
7. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. *IEEE Trans. Pattern Analysis and Machine Intelligence* 13(9), 891–906 (1991)
8. Harris, C., Stephens, M.: A combined corner and edge detector. In: *Alvey Vision Conference*, pp. 147–151 (1988)
9. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Washington (2004)

10. Koenderink, J., van Doorn, A.: Representation of local geometry in the visual system. *Biological Cybernetics* 55, 367–375 (1987)
11. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Trans. Computers* 42, 300–311 (1993)
12. Liu, C.: Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(5) (May 2004)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* (2004)
14. Lowe, D.G.: Object recognition from local scale-invariant features. In: *Proceedings of International Conference on Computer Vision*, pp. 1150–1157 (1999)
15. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: *Proceedings of Computer Vision and Pattern Recognition* (June 2003)
16. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: *Proceedings of International Conference on Computer Vision*, pp. 525–531 (July 2001)
17. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. *Proceedings of the IEEE* 69(5) (1981)
18. Poriklim, F., Tuzel, O.: Fast construction of covariance matrices for arbitrary size image windows. In: *Proceedings of IEEE International Conference on Image Processing*, pp. 1581–1584 (2006)
19. Schokopf, B., Mika, S., Burges, C.J.C., Knirsch, P., Muller, K.-R., Ratsch, G., Mola, A.J.: Input space versus feature space in kernel based methods. *IEEE Transactions on Neural Networks* 10(5), 1299–1319 (1999)
20. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
21. Tuzel, O., Porikli, F., Meer, P.: Region covariance: a fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
22. Van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns. In: Buxton, B.F., Cipolla, R. (eds.) *ECCV 1996*. LNCS, vol. 1064, pp. 642–651. Springer, Heidelberg (1996)
23. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19(7), 775–779 (1997)
24. Yanwei, P., Yuan, Y., Li, X.: Gabor-based region covariance matrices for face recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 18(7) (2008)