Emblem Detections by Tracking Facial Features

Atul Kanaujia
CBIM
Rutgers University
kanaujia@cs.rutgers.edu

Yuchi Huang
CBIM
Rutgers University
yuchuang@cs.rutgers.edu

Dimitris Metaxas
CBIM
Rutgers University
dnm@cs.rutgers.edu

Abstract

Tracking facial features across large head rotations is a challenging research problem. Both 2D and 3D model based approaches have been proposed for feature analysis from multiple views. Accurate feature tracking enables useful video processing applications like emblem detection(an event or movement that symbolizes an idea), facial expressions recognition, morphing and synthesis. A crucial requirement is generalizability of the tracking framework across appearance variations, presence of facial hair and illumination changes.

We propose a framework to detect emblems that combines active shape model with a predictive face aspect model to track features across large head movements and runs close to real time. Active Shape Model(ASM) is a deformable model for shape registration that detect facial features by combining prior shape information with the observed image data. Our view based framework represents various head poses by multiple 2D shape models and accounts for large head rotations by dynamically switching between them. Our switching variable (the current model to use) is discriminatively predicted from the SIFT descriptors computed over the bounding box of low resolution face image. We demonstrate the use of tracking framework to recognize high level events like head nodding, shaking and eye blinking.

1. Motivation

Active Appearance Model(AAM) and Active Shape Model(ASM) are generative models proposed by cootes [3] to detect facial features, and have been successfully used in the past to track faces in video. Perhaps the most relevant application of tracking facial features is for detecting emblems, facial expression recognition and synthesis.

An Emblem denotes an event or movement that symbolizes an idea. Head nodding and shaking are the prominent emblems which are frequently encountered during any conversation and denotes agreement or dissent. Emblems associated with head movement like head tilt and frequent eye blinking further convey different cognitive and emotional states of the subject during interrogation and interviewing.

A key challenge to track facial features in a generic setting is to handle local variations in feature shapes across different expressions and large head movements. The model should also be consistent due to changes in illumination, facial appearance due to skin color or presence of hair and glasses. While the past research have demonstrated successful subject specific tracking of facial features across large head rotations[17], little importance has been given to improve adaptability of the tracking framework to variations.

Therefore we encourage global shape model that uses prior shape information to constrain the parametric search for feature localization as opposed to appearance based face models. Active Shape Models (ASM), as proposed by cootes [3], is an iterative framework to combine global shape model with the localized appearance model for tracking facial features.

However traditional ASM cannot represent facial expressions accurately due to several reasons. For limited training data, the face alignment is constrained heavily due to global shape and cannot capture local deformations well. The projection subspace obtained from PCA cannot represent variations in localized parts, therefore it does not naturally decompose in subsets of localized deformations, as required for facial expression tracking.

Tracking facial expressions requires both low level modeling for feature shape variations and high level adaptation to global shapes due to head movement. A bottom up strategy that examines data at low level and tries to identify features by extracting local structures, is as inconsistent as a top down approach which puts too much constraints on the global shape thereby losing local feature shape information. Whereas both strategies have their cons: local models being less adapted to occlusion, scale and lighting variations whereas global models do not allow arbitrary variations in spatial locations of different parts.

A way to meet these requirements is by using multiple shape models which is composed of a generative localized

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

Figure 1. **Top Row** Extreme facial features alignment using global shape model learned by localized Nonnegative Matrix Factorization. The model was trained over database of 200 faces. **Bottom Row** shows the fitting results from shape model learn by PCA. Notice the inaccurate alignment of mouth and eye brow features. This occurs due to holistic basis vectors obtained from PCA that over constrains the local fitting by the global shape

feature shape tracking model and a discriminative global aspect switching model. At each step we track facial features using a specific Active Shape Model(ASM) and switch between multiple ASM models using a discriminative predictor, to account for aspect change of the face. Our 2D shape models extends Active Shape Models[3](ASM) to achieve more robustness and are applicable for tracking both subtle and extreme facial expressions. We prefer ASM model over more accurate AAM model due to its faster convergence and robustness to variations in facial appearances.

Our shape models are based on a technique known as Non-negative Matrix Factorization(NMF)[9] as opposed to PCA used in traditional ASM. NMF provides a sparse, part based representation of face shape models that correspond better to the intuitive notions of the parts of face. We provide both qualitative and quantitative results to show that NMF basis subspace can model larger variations in the facial expressions and improves the feature alignment.

2. Related Work

In vision literature, face tracking and recognition mostly involved fronto-parallel views. Efforts have been made to represent 3D face by a full 3D model [7],[6] or by using non-linear statistical models for 2D shape manifold learning [14]. An approach similar to what we propose in this paper has also been addressed in context with the AAM [4][13]. Specifically, they use multiple appearance models to represent various aspects of the face. Xiao et. al uses [17][12] AAM based 2D triangulated mesh to track face across large rotations.

Amongst 3D model based facial feature tracker, a relevant method is proposed Dornaika et al. [6] who uses particle filters to track features using a deformable model. They also use appearance model to track the features. [16] proposed a tracker that fits a 3D rigid face mask using key frame techniques and feature-based structure-from motion estimation methods to prevent drifts. The method is well suited for augmented reality applications as the 3D face mask is rigid and cannot be used to track facial expressions. Another closely related type of face models are 3D morphable models(3DMM) proposed by Brand [1] which uses appearance based triangulated 3D deformable mesh to fit the face. [2] proposed a method to track head under varying illumination conditions and modeled head as a cylinder. [7] proposed a method to track dense deformable face using statistical integration of multiple 2D image cues. Despite being powerful, the above 3D model based approaches require manual initialization and complex 3D mask model to account for large rotations and self occlusion. The global appearance based modeling is more susceptible to changes due to illumination and subject.

3. Learning localized 2D shape models using Non-negative Matrix Factorization

In order to track micro-expressions accurately, shape models need to be trained on a large number of subjects

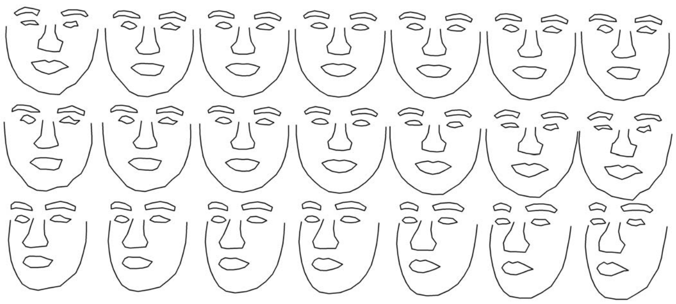

Figure 2. **Top and Middle Row** Facial shapes generated by varying the encodings of NMF basis of frontal shapes **Bottom Row** Shapes generated fromthe NMF basis of left aspect model. In all the rows, middle column is the mean shape. The shapes were obtained by decreasing (left 3) and increasing (right 3) the encodings. Unlike holistic PCA basis, NMF basis causes localized shape deformations.

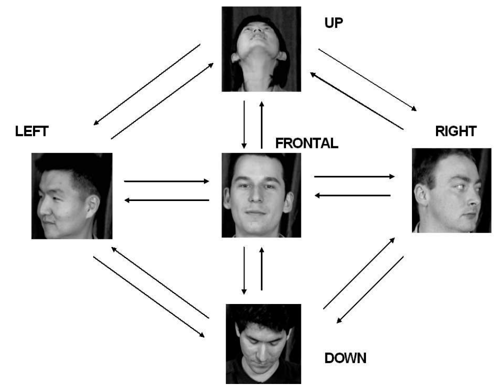

Figure 5. The 2D shape model can only switch from one face pose to another as shown in the figure. For smoother transitions, there is $\approx 10^o$ overlap in the range of rotation angles handled by each ASM model.

with different expressions. Due to the holistic representation of PCA, projection subspace cannot effectively characterize localized deformation of facial features.

3.1. Global Shape Model

Non-negative matrix factorization facilitate learning semantic information about the data [9] by factorizing data points into non-negative factor matrices. Unlike PCA which generates basis vectors having holistic representation, NMF bases vectors provide sparse localized representations of facial feature shapes.

NMF has been used effectively in the past to improve face recognition [8] [10]. The basis vectors obtained from NMF are non-overlapping and are more spatially localized compared to ICA and PCA. Non-negative Matrix Factorization(NMF) is a linear approximation of the non-negative observed data matrix $\mathbf{F} = \{\mathbf{f_1}, \mathbf{f_2}, \cdots, \mathbf{f_N}\}^\mathbf{T}$ by non-negative factors **W** and **H** as

$$\mathbf{f_i} \approx \sum_{j=1}^{M} \mathbf{w_j} h_{ji}^T = \mathbf{W}\mathbf{h_i^T} \tag{1}$$

where $\mathbf{W}$ can be thought as basis vectors $\mathbf{w_j}$ and $\mathbf{h_i}$ as the encodings. Projection subspace representations like PCA and ICA are a form of factorization methods with no constraints on the coefficients h_{ji} and basis vectors. Due to additional non-negativity constraints of the encodings, the observed data can only be expressed as a linear combination of non-negative basis vectors and coefficients. This leads to the idea of reconstructing an object by adding bases vectors that represents parts of the object. The factorization cannot be computed in a closed form and needs to be done iteratively by minimizing the error(divergence) between the observed data **F** and the reconstructed data **WH**.

$$D(\mathbf{F}||\mathbf{WH}) = \sum_{i=1}^{N}\sum_{j=1}^{M}[f_{ij}\log(WH)_{ij} - (WH)_{ij}] \tag{2}$$

subject to the constraint **W, H** ≥ 0 and $\sum_i w_i j = 1$. Seung et al. [9] came up with an iterative multiplicative updating algorithm to estimate **W** and **H**. For the data point $f_{ij} = \sum_{k=1}^{R} w_{ik}h_{kj}$

$$h_{kj} \leftarrow h_{kj} \sum_{i=1}^{N} \frac{w_{ik}f_{ij}}{\sum_{r=1}^{R} w_{ir}h_{rj}} \tag{3}$$

$$w_{ik} \leftarrow w_{ik} \sum_{j=1}^{M} \frac{f_{ij}h_{kj}}{\sum_{r=1}^{R} w_{ir}h_{rj}} \tag{4}$$

R denotes the rank of the basis matrix **W** and depends upon the number of basis vectors desired.

Several improvements in terms of sparsity and localization, have been proposed to the basic NMF algorithm [10], [8]. The improvement usually involves adding an additional constraint to the objective function to improve the non-negative decomposition. However the proposed algorithms are computationally expensive and cannot be used in our real-time framework. Encodings h_i for a new data point can be obtained by projecting it on pseudo inverse of the basis vectors **W** eqn. (5). Fig. 1 gives a qualitative comparison of face alignment results for ASM with PCA and NMF basis.

3.2. Local Profile Model

We retain the PCA model for learning 1D local profiles at feature points, due to its computational advantage. In order to avoid running iterative search to convergence every

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 **IEEE**

Figure 3. Here we demonstrate the tracking results on large head rotations and bending down poses

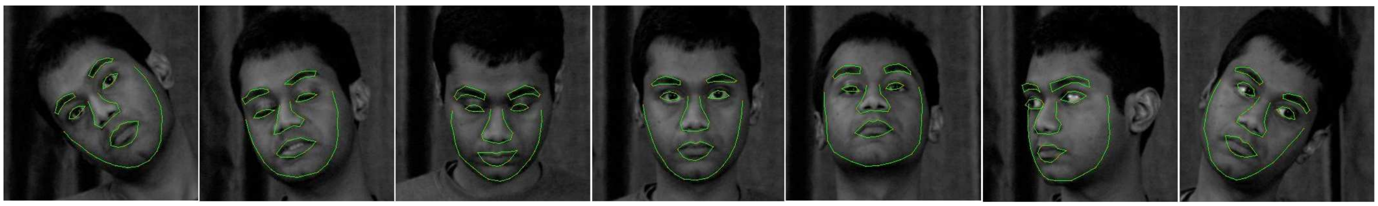

Figure 4. Our framework is robust to variations due to skin color. The above sequence illustrates accurate tracking of subject with dark skin color. The local texture model, learned from the first frame, enables robust tracking.

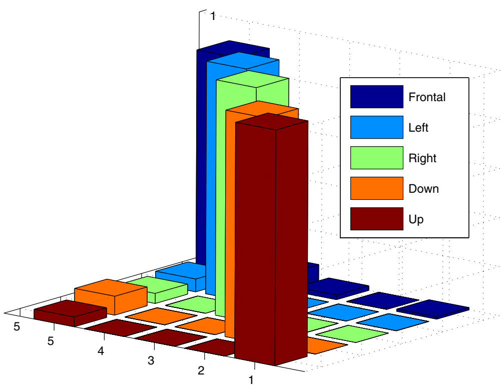

Figure 6. Bar plot representing the confusion matrix for the 5 classes of the multi-category classifier. Most of the inaccuracy is due to misclassification of Non-frontal poses to frontal pose. This does not cause tracker to fail at any point as the non-frontal 2D shape models can handle some frontal poses and vice-versa. The individual accuracy of each classifier were: Frontal -**97.3%**, Left -**91.2%**, Right - **94.6%**, Up - **92.7%** and Down - **91.8%**

time we re-initialize our 2D shape model, we adapt our profile model to the profile of the subject observed in the first frame. We run the face alignment to convergence only in the first frame and assume that the features are most accurately localized. The convergence of ASM is also improved by learning a 2D texture vector at each feature location and augmenting the profile matching cost function with the texture matching error. The texture vector encodes the gray level information at the feature points and adapts the model to the subject's face for faster and accurate search along the normal. For approximate frontal poses, it improves convergence rate and accuracy.

3.3. Global Shape Regularization

Generative face alignment in ASM proceeds as an iterative search to actively update the model parameters according to the observed image followed by the global shape regularization step to generate valid shape instances. Traditional ASM used truncation method to regularize shape by limiting the variance in a fixed range around the mean shape. With NMF framework we use continuous regularization by *Ridge Regression*

$$\mathbf{h_i} = (\mathbf{W^T W} + \lambda \mathbf{I})^{-1} \mathbf{W^T f_i} \tag{5}$$

The effect of shape regularization step (5) is to get rid of noisy distortions in the shape by preserving variance along major principal components only.

4. Discriminative Model selection using SIFT descriptors

The model for global shape variation should be sufficiently general so that it can fit a valid unseen example, but specific so as to not allow significantly different shapes. Large head rotations involves shape variations with strong non-linearities which cannot be modeled by linear statistical models like PCA and NMF. We use view based approach to represent dramatic shape variations in the dataset by multiple models. This is equivalent to representing non-linearities in the shape manifold by multiple piecewise linear models. We learn 5 different categories of ASM that model different aspects under which the face is viewed: frontal as well as with the head facing left, right, down, and up. During tracking, we discriminatively switch between these projection subspaces to account for substantial shape change. Each of the 5 subspaces are trained independently and the range of angles handled by the non-frontal

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

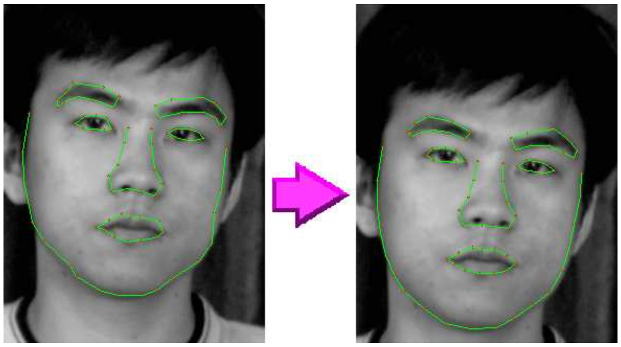

Figure 7. **Left** The feature points obtained after tracking using SSID tracker. The global shape is distorted even though the local fitting is good. **Right** Shape obtained after projecting the feature points to subspace and constraining the encodings within a variance threshold

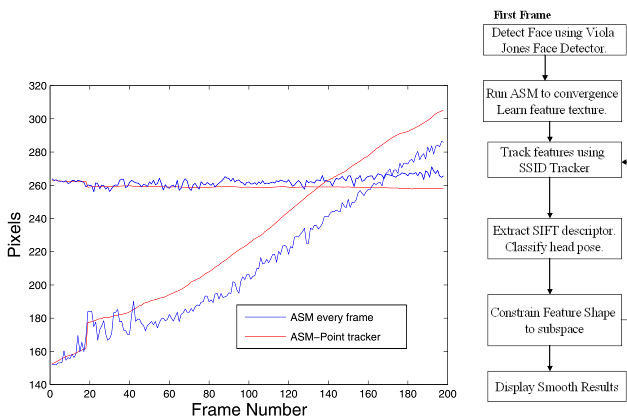

Figure 8. Here we compare the results of our tracker with ASM at every frame forthe two feature points - the outer contour and the nose tip. The blue plot shows the noisy movement of the y co-ordinates of the 2 landmark points using ASM at every frame. The red plot shows the result of our tracker. Due to inconsistent feature fitting across consecutive frames, ASM cannot be used to track facial expression at every frame . Tracking features points provides smoother and reliable feature localization to detect subtle changes in shape

subspaces overlaps with the frontal subspace in order to maintain continuity during tracking.

We use multi-class discriminative classifiers to recognize the head pose directly from the 2D descriptors extracted from the low resolution face image. Our 2D descriptor is based on scale invariant feature transform(SIFT) [11] computed over a fixed sized bounding box of the face. SIFT was primarily introduced to extract key feature points from an image and has been shown to work exceptionally well for image matching[11]. SIFT descriptor in conjunction with discriminative classifiers has been also demonstrated for pedestrians detection [5]. In effect, SIFT descriptors encodes the internal gradient information of the face to recognize the head pose with cluttered background reliably. Encodings based on histograms of gradient orientations allows us to capture essential spatial position and edge orientation information of the face. Quantizing gradient orientations into discrete values in small spatial cells and normalizing these distributions over local blocks makes the descriptor invariant to lighting changes.

The descriptors are computed over a grid of locations and the vector of the histogram values is normalized by L2-norm computed over the entire block. The normalized vector obtained for every block is concatenated to obtain the large dense descriptor vector. We combine several Relevance Vector Machine classifiers to obtain multi-category classifier using one-against-all classification. The SIFT descriptor is computed over a face bounding box(obtained from the previous frame) on a low resolution image in order to make processing real time.

To make model switching robust, we change the aspect face model only when the recognized pose is different from the current pose for 5 successive frames. We put further constraints on the switch between the models as outlined in the fig. 5.

5. Integrated Tracking Framework

Running ASM at every frame is not only computationally expensive but also causes feature points to jitter strongly. As most of features are corner points we track the features using Sum of Squared Intensity Difference(SSID) tracker across consecutive frames[15]. The SSID tracker is a method for registering two images and computes the feature displacement that minimizes the matching cost, computed as sum of squared intensity difference over a fixed sized feature window. Over a small inter-frame motion, a linear translation model can be accurately assumed.

For an intensity surface at image location $\mathbf{I}(\mathbf{x_i}, \mathbf{y_i}, \mathbf{t_k})$, the tracker estimates the displacement vector $\mathbf{d} = (\delta x_i, \delta y_i)$ from new image $\mathbf{I}(\mathbf{x_i} + \delta\mathbf{x}, \mathbf{y_i} + \delta\mathbf{y}, \mathbf{t_{k+1}})$ by minimizing the residual error over a window $\mathcal{W}$ around (x_i, y_i) [15]

$$\int_{\mathcal{W}} [I(x_i + \delta x, y_i + \delta y, t_{k+1}) - \mathbf{g}.\mathbf{d} - I(x_i, y_i, t_k)]\, d\mathcal{W} \tag{6}$$

The inter-frame image warping model assumes that for small displacements of intensity surface of image window $\mathcal{W}$, the horizontal and vertical displacement of the surface at a point (x_i, y_i) is a function of gradient vector $\mathbf{g}$ at that point. During the tracking, some features eventually lose track due to blurring or illumination changes. In order to avoid this, we re-initialize the points which have lost track by local profile search along the normal. Inaccuracies in tracking cause feature shapes to get distorted within 10-15

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 **IEEE**

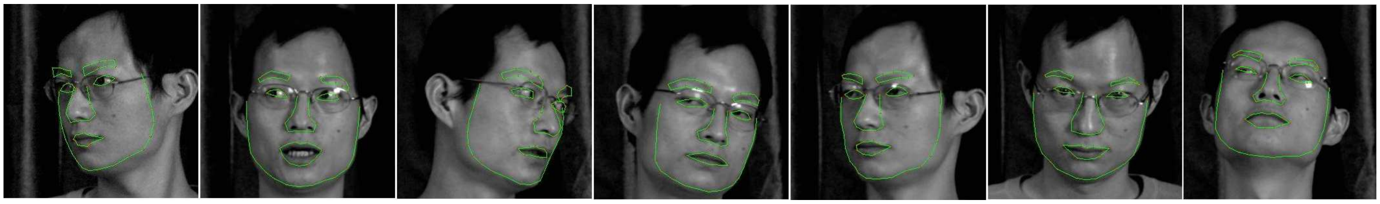

Figure 9. Wearing glasses does not affect the tracking accuracy. Tracking using ASM is more governed by the shape of the facial features and to some extent constrained by the local appearance model. The shape of the glasses are usually very different from shape of the eyes. The eyes contour are accurately detected in the above images.

frames. We ensure that the feature shapes $\mathbf{X}'$ obtained after tracking continue to be valid shape instances. The global shape parameters are estimated by aligning $\mathbf{X}'$ to the average shape of the current aspect subspace(inferred from the classifier)

$$\|\mathbf{X}' - \mathcal{T}_{\Delta x, \Delta y, s, \theta}(\bar{\mathbf{X}} + WH)\| \tag{7}$$

where $\mathcal{T}_{\Delta x, \Delta y, s, \theta}$ represents the global translation, rotation and scaling parameters[3]. The aligned shape is regularized using (5) to ensure that it is a valid face shape instance. In effect, we constrain the tracking of feature points in the NMF basis subspace to ensure that global shape of the face is preserved across tracking.

Switching between different aspect models causes discontinuities in the tracking. The final step to smooth the tracking over the past 7 frames generates smooth tracking results. Fig. 8(right) summarizes all the steps involved in the tracking framework. Fig. 8(left) compares the plot of y co-ordinates of nose tip and outer contour feature across 200 frames.

6. Emblem Detection

Robust tracking of facial features enables development of applications involving large scale video processing and high level event recognition. Detecting emblems during an interrogation provides useful cues about psychological state of the subject. In this paper we demonstrate the application of tracking results to detect head shaking, head nodding and eye blinking.

6.1. Eye Blinking

In order to detect eye blinking robustly, we need to make sure that feature points around eye do not lose track during blinking. Assuming that during eye blinking the eye region texture undergoes noticeable change, for a limited out-of plane head motion, we can construct a 2 state finite automata denoting eye close and eye open state. We assume open eyes in the first few frames of the video. During tracking the texture of open eye region(obtained for the first frame) is matched with the eye region in each frame. If there is substantial change in the texture(beyond a threshold) we assume it is due to eye blinking. For subsequent frames eye blinking is detected based on which of the 2 states the texture matching cost is minimum.

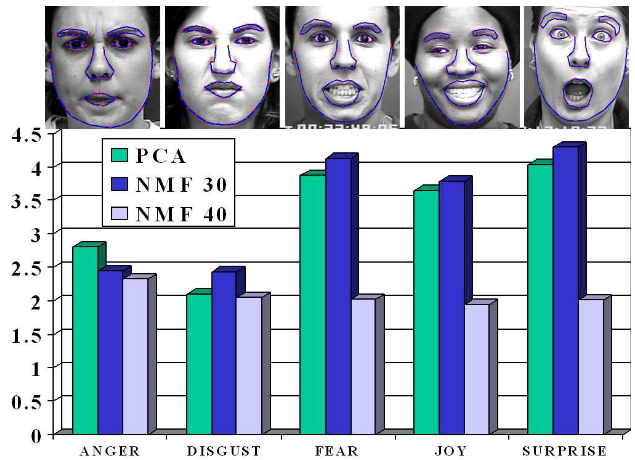

Figure 10. Comparative Results for NMF with PCA on 5 universal facial expressions from Cohn-Kanade database. The bar plot indicates the average pixel error with respect to hand labeled features. Training and testing were performed for each class of expressions separately. For PCA we used basis vectors capturing 98% of variance and ranged between 30 and 40. For NMF we compare results with 30 and 40 basis vectors. NMF show insignificant improvement for expressions involving minor shape deformation(anger and disgust). However for extreme facial deformations there is noticeable improvement.

6.2. Head Nodding and Head Shaking

Nose tip is the most stable tracked point as nose undergoes minimal deformation during tracking. Nose tip is approximated as average of lower 6 points of the nose feature. Head nodding and shaking can be recognized by detecting undulating patterns in y and x co-ordinates of nose tip respectively. Fig. 13 illustrates the nodding sequence detection plots.

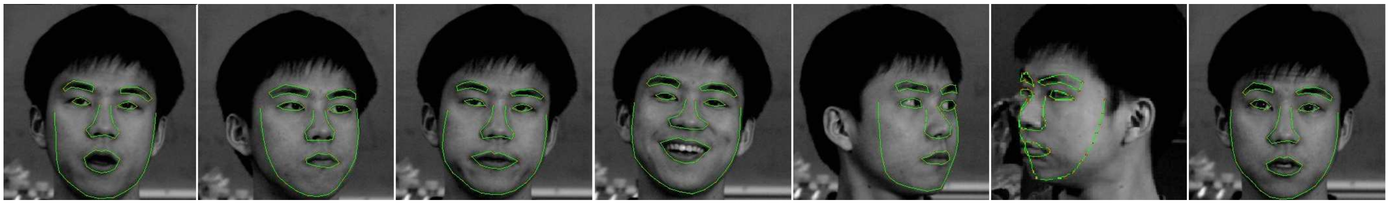

Figure 11. Tracking facial features with cluttered background. The results show no degradation for the tracking results and the facial features are tracked accurately

Figure 12. Active Shape Models are robust to presence of facial hair. In our tracking framework we adjust the local texture and profile model to the face appearance of the subject. The tracker can accurately track the lips and the outer contour in the presence of beard and moustaches

7. Experimental Results

The localized 2D shape models were trained on face images with different expressions and under varying backgrounds. Frontal model is integral to overall functioning of the tracker and was trained on 200 images while rest of the models were trained on 100 images of resolution 640×480. Our shape models contained 79 feature locations. The rank of the basis matrix $\mathbf{W}$ for Non-negative factorization was empirically adjusted to give best performance and speed.

Fig. 1 shows a qualitative comparison between the shape models learned using PCA and NMF. We run a variety of experiments to do quantitative comparison of PCA with NMF on sequences involving large facial deformation. Fig 10 compares the average pixel error due to frontal shape models based on NMF and PCA, on facial expression sequences from Cohn-Kanade database. The training was done on 75 images for each class of expression. The test subjects performed the same expressions but were not included in the training set. There is a marked improvement in feature localization for the expressions with larger shape deformation.

Tracker is initialized by detecting frontal and expression-free face in the first frame using Viola-Jones face detector. The face alignment algorithm is run to convergence to fit 2D shape model accurately to the subject. The intensity profile and the texture at feature locations are used to adjust the shape model to the subject's face. Subsequent local model search is executed for fixed number of iterations only. The customized 2D shape model converges faster and more accurately. We use 4 levels of resolution pyramid to do local search for the feature points and 2 levels of gradient pyramid for the point tracker. Once initialized, our tracker runs at 25 frames per second on PC with the configuration 3.5 GHz CPU and 3 GB RAM. For computation of the SIFT descriptors, we used 4×4 pixel cells for computing orientation of gradients and 4×4 cells in each block. For 8 angular bins, the descriptor vector was of size 128. We used low resolution face image at the the pyramid level 3(width is 1/4 of the original image) to compute the descriptor. The descriptor vector obtained by concatenating the histogram bin values over the entire bounding box of size 80×80 had a size of 3200. The SIFT descriptor based classifier to recognize head pose has an average accuracy of $\approx 93\%$. We used a set of 5 independent classifiers to model this as a multi-category problem. Fig. 6 shows the accuracy bar plots for the confusion matrix of the classifier.

At this stage we have not trained the shape models for the full profile face. The tracker loses track frequently for the full profile aspect, due to self occlusions and dramatic change in feature shapes. We re-initialize the tracker if the number of features lost at any given frame exceeds minimum threshold. The re-initialization is done by running the ASM search for fixed number of iterations. As the previous location of the feature points and head pose are available, the search converges within few iterations.

Emblem detection accuracy rely upon the feature tracking results. Large out of plane head movement require more sphisticated algorithms for nodding and shaking detection. Fig. 13 shows the ROC curves obtained from the sequences of 6 subjects. For each subject 2000 frames of an interrogation session was analysed. In all the videos the out-of-plane head movement was limited to a range of -30^o to $+30^o$.

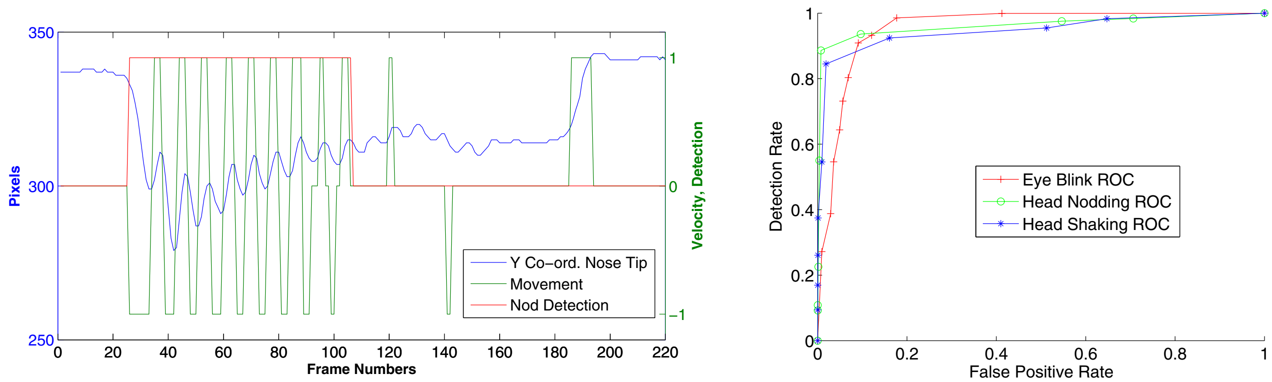

Figure 13. **(Left)** Blue plot shows the y co-ord. of nose tip. Green plot shows the velocity direction with +1 indicating downward movement and -1 indicating the upward movement. The red plot shows the detection by thresholding distance between consecutive upward and downward movements, with 1 indicating presence of nodding. Also notice that small change in Y co-ordinate is ignored to keep false positives minimal.**(Right)**ROC Curves for Eye blink detection(Area 95.46%), Head Nodding Detection(Area 96.273%) and Head Shaking detection(Area 94.693%). ROC curves were obtained by varying the thresholds involved in the detection algorithms.

8. Discussion

In this work we have proposed a generic 2D shape based framework for accurate tracking of facial features during off plane head movements. The system runs at 25 frames per second and has been demonstrated for online real-time tracking. It has been succesfully applied to online emblem detection with recognition rate of $\approx 95\%$.

The framework has been extended to offline analysis of interview videos, recorded under noisy conditions and requiring additional processing for accuracy. The results from facial analysis of the videos can be effectively applied for emblem detection and automated FACS coding for action units(AU) detection and recognition. Currently we are improving the accuracies of the framework for robust AU detection.

We provide videos to demonstrate results on different sequences involving large out of plane head movements and extreme feature shape deformations. Our tracking framework is robust to occlusions limited to less than 15 feature points. In future we plan to augment the system with an occlusion model to enable full profile face tracking and handle external obstruction.

We plan to investigate alternative algorithms to counter large illumination changes and shadowing. As our model is shape based, it is unaffected by minor variations in illumination or skin color. The tracker can also handle faces with hair and glasses to a reasonable accuracy.

References

[1] M. Brand. Morphable 3d models from video, 2001. CVPR. 2

[2] M. Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE PAMI*, 22(4), 2000. 2

[3] T. Cootes. Introduction to active shape models, 1998. Technical Report. 1, 2, 6

[4] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models, 2000. AFGR. 2

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection, 2005. CVPR. 5

[6] F. Dornaika and F. Davoine. Simultaneous facial action tracking and expression recognition using a particle filter. *ICCV*, 2005. 2

[7] S. Goldenstein, C. Vogler, and D. Metaxas. Statistical cue integration in dag deformable models, 2003. PAMI. 2

[8] P. Hoyer. Non-negative matrix factorization with sparseness constraints, 2004. JMLR. 3

[9] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization, 2000. MIT Technical Report. 2, 3

[10] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng. Learning spatially localized, parts-based representation, 2000. CVPR. 3

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints, 2004. IJCV. 5

[12] I. Matthews and S. Baker. Active appearance models revisited, 2004. IJCV. 2

[13] B. Moghaddam and A. Pentland. Face recognition using view-based and modular eigenspaces, 1994. SPIE. 2

[14] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernel pca, 1999. BMVC. 2

[15] C. Tomasi and T. Kanade. Detection and tracking of point features, 1996. Technical Report CMU-CS. 5

[16] L. Vacchetti, V. Lepetit, and P. Fua. Fusing online and offline information for stable 3d tracking in real-time, 2003. CVPR. 2

[17] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models, 2004. CVPR. 1, 2

0-7695-2646-2/06 $20.00 © 2006 **IEEE**