Resampling-down Mesh based Discriminant Filter Synthesis for Face Recognition

Vladimir A. Samokhval

United Institute of Informatics Problems National Academy of Science of Belarus, 6 Surganov st., Minsk, 220012, sam@lsi.bas-net.by, tel.: (+375) 17 284 21 64

Abstract: *Presented paper addresses face recognition algorithm by means of synthetic discriminant filters synthesized in pseudo 3-D mesh domain. The objective of the research is to construct facial descriptor in the form of linear filter, which should produce high and low outputs for intra- and inter-class recognition problem respectively. This filter can be synthesized from 3-D sparse meshes derived from a given set of images of a person. Source gray (color) images are subjected to special preprocessing to generate initial mesh. A set of meshes for given object class is the basis to construct discriminant filter. As ever the filter is created it is then used as facial descriptor, i.e. serves as personal ID for face identification. So far mesh density should be considered as a parameter of the classifier the study examines how the performance of the system depends on mesh sparseness.*

***Keywords*:** Face recognition; Pseudo 3-D mesh; Sparse mesh; Discriminant filter.

I INTRODUCTION

Automatic face image recoion is an active research topic in computer vision. This task has a rather long history [1], and still remains a very challenging problem today. The problem has been intensively researched in recent years. Researches in face recognition have been motivated by both their scientific values and wide potential applications in public security, law enforcement and commerce. Crowd surveillance, electronic line-up, store security and mug shot matching are some of the security applications. Much progress in face recognition has been made in the past few years [2]. However, face recognition remains difficult, unsolved problem in general [3, 4]. The approaches to face recognition have covered sources from 2D intensity or color images up to 4D face data [5]. With that high computational and spatial costs make limitations to use these approaches in real application systems.

Two basic approaches to facial image processing, i.e. appearance- and model-based, are now extensively exploited in research and system development. Because of digital image nature of the objects to be identified we can utilize their structural (geometric) features and their intensity/color/texture ones as well. The former features are as usual the features for model-based image recognition algorithms. The later features are used in appearance-based approach. In both approaches, extrinsic imaging parameters, such as pose, illumination, facial expression and makeup still cause much difficulties red ıcing recognition rate.

The classical methods used for face representation are the Principal Component Analysis (PCA) [6] and Linear Discriminant Analysis (LDA) [7]. The former method project the input image into eigenfaces which decorrelate image data features, whereas the later maps the input image into fisher faces which maximize the class separability. Recently, the Independent Component Analysis (ICA) [8] has been investigated in the context of face representation. However, the disadvantage of the ICA technique is that there is no natural way to identify which and how many of the ICA axes should be used to define the dimensionality reducing transformation.

The Linear Discriminant Analysis is a popular pattern recognition method, and some LDA-based face recognition systems [9, 10] have been developed in the last decade and encouraging results have been achieved. However, this method suffers from a well-known small size problem, that is the sample size is small compared with the dimension of feature vector.

Appearance-based face recognition algorithms utilize the intensity or intensity-derived features of original images. The dimensionality of feature vector used by these methods is often very high while the training sample size is relatively small. Such training data can lead the classifier to be biased and have a large variance, resulting in a poor performance [11]. To improve the performance of the "weak" classifiers, a number of approaches have been presented [12, 13, 14].

In many practical tasks it is required to represent the face images in terms of a small number of parameters of poses rather than their own face images due to its effectiveness for data representation, storage and transmission. However, the relationship between the face appearance and the pose parameters is not matched well in the original high-dimensional data space because variations of face images are characterized by a complicated non-linear manifold. The conventional wisdom in face recognition is to project the face image into a lower dimensional space. The motivation for dimensionality reduction is manifold. Image data are inherently of high dimensionality and designing recognition system in the image space would lead to a computationally complex decision rule. There is also the argument based on the peaking phenomenon, which dictates that the ratio of the training set size and pattern dimensionality should be of an order of a magnitude to prevent over-training. As training sets available for face recognition system design are invariably small, a significant reduction of dimensionality is normally thought. Moreover, the face image data are very highly correlated. The use of classical pattern recognition approaches on such data sets leads to unstable decision rules which generate extremely poorly to unseen patterns.

Reduction of dimensionality can be performed on the base of different approaches. Common used techniques are principal component analysis, multidimensional scaling (MDS). Classical MDS finds an embedding that preserves the interpoint distances, equivalent to PCA when those distances are Euclidean. The major algorithmic features of PCA and MDS are computational efficiency, global optimality, and asymptotic convergence with the flexibility to learn a broad class of manifolds.

Being consistent with appearance-based approach which does not require selection of specific facial features, we build recognition model directly from the image data. In our approach a number of facial images of given person (object class) are considered as a cluster in high-dimensional space and the separation between classes is unknown. The idea is contained in the following: given set of feature vectors (members of some object class) find vector-function which produce equal output for each of these vectors. This set of vectors is called the training data. All the other class exemplars are test set of data. The more training exemplars are available for given object class the better final result of intra-class recognition. As to inter-class separability, high value for it is not always ensured and the problem requires additional investigation. The basis of proposed approach is to model human face as approximate parameterized pseudo 3-D sparse mesh which could fill up training set well enough to achieve reliable identification.

II ALGORITHMIC PROCESSING SCHEME OVERVIEW

The most valuable variations in face images are induced by different lightning, pose and expression. Therefore the proposed algorithm requires a number of training facial images for given person obtained under different conditions. In order to eliminate/minimize noise introduced by different background, illumination and scale, special procedures were carried out at the preprocessing stage. At first, face area is limited by elliptic mask and rescaled to standard size (Fig. 1). Geometrical standardization concerns the between-eye distance and its position in the area of elliptic mask. One can choose different ways both how does it and what facial parameters have to be preserved.

To minimize illumination effects on final classification results all images are transformed into standard dynamic range of intensity. Conditional transform for image point $I(x,y)$ should yield to new intensity value $I'(x,y)$ and may be written in simple form as

$$I'(x,y) = A\ (\ I(x,y) - I_{\min}\ )/(\ I_{\max} - I_{\min}\ ) \qquad (1)$$

with constraint $I_{max} \neq I_{min}$ and amplitude A defines the dynamic range of illumination (intensity) and usually equals to 255 in correspondence with gray scale. Obtained image $I'(x,y)$ is then converted into parameterized pseudo 3-D sparse mesh.

Preprocessed images of given person p form class-specific set of feature vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m\}^p$ which are then considered as functions of a basis on which required filter has to be linear decomposed on. Formalization comes to matrix equation for linear filter, specifying conditions needed the task to be resolved.

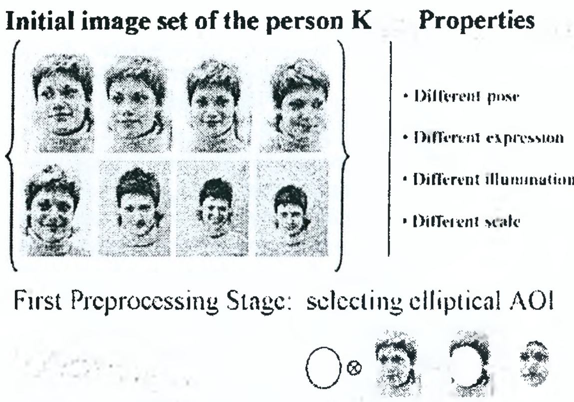

Fig.1 - Image set, its properties and first preprocessing stage.

Here should be pointed out that sparseness property of the proposed image representation by these meshes is of great importance for synthetic discriminant function algorithm. It provides the ability to vary the dimensionality of feature space used allowing both to reduce and to extend it if necessary.

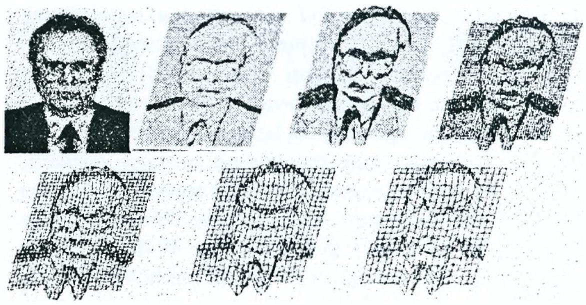

Fig. 2 - Source gray image, its initial and regular resampling down meshes (from top left to bottom right).

The question on how condensed or sparse the generated mesh should be is one of the tasks of this research. Thus one can estimate the proposed technique comparing the results obtained.

III FILTER SYNTHETISIS

Approaches with using synthetic discriminant functions for pattern recognition are good practice and suitable technique. There are many types of pattern recognition problems and there exists many ways to construct synthetic discriminant functions. Comprehensive methodology of the subject for many different cases of classification task is given by David Casasent [15]. We will concentrate our attention on such types of SDF, which produce equal output for class members. In [15] these filters are called ECP SDF or Equal Correlation Peak Synthetic Discriminant Function.

We will now consider general approach of how the synthetic discriminant function (in the form of linear filter) can be constructed.

Consider image set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m\}$ of some object class, where all $\mathbf{x}_i |_{i=1,\ldots,m}$ are $n\times1$ column vectors and n thus represents the dimensionality of the problem. To synthesize filter **f** we form linear equation

$$W\mathbf{f} = \mathbf{u}, \qquad (2)$$

where $m\times n$ matrix $W = \{\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots \mathbf{x}_m^T\}$ is constructed from feature vectors of the training set of a given object class, filter **f** is $n\times1$ vector and **u** is $m\times1$ vector of desired outputs u_i. The main idea is that of the filter **f** produces equal values for all samples from the training set, because they are members of the same object class. Thus, $u_i = u_j = u$. It can be shown, that if the filter **f** is linear combination of m training vectors, the decision of (2) is given by pseudo-inverse of W as follows

$$\mathbf{f} = W^T(WW^T)^{-1}\mathbf{u} \qquad (3)$$

Define matrix $R_{im} = WW^T$ as square $m\times m$ matrix and which is correlation matrix of images. Now equation (3) can be rewritten as

$$\mathbf{f} = W^T R_{im}^{-1}\,\mathbf{u}. \qquad (4)$$

In general case m different vectors (regardless of they are samples of the same object class) may be expected linear independent, and if $m \leq n$ matrix R_{im} is of full rank, the inversion R_{im}^{-1} exists and delivers the single decision of (4).

IV RESAMPLING DOWN MESHES

Mesh representation of initial image is intended to obtain the base for image modeling and recognition. It is aimed mainly to make the transition from intensity to geometry. Mathematically it can be written as

$$G(\mathbf{r}) = \mathbf{F}[B(\mathbf{r})], \qquad (5)$$

where $\mathbf{F}[\cdot]$ denotes operator that transforms image brightness $B(\mathbf{r})$ to shape geometry (i.e., altitude) $G(\mathbf{r})$. An extrinsic form of $\mathbf{F}[\cdot]$ can't be found directly. There exist some ways to calculate an approximate $G(\mathbf{r})$ and one can find it elsewhere or use tailor made techniques if appropriate.

Our interest in using meshes concerns mainly the ability to reduce slightly the dimensionality of the problem while preserving classification accuracy. Such technique enables to resample down mesh sparseness both in regular (like 1:2, 1:3...) as shown in Fig.2 and irregular manner. Direct way to estimate how the system performance depends on mesh sparseness is obvious and results are discussed in the next section.

V EXPERIMENTAL RESULTS

Special database of 21 individuals was prepared from scanned photos with initial resolution 300 dpi and used for computer simulation. For elliptical AOI it lead to image dimensions of approximately 240×320. Total database contains over 300 images with 14 ÷ 16 samples per class.

To evaluate recognition rate of proposed method we construct facial descriptors for all object classes. Synthetic discriminant filters were constructed from initial meshes using five training objects per class and remaining ones played as testing objects. Mean class identification value for testing sets varied between 82 ÷ 93% while delivered 100% rate for training ones. Filter performance is shown in Fig.3 where five randomly chosen objects of ten exemplars are drawn as their projection value to class specific

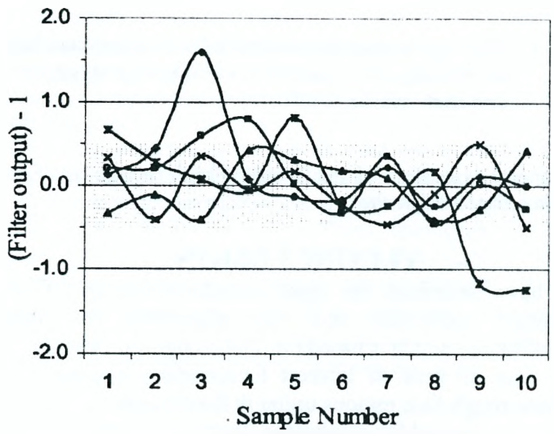

Fig. 3 - Discriminant filter performance for five randomly chosen testing objects.

synthetic filter. This value mathematically is the inner product of the filter and the image under test. The closer the value of inner product to '1' (identified as a 'member') the better classification rate. Decision boundaries for given object class are defined simply as $y = -1/2$ and $y = +1/2$. This definition leads the testing image be accepted as 'member' if filter output falls into interval]-1/2; +1/2[, and as 'nonmember' otherwise. Semitransparent gray strip marks the "true" area with correct identification results.

To evaluate the effect of mesh sparseness on the classification power initial (full) mesh is subjected to resampling-down procedure with ascending ratio. Thus we can evaluate classification performance at the every iteration of the procedure in progress. Experimental results for five chosen object classes are shown in Fig. 4 as classification curves and mean error rate curve with standard deviations. As can be seen classification rate remains practically reliable (i.e. ~ 85%) right up to the resampling ratio 1:4 and then harshly fall down. Such behavior of the error rate curve enables to conclude that efficient resolution in face recognition problem can be found within 300 ÷ 75 dpi. That leads to minimal image (elliptical AOI) size dimensions of approximately 60×80.

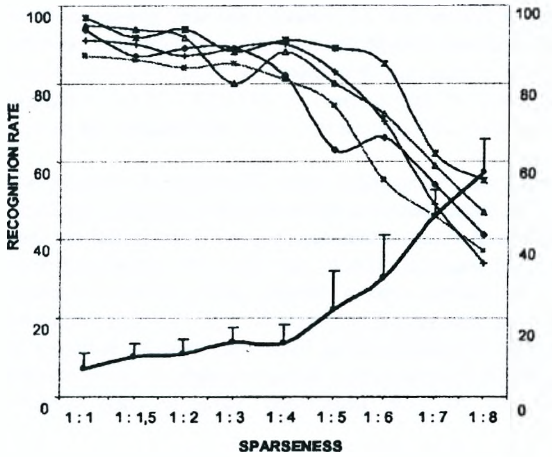

Fig. 4 - Mesh sparseness dependence of face recognition rate for five object classes and error rate with standard deviation marks (bold).

In general experimental results show high recognition rate for developed algorithm.

VI CONCLUSION

We have described the mesh based discriminant filter synthesis technique and the algorithm for face identification which is based on. The recognition algorithm first uses an area of interest localization procedure to provide rough face regions under different poses, followed by a procedure which minimizes illumination effects.

Synthetic discriminant function approach in the stage of classification can potentially deliver the very low False Acceptance Rate (FAR) for intra-class problem if the power of training set is enough. In our research the use of 5-member training set delivers reliable classification performance of the system.

In order to utilize SDF technique the 3-dimensional sparse mesh for face image representation was offered. Some obvious advantages of proposed method for data coding are the following. At first, it enables to vary the dimensionality of the problem and to fit it to obtain non-zero solution of matrix equation (4). In other words, we can guarantee correct solution until the number of training samples is less than the dimensionality of vectors included in matrix $\boldsymbol{W}$. Furthermore, in pure geometric sense the pseudo 3-D face model enables to perform (pseudo) rotations of the face not only in the image plane, but also in depth.

On the whole the question on the mesh sparseness can be defined more exactly if the initial classification scheme will be more steady-state, see Fig.3, and produce as higher recognition results as possible.

REFERENCES

[1] R. Chellappa, C.L. Wilson, S. Sirohey. Human and machine recognition of faces: A survey, *Proc. IEEE*, 5 (83) (1995), pp. 705-740.

[2] W.Y. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips,"Face recognition: A literature survey", *UMD CfAR Technical Report CAR-TR-948*, 2000.

[3] P. .J. Phillips, H. Moon, D.M. Blackburn, E. Tabassi, and J.M. Bone. The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. on PAMI*, 22 (10) (2000), pp.1090-1104.

[4] P. J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone. *FRVT 2002: Evaluation Report*, March 2003.

[5] J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, J.M. Bone, "FRVT 2002: Evaluation Report", March 2003, http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf

[6] M. Turk, A. Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), 1991.

[7] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 19 (7) (1997), 711-720.

[8] C. Liu, H. Wechsler, "Comparative assessment of independent component analysis", *Proc. the 2nd Int. Conf. Audio- and Video-based Biometric Person Authentication, Washington D.C.*, March 22-24, 1999.

[9] D. Dai, P. Yuen, "Regularized discriminant analysis and its application to face recognition", *Pattern Recognition*, 36, (2003), 845-847.

[10] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images", *J. Opt. Soc. Am. A*, 14 (1997), 1724-1733.

[11] M. Skurichina, R. Duin, "Bagging, boosting and the random subspace method for linear classifiers", Pattern Analysis and Applications, 5 (2) (2002), 121-135.

[12] L. Breiman, "Bagging predictors", Machine Learning, 24, 2, (1996), 123-140.

[13] T. Ho. The random subspace method for constructing decision forests, IEEE *Trans. Pattern Analysis and Machine Intelligence*, 8 (20) (1998), pp. 832-844.

[14] Y. Freund, R. Schapire. Experiments with a new boosting algorithm. *Proc. International Conference on Machine Learning*, pp. 148-156, 1996.

[15] D. Casasent. Unified synthetic discriminant function computational formulation, Appl. Opt., 23 (10) (1984), pp. 1620-1627.