The Effect of Facial Components Decomposition on Face Tracking

Vikas Choudhary, Sergey Tulyakov and Venu Govindaraju
Center for Unified Biometrics and Sensors, University at Buffalo, State University of New York
(vikascho, tulyakov, govind)@buffalo.edu

Abstract

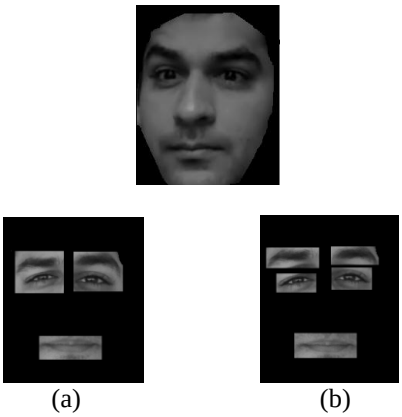
In this paper we investigate tracking of the facial motions under free head movement and changing expressions. We address this problem by examining the geometrical relationship within the facial components in order to improve their tracking. The likelihood scores of each component are combined to get the overall likelihood. Condensation has been used as the tracking algorithm. Histogram of Gradients (HOG) features have been introduced for facial features representation of the eyes, eye brows & lips which are individually classified using Support Vector Machine (SVM). We experimentally show the effect of the number of tracked components on the overall accuracy of facial tracking.

1. Introduction

Face tracking is essentially motion estimation and acts as essential pre-processing steps for high level applications in Surveillance, Biometrics (face gesture recognition), Multimedia systems (Video Games, video Conferencing) and Face recognition in videos. Face Tracking can be classified into two approaches: Global and Local. Face Tracking using Global features have been extensively studied in [1][2][3][4][5][6][7] which have used features based on Color, Contour, Shape, Texture and Probabilistic PCA. Local based approach [8][9][10][11] have used facial components, represented by features like Color, Contrast Histogram and Active Appearance model.

Generally, Face tracking algorithms approach is to improve on the tracking by using different features (local or global) and different tracking algorithms. We propose to investigate the geometrical positions of facial components (eyes, eye brows) and underlie the relative significance of them in order to enhance tracking.

There are two main important contributions in this paper. First, importance of the geometrical position within the facial components is highlighted. Second, we introduce HOG features for representing facial components along with SVMs for tracking. Depending on the expression of a person, relative position of the components vary, which affects the overall performance of tracking. Facial components can be effectively represented by the distribution of local intensity gradients or edge directions. HOG have been used by [12][13] for face recognition and human detection. We split the face into components (Fig 1) and train SVM classifier on each component.



(a) (b)

Fig1: Facial Components considered: Eyes, Eye Brows &Lips (a) contains 3-components & (b) contains 5-components.

2. Component Representation

The HOG (Histogram of Oriented Gradients) feature represents the distribution of the gradient magnitude and orientations of the pixels in an image. The descriptor is implemented by dividing the image into overlapping cells, each contributing the edge orientation for the pixels within the cell,fig(2).In this paper, we assume the centre of the image as the landmarks point. Hence, the HOG descriptor is the last part of the SIFT algorithm as we do not need to find the key point. The descriptor is scale invariant as it operates on localized cells.

For every each pixel I(x,y) gradient magnitude G(x,y) and orientation θ(x,y) is computed using the following equations:

$$px(x,y) = I(x+1, y) - I(x-1, y) \quad (1)$$

$$py(x,y) = I(x, y+1) - I(x, y-1) \quad (2)$$

$$G(x,y) = \sqrt{px(x,y)^2 + py(x,y)^2} \quad (3)$$

$$\theta(x,y) = \arctan(px(x,y)/py(x,y)) \quad (4)$$

We divide the gradients into b = 9 bins of size $2\pi/9$ in the range $[-\pi, \pi]$. Each HOG descriptor of an image is a histogram of gradients binned according their respective orientations. The combination of these histograms then represents the descriptor. In our case we have 4 overlapping cells each contributing 9 orientations. Hence, the dimensionality is 4x9 = 36.

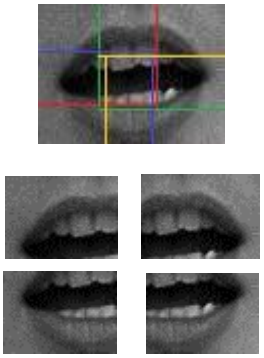


Fig2. The spatial lattice of a HOG descriptor centred on the lip. (4 overlapping cells).

A Gaussian window is centred at the image with standard deviation half the extension of the spatial bin range. The contribution of each pixel gradient to the histogram is weighted by the gradient modulus and its distance from the Gaussian window. The contribution is tri-linearly interpolated with the surrounding bins. Tri-linear interpolation distributes the weight *w* of the bin *h(x)* into two nearest neighbours (*x*1&*x*2) as follows, where b=9.

$$h(x1) = h(x1) + w\left(1 - \frac{x-x1}{b}\right) \quad (5)$$

$$h(x2) = h(x2) + w\left(\frac{x-x1}{b}\right) \quad (6)$$

Gaussian windowing and tri-linear interpolation increases the robustness of the descriptor. To handle variance in the illumination HOG feature vector v is normalized using the following equation

$$v = v / \sqrt{|v|^2 + \varepsilon^2} \quad (7)$$

ε is the small normalization constant to avoid division by zero.

3. CONDENSATION - Tracking By Parts

CONDENSATION [14] stands for "Conditional Density Propagation" and is based on particle filter techniques in order to track objects. Density represents a probability distribution of the location of the object. It is based on factored sampling, a method that aims at transforming uniform densities into weighted densities, but applied iteratively.

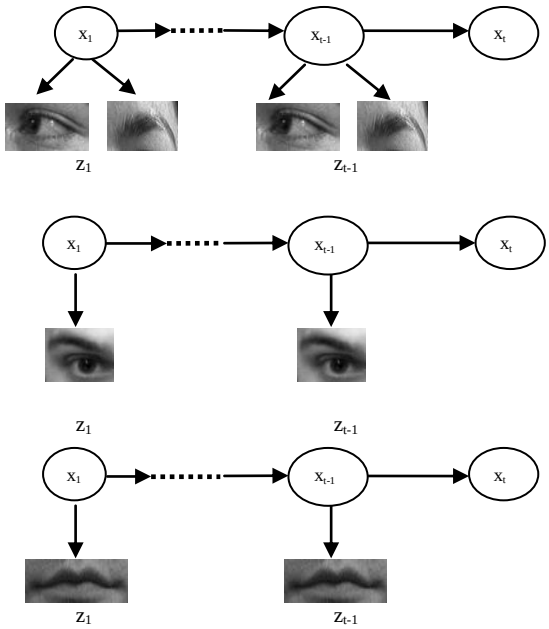


Fig 3: Bayesian Networks used in the Tracking by parts tracking algorithm. Each BN is a single chain. First BN combines the observation from eyes & brows to obtain the likelihood. Second and Third BNs compute likelihood using one component.

We have divided our tracking algorithm into two parts: 3-component and 5-component based. Each component is represented by a single chain BN as shown in Fig3. State estimation of each component is independent of each other. Eyes and eye brows are tracked together by using the first BN as we consider the spatial coherence between the two components. Using the second BN, combined patch of Eye and Eye brow is tracked. Both the left and right eye & eye brows are tracked independently. To track lips we use the third BN. Tracking by parts approach has been used by [15][16][17].
In the 3 component based approach we have used 2nd and 3rd BN, shown in Fig3. We are using the 1st and 3rd BN for the 5 component based approach. Using these combinations of Bayesian Networks we intend to study the geometrical relationship between eye and eye brows.

3.1 Algorithm

1: Use AdaBoost Classifier [18] to detect the face and separately initialize the tracking algorithm for each component (each component is initialized with 4 state variables: centre x&y coordinates and width & height)
2: Generate a set p_i of weighted particles for each component of the face.

$$p_i = (s^i_{t:n}, \pi^i_{t:n}) \quad (8)$$

$s^i_{t:n}$ is a particle and $\pi^i_{t:n}$ is weighted associated to that particle.

3: Update Weight

- Calculate HOG features for each component and pass them to the SVM classifier to compute the likelihood score.
- The joint likelihood of the weighted particles is the combination of all particles derived from each component.

$$p(z_t \mid x_t^1, x_t^2) = \prod_{i=1}^{2} p(z_t^i \mid x_t^i) \quad (9)$$

Joint likelihood for eye and eye brow shown in the first BN of fig 3 is calculated using the above equation. For rest of the components, likelihood is simply represented by $p(z_t^i \mid x_t^i)$.

Weight update step is being shown in the dashed rectangle in fig 4. Values with greater weight span across a greater range in the cumulative distribution, thus elements with higher probability are chosen several times while others with lower probability are not chosen. CONDENSATION algorithm applies this factored sampling iteratively to successive image frames in sequence, and each iteration uses prior density as the weighted sample set of the last iteration. In our experiment we have used 100 particles for every iteration. The posterior $p(x_t \mid z_{1:t})$ is calculated using Baye's rule:

$$\frac{p(z_t \mid x_t)\, p(x_t \mid z_{1:t-1})}{p(z_t \mid z_{1:t-1})}$$

$p(z_t \mid x_t)$ is computed using equation 9.

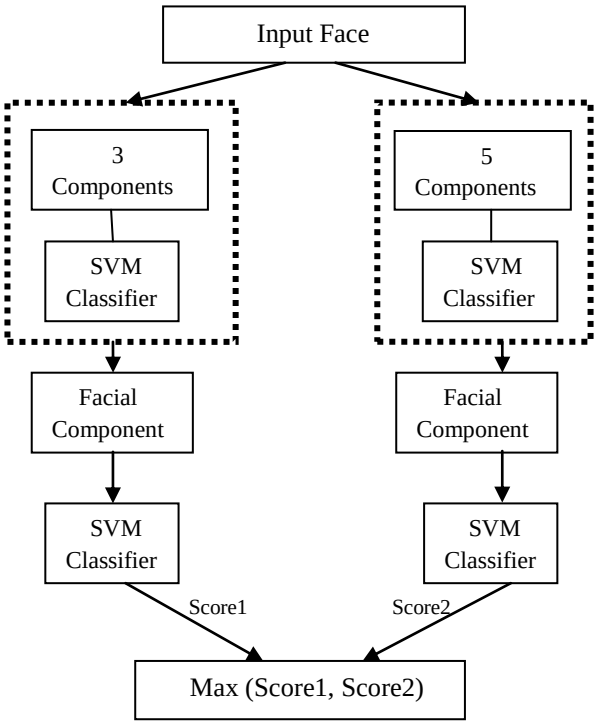


Fig 4: Flowchart of the part based tracking

After the end of the weight update step, condensation algorithm returns a mean patch of each component which is computed using the posterior calculated using Baye's rule. We compute the HOG features of each facial component and pass it to the SVM classifier (last two rectangles in Fig 4).

$$\text{Score1} = \prod_{k=1}^{3} p_k \quad (11)$$

$$\text{Score2} = \prod_{k=1}^{5} p_k \quad (12)$$

Score1 represents the product of 3 components while Score2 combines the 5 components score. Maximum of the two scores from eq. 11 & 12 represents the optimum tracking algorithm for that frame of video. Both the tracking algorithms run independently & parallel for each frame, and whenever we lose track in any one of them we reinitialize it using the step 1 of the tracking algorithm (3.1).

4. Data Set

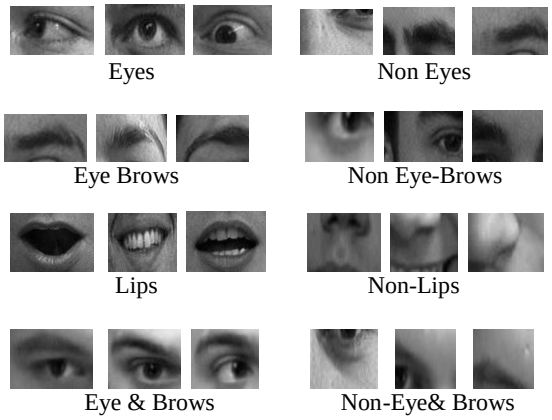


Fig3: Images from Dataset used for training the SVM.

We have extracted above images from The IMM Face Database - An Annotated Dataset of 240 Face Images [19] and videos in Mark Frank Dataset [20]. For each landmark point we have extracted 250 images to capture all variations. We have trained the 2-class SVM [21] with radial basis function kernel.

5. Results

We have performed a series of experiments on the videos sequences from Mark Franks Dataset used in [17] having 600 frames and the resolution is 720x480. To quantify the performance we have compared the performance of our algorithm separately with 3-component and 5-component based approach on Dudek sequence [22] shown in fig5. We measure accuracy every 10 frames by finding the area of intersection of the predicted area with ground truth.

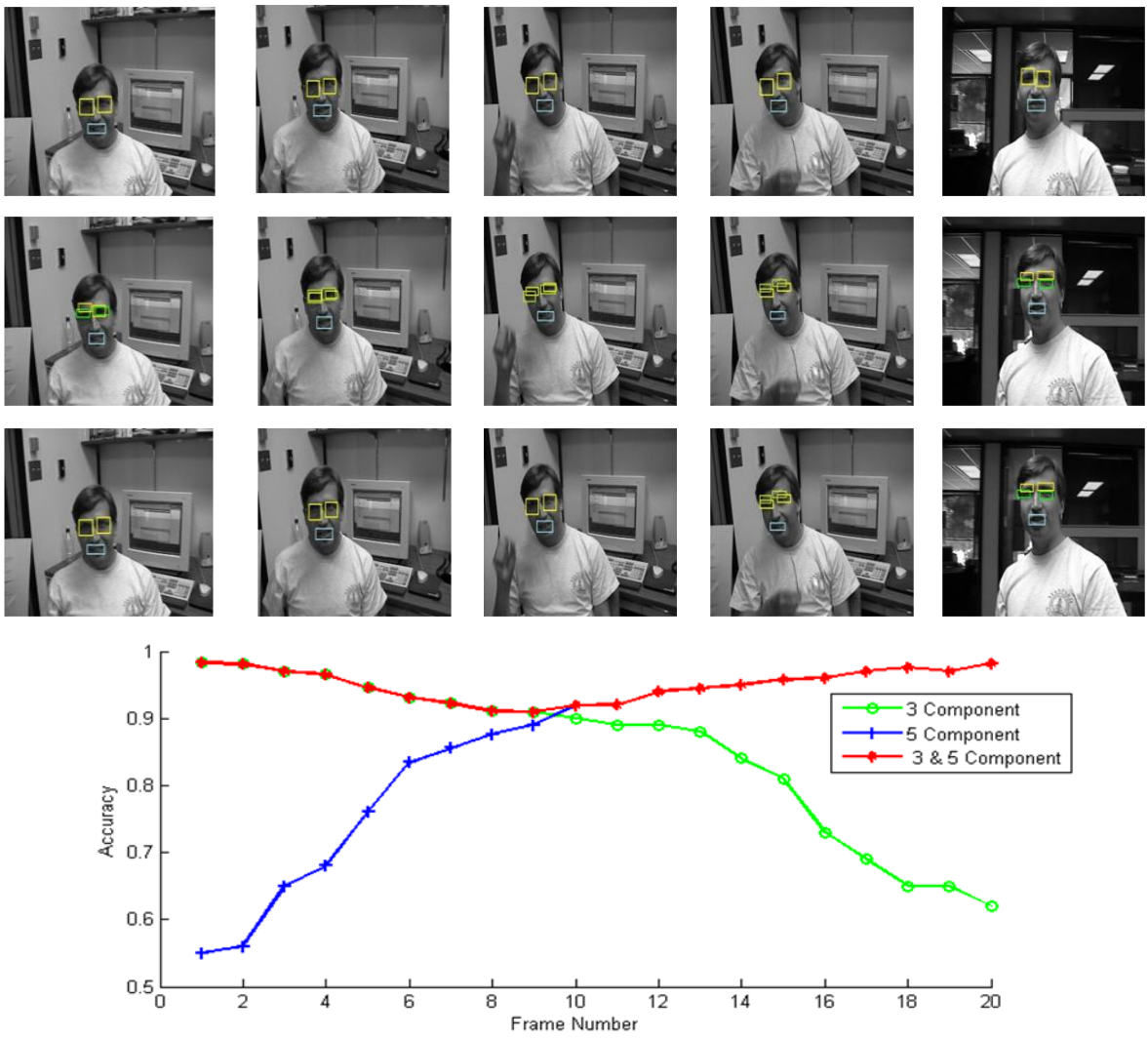


Figure 5: Tracking results on the Dudek sequence [22].(1st row : 3-component,2nd row-5component,3rd row-3&5 component) Errors are measured every 30 frames. Initially in first 3 sections the person's eye & eye brows remain geometrically close, the 3-component based approach dominates over the 5-component. In the final two sections, both components are apart, 5-component approach gives more accuracy. Overall accuracy in 3rd row is 92.45% while in 1st and 2nd its 71.4% & 77.4% respectively.

.

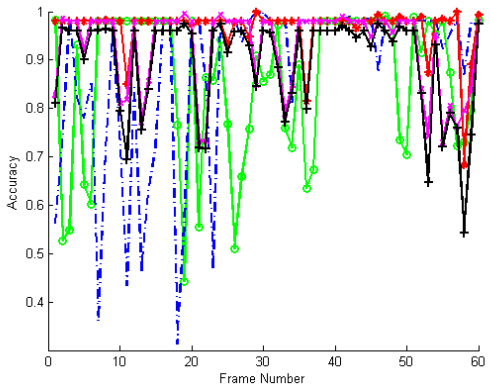
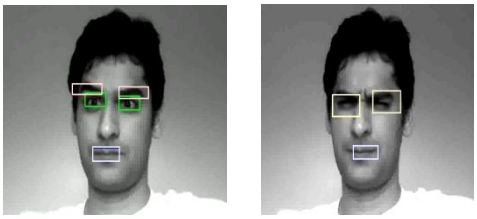


Fig 6: Results of Accuracy vs. Frame number (average accuracy of 88.67%)

Fig 6 compares illustrates accuracy computed over 5 videos from [17]. Initially, the accuracy is moderate, but once tracking algorithm stabilizes the accuracy increases. In these videos, we observed how their expression affects the tracking results,3-component based approach gives more accuracy for people who tend to bring eye & eye brow geometrically close in their expressions(illustrated in fig 7).



(a) (b)

Fig 7: An illustration of the effect of geometrical relationship between the eye and eye brows on tracking. When they are apart, as in (a), 5-component based gives more accuracy and while they come close geometrically (b) 3-component based approach gives more accurate result.

6. Conclusion

We have presented a comparison of component based approach for robust face tracking under free head movement and varying expressions. We have exploited the geometrical relationship between the eyes and eye brows by comparing their combined and separate scores. The tracking problem has been formulated as Bayesian Network. Histogram of Oriented Gradients features along with SVM has been introduced for robust tracking. In the future work, we will extend our approach by interactive collaboration of the 3-component and 5-component based methods using weight sharing.

References

[1] G. R. Bradski. "Computer Vision Face Tracking as a Component of a Perceptural User Interface". IEEE Work. On Applic. Comp. Vis., Princeton, pp. 214-219, 1998

[2] K. Schwerdt, J. Crowley, "Robust face tracking using color", in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2000, pp. 90–95

[3] X. Bing, Y. Wei, C. Charoensak, "Face contour tracking in video using active contour model", in: Proceedings of the European Conference on Computer Vision, 2004, pp. 1021–1024

[4] S. McKenna, S. Gong, R. Würtz, J. Tanner and D. Banin, "Tracking facial feature points with Gabor wavelets and shape models", Audio- and Video-Based Biometric Person Authentication. First International Conference, AVBPA'97. Proceedings, 1997, pp. 35-42.

[5] Kruger, V.; Happe, A.; Sommer, G.; , "Affine real-time face tracking using Gabor wavelet networks," Pattern Recognition, 2000. Proceedings. 15th International Conference on , vol.1, no., pp.127-130 vol.1, 2000 doi: 10.1109/ICPR.2000.905289

[6] Z. Liu and Y. Wang. "Face detection and tracking in video using dynamic programming". In ICIP, 2000.

[7] D. Ross, J. Lim, and M. H. Yang, "Adaptive probabilistic visual tracking with incremental subspace update," Proceedings of European Conference on Computer Vision, Vol. 2, pp. 470-482, 2004.

[8] I. Patras and M. Pantic, "Particle filtering with factorized likelihoods for tracking facial features," Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 97-102, 2004.

[9] Wen-Yan Chang; Chu-Song Chen; Yi-Ping Hung; , "Tracking by Parts: A Bayesian Approach With Component Collaboration," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol.39, no.2, pp.375-388, April 2009 doi: 10.1109/TSMCB.2008.2005417

[10] Jaewon Sung; Daijin Kim; , "Combining Local and Global Motion Estimators for Robust Face Tracking," Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on , vol., no., pp.345-350, 26-29 Aug. 2007

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. IEEE Trans. on Pattern Recognition and Machine Intelligence, 23(6):681-685, 2001

[12] Dalal, N.; Triggs, B.; , "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on , vol.1, no., pp.886-893 vol. 1, 25-25 June 2005

[13] Monzo, D.; Albiol, A.; Sastre, J.; , "HOG-EBGM vs. Gabor-EBGM," Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on , vol., no., pp.1636-1639, 12-15 Oct. 2008

[14] M. Isard and A. Blake, "ICONDENSATION-Unifying low-level and high-level tracking in a stochastic framework," Proceedings of European Conference on Computer Vision, Vol. 1, pp. 893-908, 1998.

[15] C. Yang, R. Duraiswami, and L. Davis, "Fast multiple object tracking via a hierarchical particle filter," Proceedings of IEEE InternationalConference on Computer Vision, Vol. 1, pp. 212-219, 2005.

[16] K. Okuma, A. Taleghani, N. De Freitas, J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," Proceedings of European Conference on Computer Vision, Vol. 1, pp. 28-39, 2004.

[17] Wen-Yan Chang; Chu-Song Chen; Yi-Ping Hung; , "Tracking by Parts: A Bayesian Approach With Component Collaboration," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.39, no.2, pp.375-388, April 2009 doi: 10.1109/TSMCB.2008.2005417

[18] Intel Corp.,Opencv Library, World Wide Web, http://www.intel.com/technology/computing/opencv

[19] The IMM Face Database - An Annotated Dataset of 240 Face Images Michael M Nordstrom, MadsLarsen, Janusz Sierakowski Mikkel B Stegmann

[20] N. Bhaskaran, I. Nwogu, M. Frank, and V. Govindaraju, "Lie To Me: Deceit Detection via Online Behavioral Learning", 9th IEEE Conference on Face and Gesture Recognition, Santa Barbara, CA.

[21] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. http://www.csie.ntu.edu.tw/~cjlin/libsvm

[22] Dudek Sequence: http://www.cs.toronto.edu/vis/projects/dudekfaceSequence.html