

THREE DIMENSIONAL FACE RECOGNITION USING WAVELET DECOMPOSITION OF RANGE IMAGES

Sina Jahanbin*, Hyohoon Choi*, Alan C. Bovik*, Kenneth R. Castleman†

* Laboratory for Image and Video Engineering, The University of Texas, Austin, Texas

† Advanced Digital Imaging Research, LLC., 2450 S. Shore Blvd., Suite 305, League City, Texas

ABSTRACT

Interest in face recognition systems has increased significantly due to the emergence of significant commercial opportunities in surveillance and security applications. In this paper we propose a novel technique to extract features from 3D face representations. In this technique, first the nose tip is automatically located on the range image, then the range data from a hexagonal region of interest around this landmark is decomposed using Barycentric wavelet kernels. The dimensionality of the extracted coefficients at each resolution level is reduced using principal component analysis (PCA). These new features are tested on 206 range images, and a high classification accuracy is achieved using a small number of features. The obtained accuracy is competitive to that of other techniques in literature.

Index Terms— 3D face recognition, Triangular B-spline wavelets, Biometry

1. INTRODUCTION

Although face recognition research first surfaced in the mid-1950s in the psychology and neuroscience communities, it was not till the 1970s that engineers started to investigate the feasibility of automatic face recognition systems. Interest in using biometrics like face images, fingerprints, and iris scans for personal identification has increased since the 1990s due to emergence of commercial opportunities; advances in biometric acquisition hardware; and most importantly increase in surveillance and security applications. Even though biometrics such as fingerprints and iris scans currently are more reliable for personal identification, the human face remains an attractive biometric because it is easily accessible, user friendly and requires minimum cooperation.

Almost all face recognition research in the early days was focused on using portrait images of the face. These portrait images (which we refer to as “2D images”) capture texture and facial color information but do not provide much information about the three dimensional (3D) structure of the face. The efforts of researchers over the past 30 years have resulted in many sophisticated and mature 2D face recognition algorithms. Unfortunately, the performance of 2D face recogni-

tion algorithms remains unsatisfactory due to pose and illumination limitations inherent in 2D images [1]. These limitations were made public by Face Recognition Vendor Test (FRVT) 2002 [2].

Recent advances in three dimensional sensor technology made quick and high quality acquisition of 3D facial surfaces possible. A “3D image” is a three dimensional representation of the face and can be rendered by a 3D mesh of points or a range image. In a “range image” pixel values reflect the distance from the sensor to the imaged surface. Face recognition using 3D representations of the face have become popular among researchers because of its potential to overcome the pose and illumination limitations of 2D images.

In recent years multiresolution surface representation has gained a lot of attention for geometric modeling, computer graphics, and surface compression; however, this important trend that is based on wavelet functions, has never been used in 3D object identification or face recognition applications.

In this paper, we present new features that are based on 3D wavelet coefficients computed on range images. The remainder of this paper is organized as follows: In Section 2 we present an overview of multiresolution surface representation using triangular B-spline wavelets. A detailed description of our feature extraction method is given in Section 3. In Section 4, face recognition accuracies are shown, followed by discussion and conclusion in Sections 5 and 6.

2. BACKGROUND CONCEPTS

2.1. B-splines

B-splines are useful mathematical tools for curve modeling and editing applications. Forsey *et al.* [3] extended this concept to surface modeling by constructing hierarchies of B-Spline bases and using their tensor products to approximate surfaces at different level of details. Alternatively, researchers like Staadt *et al.* [4] have constructed B-spline wavelets from tensor products of one dimensional B-spline wavelets and used them for multiresolution surface representation and compression. Unfortunately, tensor product B-splines (bases or wavelets) suffer limitations in seamlessly representing complicated surfaces [5]. Hence, researchers

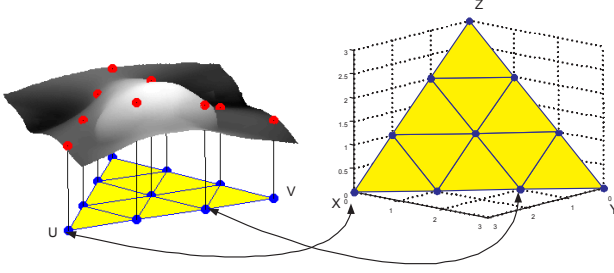


Fig. 1. 2D Barycentric plane shown in 3D cartesian space.

have focused on construction of B-splines (wavelets or bases) on triangular domains.

In [6], de Boor constructed bivariate B-spline bases on regular triangular domains, by projecting a trivariate tensor product of one dimensional B-spline bases on the Barycentric plane parameterized by Barycentric coordinates (u, v, w) , where $u+v+w = 1$. A Barycentric plane is a 2D subspace of 3D Cartesian space, spanning a diagonal 2D plane P_b shown on Fig. 1.

Inspired by this idea, Dreger *et al.* [7] constructed triangular B-spline scaling functions and wavelets based on a 3D tensor product of 1D B-spline wavelets (Chui [8]). Unlike typical 2D wavelet constructions that have one scaling function and three wavelets, one scaling function and 7 independent triangular wavelets from a projection of 3D tensor product wavelets are created. This higher number of wavelets than usual along with the 4:1 subsampling scheme employed in the implementation (section 2.3) results in an over-representation of surface in the wavelet domain.

In order to project a tensor product basis onto the Barycentric domain, that basis should be integrated along a line t which is perpendicular to the Barycentric domain [7]. This line integration will increase the degree and continuity of the resulting wavelets. For example, in our implementation, we have used 3D tensor products of Haar scaling function (φ) and Haar wavelet (ψ) to generate a set of eight linear triangular spline bases. These linear triangular bases include one scaling function ($\tilde{a} = \varphi\varphi\varphi$) and seven wavelets ($\tilde{b}_1 = \varphi\varphi\psi$, $\tilde{b}_2 = \varphi\psi\varphi$, ..., $\tilde{b}_7 = \psi\psi\psi$). Here the notation $\tilde{a} = \varphi\varphi\varphi$ implies that the linear triangular B-spline scaling function resulted from the tensor product of three one dimensional Haar scaling functions (φ) along each axis x, y, z .

Assuming that the input surfaces are defined over a uniform triangular grid, Multiresolution Analysis (MRA) can be implemented in two alternative fashions. The first scheme is to map functional values defined over the vertices of a regular triangular mesh onto an equally spaced 3D grid (Fig. 1) and perform 3D-MRA by sequentially convolving and sub-sampling along the axes x, y, z . Alternatively, multiresolution analysis can take place directly on the triangular mesh using hexagonal Barycentric filters (Barycentric-MRA). Barycentric filter kernels and convolution procedure are explained in Sections 2.2 and 2.3.

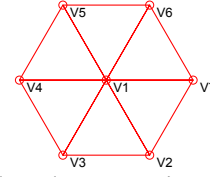


Fig. 2. A hexagonal mesh representing vertices of Barycentric filter kernels.

2.2. Barycentric Filter Kernels

The complexity of discrete convolution algorithms increases with the dimension of the data. In order to avoid 3D multiresolution analysis, Dreger *et al.* [7] developed hexagonal Barycentric filters from a two-scale relationship between 3D tensor product wavelets. Fig. 2 illustrates the structure of a hexagonal filter and the relative vertex locations.

Table 1 presents the coefficients of linear hexagonal Barycentric filters resulting from the projection of 3D tensor products of Haar scaling and wavelet functions. We have used these filters to directly perform wavelet transformation on the triangular mesh.

	V1	V2	V3	V4	V5	V6	V7
\tilde{a}	1/4	1/8	1/8	1/8	1/8	1/8	1/8
\tilde{b}_1	0	1/8	1/8	1/8	-1/8	-1/8	-1/8
\tilde{b}_2	0	-1/8	-1/8	1/8	1/8	1/8	-1/8
\tilde{b}_3	0	1/8	-1/8	-1/8	-1/8	1/8	1/8
\tilde{b}_4	1/4	-1/8	-1/8	1/8	-1/8	-1/8	1/8
\tilde{b}_5	1/4	1/8	-1/8	-1/8	1/8	-1/8	-1/8
\tilde{b}_6	1/4	-1/8	1/8	-1/8	-1/8	1/8	-1/8
\tilde{b}_7	0	1/8	-1/8	1/8	-1/8	1/8	-1/8

Table 1. Linear Barycentric filter coefficients.

2.3. Barycentric Convolution and Sub-sampling

Using hexagonal filters, a wavelet decomposition of surfaces defined over regular triangular domains can be performed on a triangular domain. Similar to conventional schemes, triangular wavelet decompositions consist of convolution and sub-sampling steps. Assuming that \tilde{c}_m represents a set of functional values at the vertices of a triangular mesh, the scaling coefficients \tilde{c}_{m+1} and the wavelet coefficients ($\tilde{d}_{m+1}^1, \tilde{d}_{m+1}^2, \dots, \tilde{d}_{m+1}^7$) for a subset of vertices present at the next coarser level mesh are computed as follows:

- Place the Barycentric filters over each vertex to calculate eight coefficients $\tilde{c}_{m+1}, \tilde{d}_{m+1}^1, \tilde{d}_{m+1}^2, \dots, \tilde{d}_{m+1}^7$ at that vertex by multiplying the corresponding filter kernel coefficients (Table 1) and scaling function coefficients (\tilde{c}_m s) of the current vertex and its neighboring vertices followed by summation.

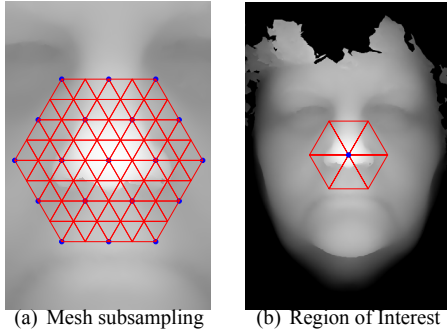


Fig. 3. Triangular Mesh a) Example of mesh subsampling: Vertices marked by blue circles are those that remain after one subsampling step. b) The displayed hexagonal region of interest represents the coarsest level of decomposition.

- Subsample the triangular mesh and save the computed coefficients $\tilde{c}_{m+1}, \tilde{d}_{m+1}^1, \tilde{d}_{m+1}^2, \dots, \tilde{d}_{m+1}^7$ only at those vertices that remain after subsampling.

Fig. 3(a) depicts an example of triangular mesh subsampling. Vertices marked by blue circles are those that remain after one subsampling step.

3. METHODS

3.1. Region of Interest Selection

In order to extract Barycentric wavelet features from range images, first a region of interest (ROI) should be selected. ROI selection is done by placing a central vertex (marked with a blue circle) of a hexagonal mesh depicted in Fig. 3(b) on the nose tip.

The nose tip is defined to be the closest point to the 3D sensor and consequently has the highest range value. In practice, the highest value of the range images is not just limited to one point. Instead it comprises a region of pixels. In this case the centroid of that region is considered as the desired fiducial point.

The mesh shown on Fig. 3(b) represents the vertices used for the coarsest level wavelet transformation. It consist of 7 vertices and 6 equilateral triangles with sides equal to 96 pixels. The subsequent finer level meshes are created by sequentially refining this hexagonal mesh. During mesh refinement, each triangle of the mesh will be divided into 4 new triangles by simply connecting midpoints of each triangle’s edge as depicted in Fig. 3(a). The finest triangular mesh used in this experiment is created by refining the original mesh six successive times. This mesh includes 3169 vertices and 6144 triangles.

3.2. Feature Extraction

Starting from the finest level, the face region surrounded by the region of interest is decomposed using Barycentric

wavelet decomposition technique (section 2.3). It should be noted that the transformation coefficients were calculated only at regular vertices (vertices with 6 neighbors). Table 2 shows the total coefficients extracted at each refinement level. For example at level 2 there are 169 regular vertices available and at each vertex 8 coefficients have been extracted.

	Level 1	Level 2	Level 3	Level 4	Level 5
Coefficients	$8 * 721$	$8 * 169$	$8 * 37$	$8 * 7$	$8 * 1$

Table 2. Number of coefficients calculated at each decomposition level.

4. RESULTS

We tested our algorithms on 206 range images corresponding to 103 individuals from Advanced Digital Imaging Research’s database. Each subject has two range images of size 501×751 with 256 gray level scales. Available range images are partitioned into two mutually exclusive sets, “gallery” and “probe”. Barycentric features were extracted from every range image in these data sets by using the procedure explained in Sections 3.1 and 3.2.

Eventually all the coefficients extracted from each refinement level were separately concatenated to create 5 augmented Barycentric feature vectors. As made clear by Table 2, each augmented feature vector represent a range image in a very high dimensional feature space. In order to decrease the dimensionality of these feature spaces, we used the Principal Component Analysis (PCA).

The basic idea is to create five new spaces and represent features extracted at each resolution level with lower dimensional feature vectors. The final decision on the best feature set is then made by comparing the performance accuracy.

The classification task is done using a nearest neighbor classifier (NN) with Euclidean distance (L_2 Minkowski metric) serving as the desired distance metric. Each curve in Fig. 4 describes the variation of classification performance for features extracted at various decomposition levels under a different PCA feature dimensionality. It is evident that by increasing the feature dimensionality, the classification performance initially improves and then saturates. Using only 5 PCA features from the first decomposition level, 81.5% accuracy is achieved. By increasing the number of PCA features from the first level, correct classification rate reaches 92.2% which is more than any combination of PCA features from higher decomposition levels. Finally the recognition rate of the first level decomposition features saturates at 94.17% when the PCA dimensionality is 30.

5. DISCUSSION

Our system’s performance is promising and competitive to the existing 3D face recognition algorithms. For example, Lee *et al.* [9] implemented a 3D face recognition system by extracting face contours at different depth values with respect to the

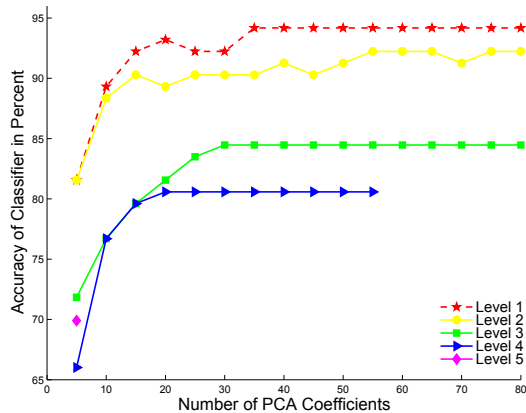


Fig. 4. Recognition performance under different dimensionality of features presented for different decomposition levels.

nose tip. They have evaluated their system using a database of 70 range images from 35 individuals and reported 94.3% correct recognition at rank-five (target individual is among top five matches). It should be noted that our system has been evaluated on a much larger database and also our recognition rates are of first rank rather than rank five recognition. The recognition rate changes dramatically between rank one and rank five.

Tsalakanidou *et al.* [10] performed PCA and extract eigenfaces from corresponding range and portrait images. A database consisting of 80 images from 40 individuals were used for evaluation. Recognition rates as high as 93% for 3D data alone, and 99% for combination of 2D and 3D data were achieved. Considering the performance of 3D features, our system performs better. It is expected that by incorporating other features, even higher performances will be achieved.

Finally, these novel features were extracted using a low computational scheme and performed satisfactory in matching 3D faces by comparing only 35 features. Thus, they are desirable for a real time face recognition system. These features are expression invariant because they are extracted from a hexagonal area around the nose tip which is rigid and does not change much by changes in facial expression. Likewise, more Barycentric features can be extracted from other fiducial points such as the eye corners. Combining features from other parts of the face with those computed from nose tip will further improve the accuracy.

6. CONCLUSION

In this paper we have introduced a set of novel features for 3D face recognition that combine multiresolution surface representation with PCA. In this method, the range data from a hexagonal region of interest is decomposed using Barycentric wavelet kernels. The dimensionality of the extracted coefficients at each resolution level is reduced using principal component analysis (PCA). Experimental results on a database of 206 range images showed that by using only 35 PCA coeffi-

cients in a simple nearest neighbor classifier, recognition rates as high as 94.17% are achievable. In the future, it is desirable to research the performance of higher order Barycentric filters, rather than linear ones. Improving the classification and discrimination method is another research topic for future work.

7. REFERENCES

- [1] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.
- [2] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone, "Face recognition vendor test 2002: overview and summary," NISTIR 6965, National Institute of Standards and Technology, Tech. Rep., 2003. [Online]. Available: www.frvt.org
- [3] D. Forsey and R. Bartels, "Hierarchical b-spline refinement," in *Computer Graphics*, vol. 22, no. 4. Comput. Graphics Lab., Waterloo Univ., Ont., Canada: ACM, 1988, pp. 205–212.
- [4] O. Staadt, M. Gross, R. Weber, R. Yagel, and H. Hagen, "Multiresolution compression and reconstruction," in *Proceedings. Visualization '97 (Cat. No.97CB36155)*, R. Yagel and H. Hagen, Eds. IEEE Comput. Soc. Tech. Committee on Comput. Graphics, 1997., pp. 337–346.
- [5] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, *Wavelets for Computer Graphics*. Morgan Kaufmann Publishers, Inc., 1996.
- [6] C. de Boor, "On the evaluation of box splines," *Numerical Algorithms*, vol. 5, no. 1–4, pp. 5–23, 1993.
- [7] A. Dreger, M. Gross, and J. Schlegel, "Construction of multiresolution triangular b-spline surfaces using hexagonal filters," in *Visual Computer*, vol. 16, no. 6. CGS, 2000, pp. 339–356.
- [8] C. K. Chui and E. Quak, "Wavelets on a bounded interval," vol. 105, pp. 53–75, 1992.
- [9] Y. Lee and T. Yi, "3d face recognition using multiple features for local depth information," in *Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference focused on*, vol. 1, 2-5 July 2003, pp. 429–434vol.1.
- [10] F. Tsalakanidou, D. Tzovaras, and M. G. Strintzisa, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, pp. 1427–1435, June 2003.