Recognising Facial Expressions Using Spherical Harmonics

James Sharpe and Edwin R. Hancock

Department of Computer Science, University of York

Abstract. This paper explores whether facial expressions can be recognised by using the distribution of surface normal directions in the extended Gaussian image (EGI). We work with range images and extract surface normals using a mesh fitting technique. Our representation of the surface normals is based on the co-efficients of spherical harmonics extracted from the EGI. We explore whether the co-efficients can be used to construct shape-spaces that capture variations in facial expression using a number of manifold learning techniques. Based on a comparison of various alternatives, the best results are given by the diffusion map.

1 Introduction

Recently, it has been shown that statistical models based on the distribution of surface normals can offer a powerful means of representing an recognising facial shape. The reasons for this are two-fold. First, the needle map (or Gauss map) offers a representation that is rich in terms of differential geometry, and hence can potentially be used to capture subtle variations in facial shape. Secondly, surface reflectance is determined by the relative orientation of the surface and the light source. Hence, given a description of skin reflectance then a surface normal model can be fitted to image brightness data to recover 3D shape from a 2D facial image.

There are a number of approaches to capturing the statistics of surface normal direction. For instance, Smith and Hancock [1] project the surface normals into a tangent space to construct a statistical model using principal geodesic analysis. This work has recently been extended to gender recognition [2], but has proved too cumbersome for expression recognition. Bronstein, Bronstein and Kimmel [3] develop a spherical embedding, that allows faces to represented in a manner that is invariant to expression. Parameterising the distribution of surface normals, Kazhdan et al. [4] use the fact that the spherical harmonics of a given frequency form a subspace which is a rotationally invariant and which can be applied to the extended gaussian image (EGI). to create a rotationally invariant shape descriptor.

The aim in this paper is to take this work one step further and to explore whether spherical harmonics can be used for expression recognition. Specifically, we aim to explore whether the co-efficients can be used to construct shape spaces that allow different individuals and their expressions to be distinguished. To this

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 157–166, 2008.

end, we explore different ways of embedding the spherical harmonic co-efficients into a low dimensional pattern space. The most effective of these is the diffusion map [5], but we also provide comparison with kernel principal components analysis [6], local linear embedding [7], heat kernel embeddings and [8], commute time embeddings [9,5].

2 Representing Variations in Surface Normal Direction

In this section we describe our method. We commence from the extended Gaussian image which we parameterise using spherical harmonics. Using the co-efficients of the spherical harmonics, we construct shape-spaces using the diffusion map.

Extended Gaussian Image: The Extended Gaussian Image(EGI) [10] is a shape descriptor that represents the distribution of surface normals as data on a Gaussian Sphere. The Gaussian image is the mapping of surface normal data onto a unit sphere. The EGI is then an extension of this idea where an additional weight is assigned to each point on the sphere equal to the area of the surface that has the given normal. Computationally the EGI is formed by constructing a histogram from the surface normal data. This is achieved by representing the normal data as (θ, ϕ) pairs and then constructing a histogram.

Spherical Harmonics: Spherical Harmonics [11] arise as part of the general solution for the heat equation in spherical coordinates and are expressed in terms of Legendre polynomials. The Legendre polynomials are a set of special functions that are obtained when solving the heat equation. Specifically, they emerge as a result of solving associated Legendre equation:

$$(1-s^2)\frac{d^2\Theta}{ds^2} - 2s\frac{d\Theta}{ds} + \left(\mu - \frac{m^2}{1-s^2}\right)\Theta = 0 \tag{1}$$

Since this is a second-order differential equation, the power series is of the form $\Theta(s) = \sum_{n=0}^{\infty} a_n s^n$. where $s = cos\theta$. The solution $\Theta(s)$ is a polynomial of degree l, referred to as the Legendre Polynomial of degree k, denoted by $P_l(s)$. The Legendre Polynomials solve the Legendre equation (Eqn 1) for the case $m = 0$. To solve for $m \neq 0$ we can use the associated Legendre functions:

$$P_l^m(\cos\theta) = \sin^m\theta P_l^(\cos\theta) \qquad m = 1, 2, \ldots, l \tag{2}$$

With these ingredients the spherical harmonics are defined as

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\exp[im\varphi]$$

A Basis for the 2-Sphere: A complex valued function $f(\theta, \varphi)$ can be decomposed as a set of coefficients $a_{l.m}$ using the series expansion

$$f(\theta,\varphi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{l.m}Y_m^l(\theta,\varphi)$$

This is an infinite sum over the set of all spherical harmonics. A band limited function is a function that can be approximated by the finite set of terms

$$f(\theta,\varphi)=\sum_{l=0}^{b}\sum_{m=-l}^{l}a_{l.m}Y_m^l(\theta,\varphi)$$

Using the band limited representation we can calculate the coefficients a_{lm} by multiplying both sides by the complex conjugate $Y_m^{l*}(\theta,\varphi)$ and integrating over the solid angle Ω to give:

$$a_{lm}=\int_{\Omega}f(\theta,\varphi)(Y_m^l)^*(\theta,\varphi)d\Omega$$

Over the unit 2-sphere this becomes:

$$a_{lm}=\int_0^{\pi}d\varphi\int_0^{2\pi}f(\theta,\varphi)(Y_m^l)^*(\theta,\varphi)\sin(\theta)d\theta$$

Spherical Embedding: Recent work on expression-invariant face recognition using spherical harmonic embeddings [3] has been based on the application of the work of Kazhdan et al. [4]. Kazhdan et al. use the fact that the spherical harmonics of a given frequency l, form a subspace $V_l=\mathrm{Span}(Y_l^{-l},Y_l^{-l+1},\ldots,Y_l^{l-1},Y_l^l)$ which is a representation for the rotation group. It is this property which ensures the description is rotationally invariant. By taking each set of basis functions, summed over a given frequency, the L2-norm for the coefficients is invariant to rotation [4]. This parameterisation can be applied to the extended gaussian image (EGI) to create a rotationally invariant shape descriptor. There is however a degree of information loss which means that in general for two values with the same energy for a given frequency ($l>2$), there are multiple representations that do not have a rotation defined between them. Kazhdan et al use this representation for expression invariant recognition, and so the information that is lost by the transform may be salient. Moreover, of the transform that is used to construct shape-spaces via manifold embedding, then the modes of shape variation may not be reliably captured.

Diffusion Map: Diffusion maps are coordinates constructed from the eigenfunctions of Markov matrices. Coifman and Lafon recognised that the majority of the existing manifold learning techniques are simply special cases of diffusion processes[5]. They provide examples of how diffusions relate to the Laplace-Beltrami operator on manifolds which Levy [12] shows as a good basis for functions of geometry and topology of objects. Coifman et al. also shows how the diffusion map is robust to noise, a useful feature for a pattern recognition algorithm, since noise is an inherent problem.

Suppose that the jth range-image can be paramterised using the vector of spherical harmonic co-efficients A_j. From the T range images available to use we compute the $T\times T$ similarity weight matrix W with elements $W(i,j)=\exp[-\sigma(A_i-A_j)^T(A_i-A_j)]$. The associated Laplacian matrix is $L=D-W$,

where D is the $T \times T$ diagonal degree matrix with elements $D(i,i) = \sum_{j=1}^{T} W(i,j)$. The diffusion map commences from the random walk on a graph which has transition probability matrix $P = D^{-1}W$. Although P is not symmetric, it does have a right eigenvector matrix Φ, which satisfies the equation

$$P\Phi = \Lambda\Phi \tag{3}$$

Since $P = D^{-1}W = D^{-1}(D-L) = I - D^{-1}L$, $L\Phi = (I-\Lambda)D\Phi$. The embedding co-ordinate matrix for the diffusion map is $X = \Lambda^t \Psi^T$, where t is real. For the embedding, the diffusion distance between a pair of nodes is $d_t^2(u,v) = \sum_{i=1}^{m} (\lambda)_i^{2t} \left(\phi_i(u) - \phi_i(v)\right)^2$.

The diffusion map algorithm has two input parameters; σ, which is used as the variance value when constructing the Markov chains from a Gaussian distribution and t which is the number of steps of the diffusion process. The larger the value of σ, the more weight is given to distant points. When t is set to 0 then the diffusion map is equivalent to the Graph Laplacian, when set to $\frac{1}{2}$ it is equivalent to the Fokker-Plank propagator and when set to 1 the diffusion map is equivalent to the Laplace-Beltrami operator.

3 Experiments

Range Data Collection: We have collected expression data using a Cyberware 3030 whole-head scanner. The Cyberware 3030 scanner is mounted on a PS motion platform and rotates around the subject whilst shining a low intensity laser to create the digitized points. The scan process takes about 30 seconds per scan, during which time the subject is required to remain stationary. For some expressions this is difficult to achieve and so multiple attempts were required for these scans. The scans begin at the back of the head so the effect of any slight movements of the subject during the scan are minimized as we are not interested in the area at the rear of the head where the join occurs. Movement of the subject can be seen visually in the output by stretching of features or distortion of the face. Another obstacle to overcome was the area under the chin often had holes in the scan due to the chin occluding the area from the laser during the scan. This was not a problem with all subjects but for those where this occurred the subject was asked to tilt their head back so that the laser was not obstructed by the chin.

To simplify the identification process the collected range data was edited to remove hair, ears and neck. Subjects were chosen without glasses or facial hair. Ten subjects were collected in five different facial poses: anger, neutral, smile, sad and surprise, however one subject's data was not usable due to their hair covering too much of the face and thus not producing usable meshes after the editing process. The nine subjects that were used are shown in their edited forms in figure 1. Each row in the figure corresponds to a facial expression; anger, neutral, smile sad and surprise; and each column, an individual subject.

The EGI can then be used as the input for the spherical decomposition described above. The resulting shape descriptors reconstructed from their spherical

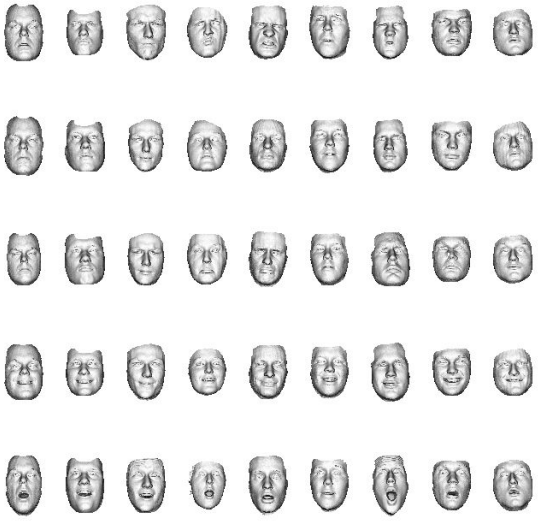
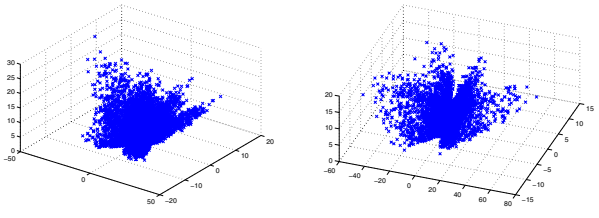


Fig. 1. Collected facial expressions



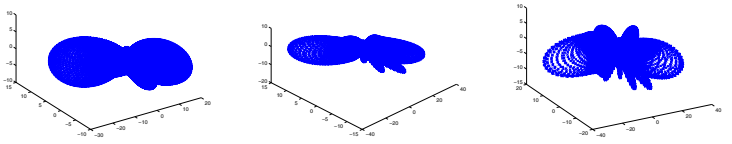
(a) (b)

Fig. 2. Shape Descriptors from histogram data

decompositions can be seen in Figure 3. As the spherical harmonic frequency is increased, higher frequency data can be represented as can be visually observed. It is important to verify that this representation, represents the original data so we will begin by looking at the accuracy of the produced shape descriptor.

Shape Descriptor Accuracy: The average error rates for the reconstruction of the shape descriptor from its spherical decomposition can be seen in Figure 4. The error value is calculated as the squared difference between the reconstructed shape descriptor and the actual shape descriptor, normalised by the number of samples. The shape descriptor was sampled with 10000 bins, with both θ and φ quantised into 100 bins each. The data was averaged over the 45 scans that were collected. The error shows an approximately logarithmic reduction in error as the frequency is increased.

Spherical Harmonic Decomposition: The surface normals for the mesh are converted from Cartesian format to spherical angles in the ranges $0 < \theta < 2\pi$ and



(a) 10 (b) 20 (c) 30

Fig. 3. Spherical Harmonic shape descriptors

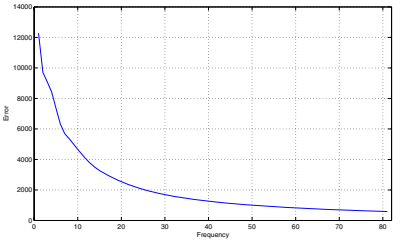


Fig. 4. Error in shape descriptor

$0 < \varphi < \pi$. An Extended Gaussian Image is calculated from these angles which is then used as the descriptor to decompose into spherical harmonics. Calculation of the Spherical Basis functions $Y_m^l : l < n, -l < m < l$ involves evaluating the function at each of the sample points (θ, φ) taken to be the centre point of each of the histogram bins calculated as part of the creation of the shape descriptor. The coefficients calculated from the decomposition are then arranged into a vector to create a vector representation of the spherical harmonic decomposition.

Embeddings: The embeddings were constructed using spherical harmonics decomposed with a bandwidth of 60. Above this the computational time becomes prohibitive within Matlab. The dimensionality of the embeddings can be determined by examining the magnitudes of the eigenvalues produced. By computing the cumulative percentage represented by the eigenvalues, a value for the dimensionality of the data can be determined. In the case of PCA (figure 5) we can see that 90% of the variation is accounted for in the first 5 eigenvalues. However, we observe a poor level of clustering for this embedding.

The degree of clustering can be visualised using the pairwise distance matrix with elements $DD(i,j) = (A_i - A_j)^T(A_i - A_j)$. In the ideal case, there should be a block structure along the diagonal when the vectors of the same class are grouped close to each other. The clustering for the spherical harmonics is shown in figure 6. The vectors are grouped by expression to determine if there is any visible block structure. There is no diagonal block structure present in this matrix which suggests a lack of clustering for the raw spherical harmonics.

We have explored the effect of varying the parameters of the diffusion map on the resulting embedded distance matrix. The value of σ controls effect of distance. Calculating the distance matrix for values of σ from 1 to 15 and varying t between 0 and 1 for each sigma value produces an array of varying distance

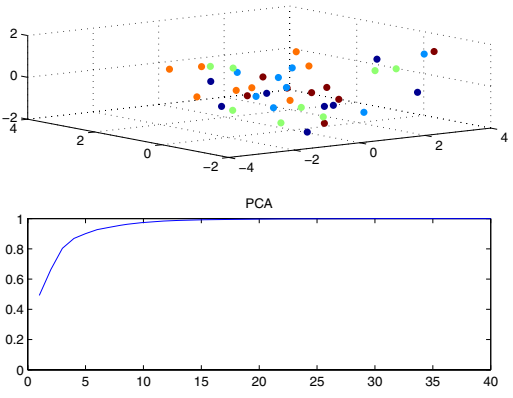


Fig. 5. Data embedded using PCA (top). Cumulative percentage of eigenvalues (bottom).

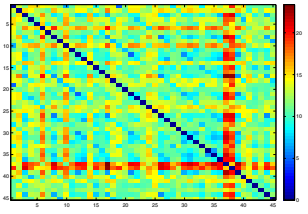


Fig. 6. Pairwise distance matrix for spherical harmonic coefficient vectors, ordered by expression

matrices of differing quality. Most are similar in structure to figure 9(e), although the result from figure 7 with t set to 0.8 and σ to 10 gives a better block diagonal structure.

Figure 8 shows the embedded points after applying the diffusion map, each colour represents a different facial expression. There is some clustering of points, as expected from the distance matrix, but there is more than one cluster for each class.

We have also applied Kernel PCA (using a polynomial kernel), heat kernel embedding, commute time embedding and LLE to the expression data. Figure 9 shows the distance matrices for the embedded points. These all give poorer results than the diffusion map. Although within the blocks the distances increase with each subject, and so would not be useful for recognition, but the diffusion map algorithm has parameters which can be adjusted to produce different results. It is interesting to note that in the KPCA distance matrix (Figure 9(a)) there is a regular pattern of banding of larger distances, these bands all belong to one subject, if the matrix was grouped by subject this would be more apparent. This suggests that this subject has some particular features that are not present in the other subjects. This additional structure may reveal itself with a larger sample size, as more subjects are likely to exhibit the characteristic,

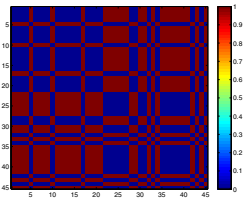


Fig. 7. Diffusion map distance matrix with alpha = 0.8 and sigma = 10

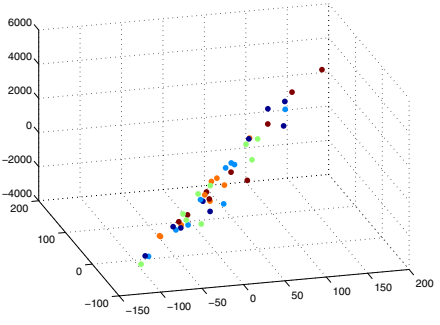
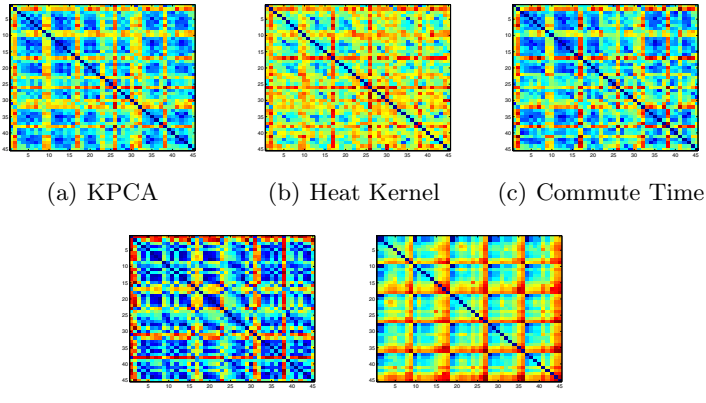


Fig. 8. Embedded data using diffusion map



(a) KPCA (b) Heat Kernel (c) Commute Time

(d) LLE (e) Diffusion Map

Fig. 9. Pairwise distance matrices for various embedding methods

enabling the algorithm to construct a better manifold embedding due to the extra information.

Classifier: A k-nearest neighbour algorithm is used to classify a given sample point. For each point to be tested, the embedding is applied and then the nearest k neighbours are found. A k-nearest neighbour classifier is suitable for this task

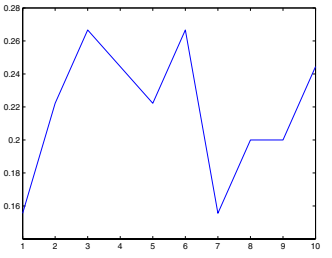


Fig. 10. Classifier accuracy as k varies

because there is a small training set and so storage for the classifier is reasonable. To test the performance of the classifier on the diffusion map embedding, each point was removed from the dataset and then the k-nearest neighbour algorithm applied. The percentage of correct classifications is then recorded. The results can be seen in figure 10 for values of k ranging from 1 to 10. The best performance is for 3 nearest neighbours which by observing the clustering in figure 8 is sensible since the clusters are generally at least 3 elements. The sample size is not large enough to do more sophisticated error analyses of the classifier.

4 Conclusions and Further Work

In this paper we have explored using the set of spherical harmonics coefficients of a model's normal map as a shape descriptor to be used in manifold learning techniques, specifically attempting to identify differences in facial expressions. We have shown that application of the diffusion map to the spherical harmonic decomposition of the surface normals of the model provides some block structure which can be utilised in classifying facial expressions. For a given expression there may be multiple clusters, although determining the differences between these clusters is an open question. A larger database for training the algorithm to create a classifier would enable a more accurate classifier. This paper only considers a static 3D representation as source data. As facial expressions are temporal in nature, more success may be achieved by looking at the time derivatives of the statistics of the normal maps.

References

1. Hancock, E.R., Smith, W.A.P.: Recovering facial shape using a statistical model of surface normal direction. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1914–1930 (2006)
2. Wu, J., Smith, W.A.P., Hancock, E.R.: Weighted principal geodesic analysis for facial gender classification. Progress in Pattern Recognition, Image Analysis and Applications, 331–339 (2007)
3. Bronstein, A., Bronstein, M., Kimmel, R.: Expression invariant face recognition via spherical embedding (2005)

4. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation invariant spherical harmonic representation of 3d shape descriptors (2003)
5. Coifman, R.R., Lafon, S.: Diffusion maps. Applied and Computational Harmonic Analysis, 5–30 (2006)
6. Kim, K.I., Jung, K., Kim, H.J.: Face recognition using kernel principal component analysis. IEEE Signal Processing Letters 9(2) (2002)
7. Rowels, S., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
8. Bai, X., Wilson, R.C., Hancock, E.R.: Manifold embedding of graphs using the heat kernel. Mathematics of Surfaces, 34–49 (2005)
9. Qiu, H., Hancock, E.R.: Graph embedding using commute time. Structural, Syntatic and Statistical Pattern Recognition, 441–449 (2006)
10. Horn, B.K.P.: Extended gaussian images. Proceedings of the IEEE 72(12), 1671–1686 (1984)
11. Hobson, E.W.: The Theory of Spherical and Ellipsoidal Harmonics. Chelsea Publishing Company (1965)
12. Levy, B.: Laplace-beltrami eigenfunctions: Towards an algorithm that understands geometry (2006)