LARGE SCALE LEARNING OF ACTIVE SHAPE MODELS

Atul Kanaujia and Dimitris N. Metaxas

Department of Computer Science,Rutgers University
{*kanaujia,dnm*}*@cs.rutgers.edu, http://www.cs.rutgers.edu/˜kanaujia,dnm*

ABSTRACT

We propose a framework to learn statistical shape models for faces as piecewise linear models. Specifically, our methodology builds upon primitive active shape models(ASM) to handle large scale variation in shapes and appearances of faces. Non-linearities in shape manifold arising due to large head rotation cannot be accurately modeled using ASM. Moreover overly general image descriptor causes the cost function to have multiple local minima which in turn degrades the quality of shape registration. We propose to use multiple overlapping subspaces with more discriminative local image descriptors to capture larger variance occurring in the data set. We also apply techniques to learn distance metric for enhancing similarity of descriptors belonging to the same class of shape subspace. Our generic algorithm can be applied to large scale shape analysis and registration.

Index Terms— Active Shape Models, SIFT, Relevance Component Analysis, Anderson Darling Statistics

1. INTRODUCTION

Recent research in shape analysis and registration have proposed improved methodologies for searching in highly non-linear Riemannian manifold for the globally optimal shape. Whereas [1, 2] have used sampling based techniques to optimize regularized shape matching cost function, [3] have proposed improved continuous shape regularization for more stable and optimal solution. In this work we propose several improvements over the past shape registration techniques. Unlike previous works, shape analysis is not performed in the common tangent space. This removes the restriction that all shapes should be in the vicinity of the mean shape. In addition, we propose a framework to learn non-linear shape manifold as overlapping subspaces. In this respect our work follows from [4, 5]. The number of clusters is learned directly from the data using normality test for clusters [6]. Finally we improve upon the likelihood function by using more discriminative descriptors and a learned distance metric to enhance correlation between features belonging to the same shape cluster [7]. We demonstrate the algorithm on the face alignment problem by accurately localizing faces with shapes that are far from the mean shape in the shape space.

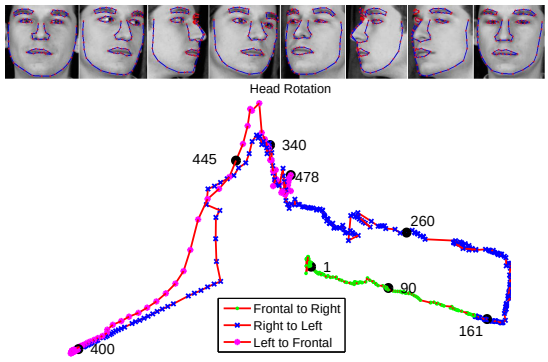

Fig. 1. 3D isomap embeddings for the tracked shapes of a full profile head movement(right turn followed be full left turn and back). The intrinsic non-linearity of the shape manifold makes linear models ineffective in dealing with large variations. The marked frame numbers are shown in the top row.

2. LEARNING NON-LINEAR SHAPE MANIFOLD

The active shape learning has been formulated as the posterior optimization with the global prior shape model and the local image likelihood model. For the shapes $\mathbf{S}$ learned as a set of $\mathbf{N}$ landmark point locations $\mathbf{S} = \{x_1, y_1, \cdots, x_N, y_N\}$, a PCA subspace is learned that captures the relevant variance in shapes (95%) by projecting the data set onto eigenvectors $\mathbf{P}$ with largest eigenvalues

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{P} * \mathbf{b} + \epsilon \tag{1}$$

where $\mathbf{X} = \mathbf{\Phi}(\mathbf{S})$, $\mathbf{\Phi}$ being the linear transformation for global scaling, translation, rotation and linearizing the shape. Planar shape distribution lies on highly non-linear Riemannian manifold. Fig. 1 shows an isomap embedding in 3D of the tracked shapes across a full head rotation [1]. The distance metric on the non-linear manifold is approximated as procrustes distance by projecting the shapes onto the tangent plane of the mean shape. The shape model learned in the tangent space is an accurate representation of the shapes in the vicinity of the mean shape $\bar{\mathbf{X}}$. However for the shapes far away from the mean shape, the large scaling (fig 2*(left)*) of the shape vectors causes the learned PCA subspace to distort and generate unrealistic shapes. Kernel methods [8] tar-

[1]video http://www.cs.rutgers.edu/∼kanaujia/Data/Video.zip

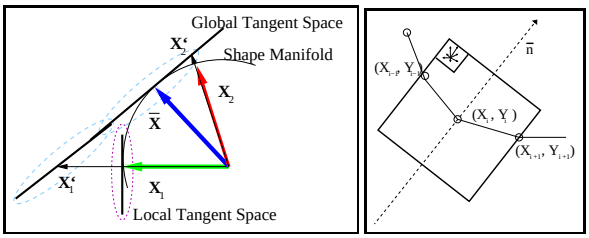

Fig. 2. *(Left)* The shape $\mathbf{X_1}$ is projected to tangent space by rescaling as $\mathbf{X_1}/(\mathbf{X_1} \bullet \bar{\mathbf{X}})$. For shapes $(\mathbf{X_1})$ farther of $\bar{\mathbf{X}}$, the rescaling causes unrealistically large variance in the tangent space thereby distorting the PCA subspace. *Right* The SIFT descriptor is computed over a patch along the normal vector at the landmark.

get this problem by projecting the shapes into features space where linear methods can be applied. These methods suffer from two principal drawbacks that prevent their applications to large scale shape analysis. Firstly kernel methods are inclined to overfitting due to more parameters and hence are not robust to outliers. Secondly kernel methods require pre-image mapping for projecting the shapes back from the feature space to the image space. This introduces additional inaccuracies in the shape model.

2.1. Preserving Non-linearity in Shapes

To address this problem for large set of shapes, we propose to learn the non-linear shape manifold as multiple overlapping linear subspaces. The original shapes are first projected to a global tangent space (fig. 2) so that the euclidean distance can be used for clustering. The shapes are aligned to a reference shape iteratively by computing $\mathbf{\Phi_i(S)} = \gamma\mathbf{RS} + \mathbf{T}$ where the γ is the scaling factor, $\mathbf{R}$ and $\mathbf{T}$ as the rotation and the translation matrices respectively. The aligned shapes are clustered using Gaussian Mixture Model. Based on the class resposibilites in the tangent space, the original shapes are grouped into multiple clusters with subspaces learned within each cluster independently.

In order to ensure smooth manifold during shape search, adjacent subspaces should overlap sufficiently. The amount of overlap can be controlled by variance flooring during the EM algorithm for clustering the data set. In addition we artificially add 15% of cluster points from the neighboring clusters. This overlapping in the global tangent space ensures that shapes in the original image space generate overlapping linear subspaces.

2.2. Learning Number of Clusters

Key to good cluster model holds in choosing the number of clusters (to avoid oversized or undersized clusters) which is a hard algorithmic problem and is usually done by cross-validation. However for shape analysis, it is also difficult to determine

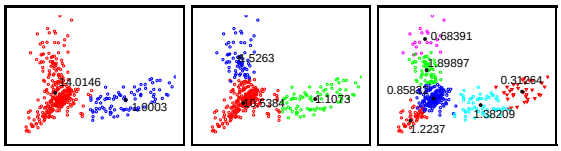

Fig. 3. Anderson Darling($\mathbf{AD^2}$) statistics for the clusters for the iterations 1, 2 and 3. The cluster centers are split if the $\mathbf{AD^2}$ statistics is greater than the critical threshold (based on the desired significance level). The final result is as shown in fig. 5. Only the first two principal components are shown

the optimal number of clusters due to absence of any reliable evaluation technique. Moreover, learning a linear PCA model in the cluster entails shapes to have gaussian distribution. We determine optimal number of clusters based on normality statistics of the cluster distribution. A number of goodness-of-fit tests exist for gaussian distribution e.g. Komolgorov-Smirnov, Anderson-Darling, Shipiro-Wilk and Von Mises. The most popular normality test is Anderson-Darling $(\mathbf{AD^2})$ that determines when it is unlikely that the current data is not generated from the gaussian distribution. The $\mathbf{AD^2}$ test is a 1D normality test and uses empirical cumulative distribution function(CDF) to compute the statistics $\mathbf{AD^2}$ of N shapes projected onto 1D vector $\mathbf{V}$ that preserves the intrinsic structure of the data. The vector $\mathbf{V}$ is obtained as the line joining 2 centers obtained by running 2 cluster k-means on the data. The projected shapes are ordered according to the scalars $\langle \mathbf{X_i}, \mathbf{V} \rangle / \|\mathbf{V}\| = w_i$.

$$\mathbf{AD^2}(\psi) = -N - \left(\frac{1}{N}\right) \sum_{i=1}^{N} (2i-1) * \{ln(\psi(\frac{(w_i - \mu)}{\sigma})) + ln(1 - \psi(\frac{(w_{N-i+1} - \mu)}{\sigma}))\} \quad (2)$$

where $\psi(x) = \frac{1}{2}\{1 + erf(\frac{x-\mu}{\sigma})\}$ is the normal CDF. This test need to be modified for small samples as $\mathbf{AD^2_m} = \mathbf{AD^2}\{1 + \frac{0.75}{N} + \frac{2.25}{N}\}$ The test compares the $\mathbf{AD^2_m}$ statistics against the standard critical values $\mathbf{AD^2_{crit}} = \{0.631, 0.752, 0.873, 1.035\}$ depending upon the corresponding desired significance level $\alpha = \{0.1, 0.05, 0.025, 0.01\}$ and rejects the hypothesis that the distribution is normal if the value exceeds the critical value. In our experiments we used $\alpha = 0.01$ with critical value 1.035. The algorithm starts from a single center and iteratively splits the centers until it becomes unlikely that the current distribution is not a gaussian. Fig. 3 shows the iterative splitting along with the $\mathbf{AD^2_m}$ test statistics for each cluster(computed by splitting the cluster). The split centers of the clusters with $\mathbf{AD^2_m}$ statistics larger than the critical values are retained in the next iteration. The final clusters shown in fig. 5 are obtained by further splitting clusters (fig. 3*(right)*) that are not gaussian.

2.3. Image Likelihood

The image likelihood $P(I|X, \mathcal{M})$ is modeled as the probability of local descriptors at the landmark points conditioned

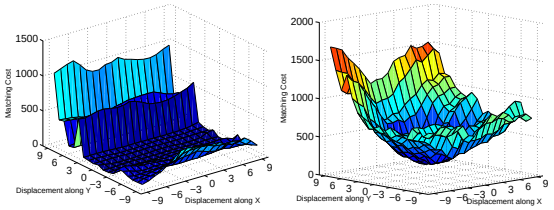

Fig. 4. *(Left)* Gradient profile matching cost of a landmark point over a window of size 19x19. Notice the multiple minima resulting in poor alignment of shapes. *(Right)* SIFT descriptor matching cost for the same landmark point

on the learned models $\mathcal{M}$ (the shape and the local descriptor models). With the prior shape model $P(X|\mathcal{M})$ the posterior can be expressed as

$$P(X|I,\mathcal{M}) \propto P(I|X,\mathcal{M}) * P(X|\mathcal{M}) \tag{3}$$

The posterior maximization however suffers from the difficulty of getting stuck at local minima. A number of works exist that try to alleviate this problem by either sampling from the prior and evaluating the likelihood[1, 2] or improving the shape regularization methods[3]. We adopt simpler approach to improve the likelihood model by using more discriminative SIFT(scale invariant feature transform) descriptors that are distinctive enough to differentiate between landmarks, in order to avoid multiple minima and yet invariant to within-class variations. SIFT descriptors encodes the internal gradient information of a patch around the landmarks thus capturing essential spatial position and edge orientation information of the landmark. Quantizing gradient orientations into discrete values in small spatial cells and normalizing these distributions over local blocks makes the descriptor invariant to affine changes in illumination and contrast. The descriptors are computed over a grid of cells and the vector of the histogram values is normalized by L2-norm computed over the entire block. In order to make the descriptor rotation invariant, the gradient orientations are always computed relative to the normal vector (fig. 2 *(right)*) at the landmark point. Unlike the gradient profile cost function, that has multiple minima (fig. 4), the SIFT matching cost function is remarkably unimodal.

2.4. Distance Metric Learning

The local descriptors tend to be highly correlated for a particular viewpoint e.g. SIFT descriptors for landmarks on the outer contour for a frontal face differ markedly from the profile face. Local image descriptors of the shapes belonging to the same cluster cannot be used for learning the likelihood model as small clusters may be too restrictive and hinder the shape search. On the other hand using all the data points may generate extraneous variance in the training also causing inaccuracies. Instead, we apply an alternate strategy

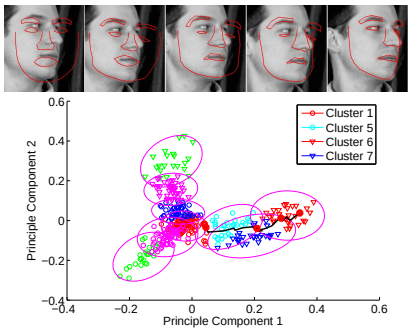

Fig. 5. Trajectory showing search for the optimal shape across clusters. The red circles denote the frames in the top row

to selectively downweight values of spatial cells by learning full rank Mahalanobis metric in the descriptor space using Relevance Component Analysis(RCA). RCA downscales the global variance and enhances similarity between descriptors belonging to the same shape cluster, by giving larger weights to the relevant cells of the SIFT blocks. It finds independent linear mappings $f : \Omega_i(\mathbf{S}) \rightarrow [\mathbf{Y_i} = \mathbf{A_i} * \Omega_i(\mathbf{S})]$ that maximizes the mutual information $I(\Omega_i(\mathbf{S}), \mathbf{Y_i})$ between $\mathbf{Y_i}$ and $\Omega_i(\mathbf{S})$(the SIFT descriptors for the i^{th} landmark of shape $\mathbf{S}$). For M shape clusters, we compute the reweighings $\mathbf{A_i}$ that minimizes within class variance

$$\min_{\mathbf{A_i^T A_i}} \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{N_j} \|\Omega_i(\mathbf{S_{jk}}) - \overline{\Omega_{i,j}(\mathbf{S})}\|_{\mathbf{A_i^T A_i}} \text{ s.t. } |\mathbf{A_i^T A_i}| \geq 1 \tag{4}$$

$\overline{\Omega_{i,j}(\mathbf{S})}$ is the mean of descriptors for cluster j. The reweighted descriptors are used to learn the appearance for landmarks.

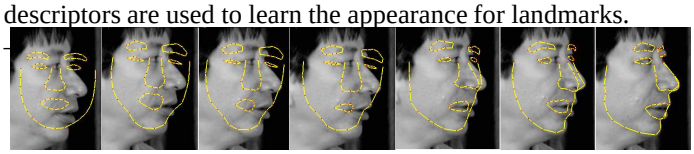

Fig. 6. Fitting the Profile Shape

2.5. Shape Search

The aligned shape is obtained by maximizing the posterior(3) by either EM algorithm [3, 9] or sampling from the distribution and evaluating the likelihood[2, 1]. These methods cannot be applied to our framework as the cluster based prior shape model does not have a functional form. Since the shapes do not lie on a common tangent space, the euclidean metric cannot be used to compare shapes. We maximize the posterior using alternating optimization of likelihood by sampling along the normal of the landmarks followed by shape regularization using the cluster based prior model. The shape regularization is done as follows - (1) project the shape onto global tangent space and compute class conditionals, (2) project the shape to local tangent space of the cluster with maximum likelihood and constrain the shape to lie within its subspace.

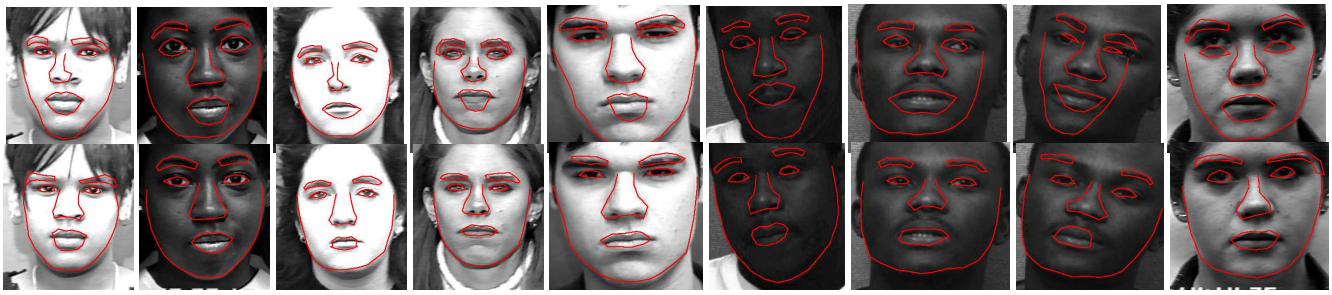

Fig. 7. *(Top)* Facial feature localization using ASM with gradient profiles. *(Bottom)*Localization using local descriptors as SIFT features. Notice the accurate localization of eye features due to SIFT descriptors

The overlapping between the clusters ensures smooth traversal across subspaces during search. Fig. 5 illustrates the algorithm showing the trajectory of the shape search and fig. 6 shows the iterative steps for fitting shape to a profile face.

3. EVALUATION

The prior ASM model is learned using 1029 labeled images (79 landmark points) in various head poses. We use coarse-to-fine search over 4 levels of gaussian scale pyramid. The SIFT block contained 4x4 cells with 4x4 pixels and 8 gradient orientation bins thus having descriptor size 128. The orientation quantization was done using angular interpolation with cutoff value as 0.2 to minimize the effect of extreme non-linear illumination changes. We tested on 342 unseen images spanning 3 category of poses - Left facing, Frontal and Right facing. Table 1 compares the average errors in pixels. The likelihood function enables more accurate alignment compared to the gradient profile used in conventional ASM. The RCA improves the alignment accuracy for the left and right facing poses. Fig. 7 provides qualitative comparisons of shape registration in extreme conditions of illumination and skin color. The SIFT descriptors are not only robust to variations in skin color, that are common among subjects, but also to the changes in illumination.

Algorithm	**Frontal**	**Left**	**Right**
Gradient profile	10.24	12.46	11.89
SIFT Descr.	7.09	10.11	9.13
SIFT Descr. + RCA	8.14	8.93	7.76

Table 1. Average errors in pixels for different algorithms.

4. CONCLUSION

In this work [2] we have advocated use of cluster based approach to learn non-linear shape manifold. This combined with the robust likelihood function allows us to scale the face registration algorithms to larger database having more variation in shapes and appearance. Empirically we observed that the system improves the accuracy of shape alignment and provides a groundwork for a generic shape registration framework.

[2]Protected by patenting and trademarking office (provisional patent #60/874,451). No part of this technology may be reproduced or displayed in any form without the prior written permission of the authors

5. REFERENCES

[1] J. Tu, Z. Zhang, Z. Zeng, and T. Huang, *Face Localization via Hierarchical CONDENSATION with Fisher Boosting Feature Selection*, CVPR, 2004.

[2] L. Liang, F. Wen, Y. Xu, X. Tang, and H. Shum, *Accurate Face Alignment using Shape Constrained Markov Network*, CVPR, 2006.

[3] Y. Zhou, L. Gu, and H. Zhang, *Bayesian Tangent Shape Model: Estimating Shape and Pose Parameters via Bayesian Inference*, CVPR, 2005.

[4] C. Bregler and S. Omohundro, *Surface Learning with Applications to Lipreading*, NIPS, 1994.

[5] T. Heap and D. Hogg, *Improving specificity in pdms using a hierarchical approach*, BMVC, 1997.

[6] G. Hamerly and C. Elkan, *Learning the k in k-means*, NIPS, 2003.

[7] A. Bar-hillel, T. Hertz, N. Shental, and D. Weinshall, *Learning distance functions using equivalence relations*, ICML, 2003.

[8] S. Romdhani, S. Gong, and A. Psarrou, *A Multi-View Nonlinear Active Shape Model Using Kernel PCA*, BMVC, 1999.

[9] Y. Zhou, W. Zhang, X. Tang, and H. Shum, *A Bayesian Mixture Model for Multi-view Face Alignment*, CVPR, 2005.