Robust Detection and Tracking of Human Faces with an Active Camera

Dorin Comaniciu   Visvanathan Ramesh

Imaging and Visualization Department, Siemens Corporate Research

755 College Road East, Princeton, NJ 08540

{comanici, vramesh}@scr.siemens.com

Abstract

We present an efficient framework for the detection and tracking of human faces with an active camera. The Bhattacharyya coefficient is employed as a similarity measure between the color distribution of the face model and face candidates. The proper derivation of these distributions allows the use of the spatial gradient of the Bhattacharyya coefficient to guide a fast search for the best face candidate. The optimization, which is based on mean shift analysis, requires only a few iterations to converge. Scale changes of the tracked face are handled by exploiting the scale invariance of the similarity measure and the luminance gradient computed on the border of the hypothesized face region. The detection and tracking modules are almost identical, the difference being that the detection involves mean shift optimization with multiple initializations. Our dual-mode implementation of the camera controller determines the pan, tilt, and zoom camera to switch between smooth pursuit and saccadic movements, as a function of the target presence in the fovea region. The resulting system runs in real-time on a standard PC, being robust to partial occlusion, clutter, face scale variations, rotations in depth, and fast changes in subject/camera position.

1 Introduction

The task of real-time detection and tracking of human faces is a key component of video surveillance and monitoring systems [7]. It provides input to high-level processing such as recognition [21], access control, or re-identification, or is used to initialize the analysis and classification of human activities [1, 13].

The head is one of the most easily recognizable human parts, having a relatively constant color composition in certain color sub-spaces [6, 27], a projection of elliptical shape in the image frame, and distinct local features. Although the illumination conditions can thoroughly affect the face appearance [16] and occlusions can modify its perceived shape, the color and shape are the visual clues most often used in detect and track faces [4, 5, 10, 11, 12, 15, 18, 22, 26] (see [24] for a discussion).

To achieve robustness to out-of-plane rotations of the head, the color distribution of a face model (target) is employed instead of raw image pixels. The location of the target in the new frame is predicted based on the past trajectory, and a search is performed in its neighborhood for image regions (target candidates) whose distribution is similar to that of the model. In single hypothesis tracking the best match determines the new location estimate, however, more complex strategies also exist to form multiple hypothesis [2].

The exhaustive search in the neighborhood of the predicted target location for the best target candidate is a computationally intensive process. Although the computing machinery becomes faster and faster, the complexity of the tracker is critical for most applications, since only a small percentage of a computer resources are allocated for tracking, the rest being reserved for high level tasks. In addition, the face detector should be fast enough to insure a fast tracker (re)initialization.

This paper presents a system that employs a new framework for the efficient detection and tracking of faces with an active camera. We compute the Bhattacharyya coefficient as the similarity measure between the face model (its color distribution derived in an intensity normalized space) and the target candidates, and use the spatial gradient of this measure to guide a fast search for the best candidate. The optimization we propose, based on mean shift analysis [8], achieves convergence in only a few iterations, being thus well suited for real-time tracking. To adapt to the scale changes of the target we exploit the scale invariance property of the Bhattacharyya coefficient as well as the gradient information on the border of the hypothesized face region.

The system operates either as a face detector, waiting for a face to appear in the active frame, or as a tracker, until the tracked face/head is completely occluded or exits the scene. The detection module is based on the same principle as the tracker, employing the mean shift optimization with multiple initializations (in different locations of the current image frame).

The active camera is driven by a two mode controller, being capable of both smooth pursuit and fast saccades, according to the position of the target in the current frame. The resulting detection and tracking system was tested on different and multiple persons and proved to

be robust to partial occlusion, significant clutter, target scale variations, rotations in depth, and fast changes in camera and subject positions.

The organization of the paper is as follows. Section 2 presents the Bhattacharyya coefficient as a color-based similarity measure. The derivation of the color histograms which incorporate spatial information is given in Section 3. Section 4 formulates the optimization problem for face localization in the neighborhood of a predicted location. The tracking process and scale adaptation scheme are discussed in Section 5. Section 6 describes the details of face detection for tracker initialization. The detection and tracking experiments are presented in Section 7 and some conclusions are drawn in Section 8.

2 A Similarity Measure for Color-Based Matching

As stated before, the single hypothesis tracking assumes the search of a neighborhood of the predicted target location in the current frame for the face candidate that is the most similar to the face model. The similarity measure we develop is based on color information. The feature $\mathbf{z}$ representing the color of the face model is assumed to have a density function $q_{\mathbf{z}}$, while the face candidate centered at location $\mathbf{y}$ has the feature distributed according to $p_{\mathbf{z}}(\mathbf{y})$. Then, the problem is to find the discrete location $\mathbf{y}$ whose associated density $p_{\mathbf{z}}(\mathbf{y})$ is the closest to the target density $q_{\mathbf{z}}$.

We estimate the similarity between two densities according to the Bhattacharyya coefficient, whose general form is defined by [17]

$$\rho(\mathbf{y}) \equiv \rho\left[p(\mathbf{y}), q\right] = \int \sqrt{p_{\mathbf{z}}(\mathbf{y})q_{\mathbf{z}}}\, d\mathbf{z} \ . \qquad (1)$$

The similarity measure (1) is valid for arbitrary densities, being invariant to the scale of the target [1]. We implicitly assume that the color distribution of the target is invariant to scale, although the quantization effects We recently showed [9] that the metric derived from the Bhattacharyya coefficient is superior to other measures such as histogram intersection [23], Bhattacharyya distance, Fisher linear discriminant or Kullback divergence. Some theoretical properties of the Bhattacharyya coefficient such as its relation to the Fisher measure of information and explicit forms for various distributions are given in [17].

The evaluation of (1) from sample data assumes the estimation of the densities p and q. Due to the computational complexity constraint we use the density estimates derived from a simple histogram formulation. The discrete density $\hat{\mathbf{q}} = \{\hat{q}_u\}_{u=1\dots m}$ (with

$\sum_{u=1}^m \hat{q}_u = 1$) is estimated from the m-bin histogram of the face model, while $\hat{\mathbf{p}}(\mathbf{y}) = \{\hat{p}_u(\mathbf{y})\}_{u=1\dots m}$ (with $\sum_{u=1}^m \hat{p}_u = 1$) is estimated at a given location $\mathbf{y}$ from the m-bin histogram of the face candidate. Therefore, the sample estimate of the Bhattacharyya coefficient is given by

$$\hat{\rho}(\mathbf{y}) \equiv \rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}. \qquad (2)$$

3 Including Spatial Information in Color Histograms

This section shows how to build from image regions color histograms that incorporate spatial information. We assume the existence of a known image region containing the face model and multiple image regions representing the face candidates.

Face Model   Let us denote by $\{\mathbf{x}_i^\star\}_{i=1\dots n}$ the pixel locations of the face model, centered at $\mathbf{0}$. Let $b : R^2 \to \{1\dots m\}$ be function which associates to the pixel at location $\mathbf{x}_i^\star$ the index $b(\mathbf{x}_i^\star)$ of the histogram bin corresponding to the color of that pixel. The probability of the color u in the face model is computed by employing a convex and monotonic decreasing function $k : [0, \infty) \to R$ which assigns a smaller weight to the locations farther from the center of the face [2]. This weighting increases the robustness of the estimation, since the peripheral pixels are the least reliable, being often affected by occlusions (clutter) or background. By assuming that the generic coordinates x and y are normalized with h_x and h_y, respectively, we can write

$$\hat{q}_u = C \sum_{i=1}^n k(\|\mathbf{x}_i^\star\|^2)\delta\left[b(\mathbf{x}_i^\star) - u\right] \ , \qquad (3)$$

where δ is the Kronecker delta function. The normalization constant

$$C = \frac{1}{\sum_{i=1}^n k(\|\mathbf{x}_i^\star\|^2)} \ , \qquad (4)$$

results from the condition $\sum_{u=1}^m \hat{q}_u = 1$ and by taking into account that the summation of delta functions for $u = 1\dots m$ is equal to one.

Face Candidates   Let us denote by $\{\mathbf{x}_i\}_{i=1\dots n_h}$ the pixel locations of the face candidate, centered at $\mathbf{y}$ in the current frame. Employing the same weighting function k, the probability of the color u in the face candidate is given by

$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right) \delta\left[b(\mathbf{x}_i) - u\right] \ . \qquad (5)$$

The scale of the face candidate (i.e., the number of pixels) is determined by the constant h which plays the

---

[1] In practice, the uncertainty in the density estimates is influenced by the number of available samples, i.e., is scale dependent.

[2] The function k is called a kernel profile when used in the context of density estimation [9].

same role as the bandwidth (radius) in the case of kernel density estimation [8]. By imposing the condition that $\sum_{u=1}^{m} \hat{p}_u = 1$ we obtain the normalization constant

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k(\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\|^2)}. \qquad (6)$$

Observe that C_h does not depend on $\mathbf{y}$, since the pixel locations $\mathbf{x}_i$ belong to a regular lattice, $\mathbf{y}$ being one of the lattice nodes. Thus, C_h can be precalculated for a given kernel and different values of h.

4 Optimization Problem for Face Localization

The search for the new face location in the current frame starts at the predicted location $\hat{\mathbf{y}}_0$ of the face computed from the previous frame. Therefore, we compute first the color probabilities $\{\hat{p}_u(\hat{\mathbf{y}}_0)\}_{u=1...m}$ of the face candidate at location $\hat{\mathbf{y}}_0$ in the current frame.

To maximize the Bhattacharyya coefficient (2), we start with the Taylor expansion of $\hat{\rho}(\mathbf{y})$ around the values $\hat{p}_u(\hat{\mathbf{y}}_0)$, which yields

$$\hat{\rho}(\mathbf{y}) \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(\mathbf{y})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}} \quad (7)$$

Introducing now (5) in (7) we obtain

$$\hat{\rho}(\mathbf{y}) \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{C_h}{2}\sum_{i=1}^{n_h} w_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)$$

where $\qquad\qquad\qquad\qquad\qquad\qquad\qquad (8)$

$$w_i = \sum_{u=1}^{m} \delta\left[b(\mathbf{x}_i) - u\right]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}}. \qquad (9)$$

Hence, to maximize (2) the second term in equation (8) has to be maximized, the first term being independent of $\mathbf{y}$. The second term represents the density estimate computed with kernel profile k at $\mathbf{y}$ in the current frame, with the data being weighted by w_i (9). The maximization can be efficiently achieved based on the mean shift iterations (see [8]), using the following algorithm. Figure 1 shows a block diagram of the algorithm.

Location Optimization

Given the distribution $\{\hat{q}_u\}_{u=1...m}$ of the face model and the predicted location $\hat{\mathbf{y}}_0$ of the face:

1. Compute the distribution $\{\hat{p}_u(\hat{\mathbf{y}}_0)\}_{u=1...m}$, and evaluate
$$\hat{\rho}(\hat{\mathbf{y}}_0) = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} .$$

2. Derive the weights $\{w_i\}_{i=1...n_h}$ according to (9).

3. Derive the new location of the face [8]
$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g\left(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h}\right\|^2\right)} . \qquad (10)$$

Update $\{\hat{p}_u(\hat{\mathbf{y}}_1)\}_{u=1...m}$, and evaluate
$$\hat{\rho}(\hat{\mathbf{y}}_1) = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_1)\hat{q}_u} .$$

4. While $\quad \hat{\rho}(\hat{\mathbf{y}}_1) < \hat{\rho}(\hat{\mathbf{y}}_0)$
   Do $\quad\quad \hat{\mathbf{y}}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$.

5. If $\quad \|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\| < \epsilon \quad$ Stop.
   Otherwise $\quad\quad$ Set $\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1$ and go to Step 1.
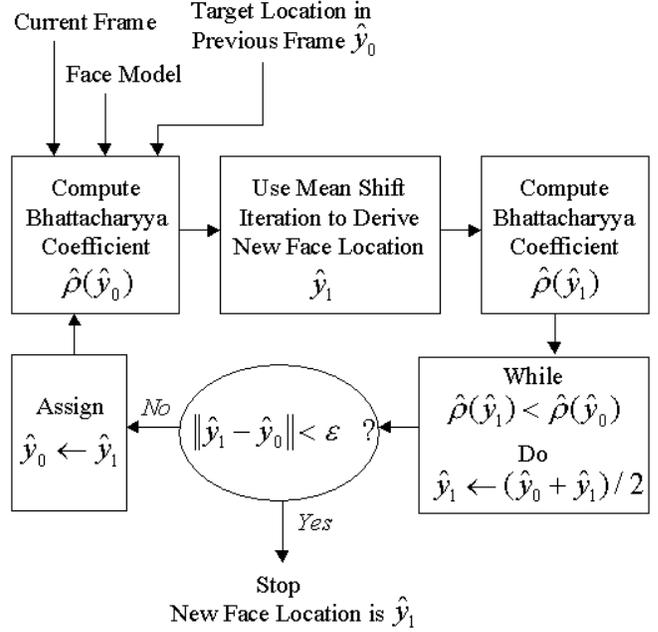

Figure 1: Block diagram of the algorithm for location optimization that maximizes the Bhattacharyya coefficient.

The optimization from above employs the mean shift vector in Step 3 to increase the value of the approximated Bhattacharyya coefficient $\tilde{\rho}(\mathbf{y})$, given by the right side of equation (7). Since this operation does not necessarily increase the value of $\hat{\rho}(\mathbf{y})$, the test included in Step 4 is needed to validate the new location of the face. However, practical experiments (tracking different faces, for long periods of time) showed that the Bhattacharyya coefficient computed at the location $\hat{\mathbf{y}}_1$ defined by (10) was almost always larger than the coefficient corresponding to $\hat{\mathbf{y}}_0$. The termination threshold ϵ used in Step 5 is derived by constraining the vectors representing $\hat{\mathbf{y}}_0$ and $\hat{\mathbf{y}}_1$ to have the same integer coordinates.

5 Tracking and Scale Adaptation

The face tracking process assumes for each frame the execution of the location optimization algorithm described in Figure 1. Thus, given the face model, the new location of the face in the current frame maximizes the value of Bhattacharyya coefficient in the neighborhood of the previous location estimate.

3

To handle scale changes of the target, the scale invariance property of (2) is exploited. Thus, in addition to the location optimization, a second search is performed, this time function of the bandwidth h. More specifically, the tracker looks for the bandwidth h that determines the face region in the current frame associated with the largest Bhattacharyya coefficient. This task can be implemented efficiently through a random search for scale variations. We simply modify the bandwidth h of the kernel profile with a random fraction (limited to $\pm50\%$), and let the mean shift based algorithm to converge again.

The decision on scale and location is also based on the luminance gradient [25, p.312] computed on the border of the hypothesized face region. As suggested in [4] the mean value of the gradient magnitude computed on the face border, in a direction perpendicular to the border is an important clue for face localization. However, since highlights in the image can induce unreasonably large values of the local gradient, we use a trimmed mean measure which calculates the mean only for points between the 10th and 90th percentiles [3].

Finally, out of the three sets of results (coming from three optimizations at different scales) the system chooses the bandwidth and location yielding the largest combined color and gradient measure, denoted by γ. The value of γ is between 0 and 1 and is obtained as the mean of the Bhattacharyya coefficient value and normalized gradient measure ϕ. An IIR filter is used to derive the new bandwidth based on the current measurements and old bandwidth.

6 Tracking Initialization

6.1 Face Model

The face model is the same for all experiments presented in this paper. It has been obtained from only one subject, by computing the mean histogram of his face instances recorded in the morning, afternoon and at night. The office where the experiments were performed has a large window on one side, therefore, both outdoor and indoor illumination condition were tested. The histograms were computed in the intensity normalized RG space [26, 27] with 128×128 bins.

6.2 Face Detection

The face detector employs the same ideas as the tracking procedure, however, it runs the location optimization with different initializations. For the current settings of the system (320×240 pixel images with subjects at a distance between 30cm to 3m from the cam-

Figure 2: *Three people* image and initialization ellipses used for face detection.

era) we use five initial regions of elliptical shape with semi-axes (normalization constants) $(h_x, h_y) = (37, 51)$, as shown in Figure 2. This arrangement guarantees that at least one initial ellipse is in the basin of attraction of a face of typical size.
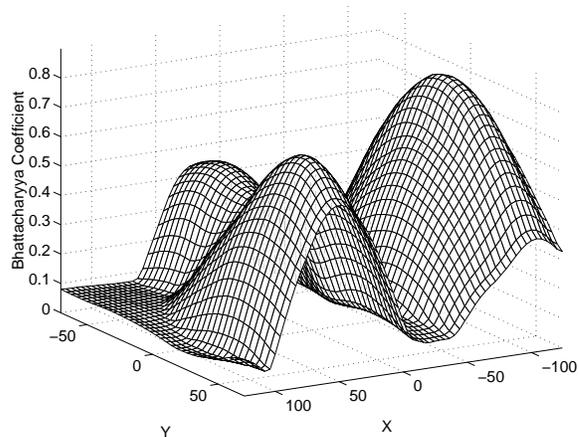

Figure 3: Values of the Bhattacharyya coefficient corresponding to the image shown in Figure 2. One can identify the three peaks of the surface, one for each face in the image.

Figure 3 presents the surface obtained by computing the Bhattacharyya coefficient for the entire image from Figure 2. One can easily identify the three peaks of the surface, one for each face in the image. The advantage of our method should be obvious from Figure 3. While most of the tracking approaches based on regions must perform an exhaustive search of the image (or a given neighborhood) to find the maximum value of the similarity measure, our algorithm exploits the gradient of the surface to climb to the closest peak. With proper multiple initializations, the highest peak is found very fast.

6.3 Detection and Failure Thresholds

The thresholds that control the target detection and tracking failure are dependent on the current environ-

---

[3] The behavior of the tracker in the presence of highlights or other large gradient sources improved considerably with the use of the trimmed mean.

ment the tracker operates. Although we plan a detailed analysis of the detection and failure events [14], at this moment the two thresholds are fully adjustable by the user according to the background and illumination conditions. Their default values were set experimentally. For all the results presented in the paper, detection was hypothesized whenever the combined measure γ was larger than $T_d = 0.2$ during the detection stage, while tracking failure was declared whenever γ was smaller than $T_f = 0.3$.

7   Camera Control

The adequate control of the pan, tilt, and zoom camera is an essential phase of the tracking process. The camera should execute fast saccades in response to sudden and large movements of the target while providing a smooth pursuit when the target is quasi-stationary [19, 20]. We implemented this type of control which resembles that of the human visual system. The fovea subimage occupies laterally about 6 degrees of the camera's 50 degrees field of view, at zero zoom.

However, contrary to other tracking systems that suspend the processing of visual information during the saccades movements [3], our visual face tracker is sufficiently robust to deal with the large amount of blurring resulting from camera motion. As a result, the visual tracking is a continuous process that is not interrupted by the servo commands. Note that a standard $RS - 232C$ interface is used to communicate with the $SonyEVI - D30$ camera.

8   User Interface

The user interface of the system (shown in Figure 4) is developed in Java and employs a standard JNI (Java Native Interface) to call fast native methods for video acquisition and camera control. The visualization of the tracked face can be selected as an upper-left window, superimposed ellipse, centered window with black background, or the entire frame can be shown, after both the subject and camera motions have been compensated.

9   Experiments

A first set of results are presented in Figure 5, demonstrating face tracking along a sequence of 1059 frames. The subject's face has been detected within a few frames after entering the field of view of the camera (frame 42). The detected face is shown in the small upper-left window. The camera is then tracking the face during walking (frames 69 and 147) and turnings (frames 99 and 177). The subject tries to escape the tracker by performing fast lateral movements (frame 267) or hand waiving (frame 309). Observe the blurring that accompanies the these movements, without affecting the tracker. Next, the subject tries to hide
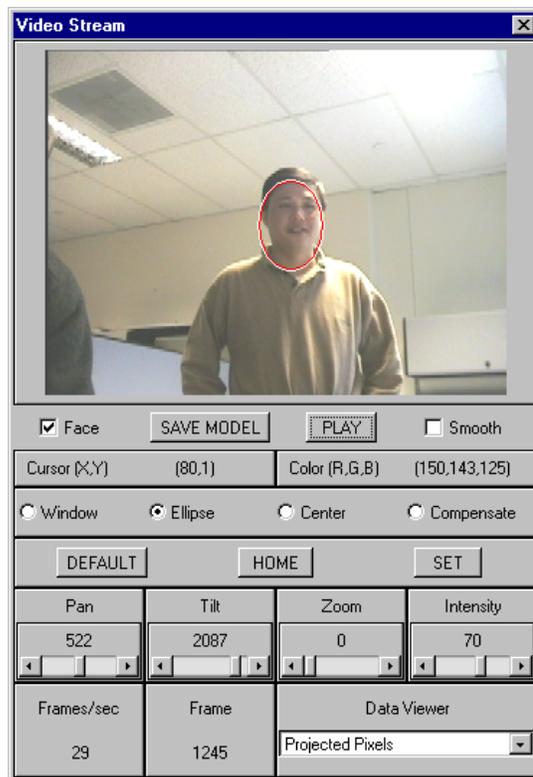

Figure 4:   User interface.

behind a chair (frames 552 and 582), but only when the head is completely occluded the tracker fails (frame 654). However, once the occlusion is terminated, the face is immediately recovered. Finally, one can see the scale adaptation working when the subject approaches the camera.

The histogram of the number of mean shift iterations necessary for each frame to perform all optimizations is given in Figure 6. The mean number of iterations, 4.35 is very low, showing the low computational complexity of the method. Figure 7 presents the evolution of the combined measure γ along the sequence from Figure 5. Observe the value drops corresponding to the turnings, fast motion, or occlusion of the target.

The separate plots for the Bhattacharyya coefficient and normalized gradient measure are given in Figure 8. Observe that the gradient measure carries much less information than the Bhattacharyya coefficient. This justifies our approach in using the gradient information only combined with color information, and only for scale control.

Figure 9 shows samples from another tracking sequence demonstrating the detection of the subject (frame 108) and tracking. The subject is walking (frame 174 and 282) and turning while sitting on a chair (frames 846, 1041, 1080). Note that the same

face model has been used, although the subject has a different skin color.

The sequence presented in Figure 10 shows face detection and tracking in a crowded environment. The frames 747 and 780 underline potential problems caused by the occlusion of the tracked face with another face. The tracker is distracted and starts pursuing the new face. Unfortunately, no safe solution exists for this event, excepting a more complex characterization of the tracked face and its permanent re-identification. This is however a computationally intensive process. Note that Kalman filter type prediction is almost useless in this scenario, due to the random walk of the subjects.

Finally, the sequence from Figure 11 shows the capability of the tracker to handle scale changes, subject turnings (frame 150), in-plane rotations of the head (frame 498), and foreground/background saturation due to back-light (frame 576).

10   Conclusion

We presented a real-time system for the detection and tracking of human faces with an active camera. The core component of the system is based on a new tracking framework that involves low computational cost. The paper demonstrated the robustness of the system to scale variations, fast subject movements and camera saccades, partial occlusion, out-of-plane rotations, and crowded scenes. Our current work focuses on techniques that prune the face candidates for cases when the background color is similar to the skin color.

References

[1] J.K. Aggarwal, Q. Cai, "Human Motion Analysis: A Review," *Computer Vision and Image Understanding*, 73:428–440, 1999.

[2] Y. Bar-Shalom, T. Fortmann, *Tracking and Data Association*, Academic Press, London, 1988.

[3] J. Batista, P. Peixoto, H. Araujo, "Real-Time Active Visual Surveillance by Integrating Peripheral Motion Detection with Foveated Tracking," *IEEE Workshop on Visual Surveillance*, Bombay, India, 18–25, 1998.

[4] S. Birchfield, "Elliptical Head Tracking using intensity Gradients and Color Histograms," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Santa Barbara, 232–237, 1998.

[5] G.R. Bradski, "Computer Vision Face Tracking as a Component of a Perceptual User Interface," *IEEE Work. on Applic. Comp. Vis.*, Princeton, 214–219, 1998.

[6] J. Cai, A. Goshtasby, "Detecting Human Faces in Color Images," *Image and Vis. Computing*, 18:63-75, 1999.

[7] R.T. Collins, A.J. Lipton, T. Kanade, "A System for Video Surveillance and Monitoring," *American Nuclear Society Eight Intern. Topical Meeting on Robotics and Remote Systems*, 1999.

[8] D. Comaniciu, P. Meer, "Mean Shift Analysis and Applications," *IEEE Int'l Conf. Comp. Vis.*, Kerkyra, Greece, 1197–1203, 1999.

[9] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift, To appear, *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Hilton Head Island, South Carolina, 2000.

[10] J.L. Crowley, F. Berard, "Multi-Modal Tracking of Faces for Video Communications," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Puerto Rico, 640–645, 1997.

[11] T. Darrell, G. Gordon, M. Harville, J. Woodfill, "Integrated Person Tracking using Stereo, Color, and Pattern Detection," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Santa Barbara, 601–609, 1998.

[12] P. Fieguth, D. Terzopoulos, "Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Puerto Rico, 21–27, 1997.

[13] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, 73:82–98, 1999.

[14] M. Greiffenhagen, V. Ramesh, D. Comaniciu, "Statistical Modeling and Performance Characterization of a Real-Time Dual Camera Surveillance System," To appear, *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Hilton Head Island, South Carolina, 2000.

[15] M. Isard, A. Blake, "Condensation - Conditional Density Propagation for Visual Tracking," *Intern. J. Comp. Vis.*, 29(1):5–28, 1998.

[16] D.W. Jacobs, P.N. Belhumeur, R. Basri, "Comparing Images Under Variable Illumination," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, Santa Barbara, 610-617, 1998.

[17] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," *IEEE Trans. Commun. Tech.*, COM-15:52–60, 1967.

[18] S.J. McKenna, Y. Raja, S. Gong, "Tracking Colour Objects using Adaptive Mixture Models," *Image and Vision Computing*, 17:223–229, 1999.

[19] D.W. Murray, K.J. Bradshaw, P.F. McLauchlan, I.D. Reid, P.M. Sharkley, "Driving Saccade to Pursuit using Image Motion," *Intern. J. Comp. Vis.*, 16(3):204–228, 1995.

[20] H.P. Rotstein, E. Rivlin, "Optimal Servoing for Active Foveated Vision," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, San Francisco, 177–182, 1996.

[21] H.A. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Anal. Machine Intell.*, 20(1):23–38, 1998.

[22] K. Sobottka, I. Pitas, "A Novel Method for Automatic Face Segmentation, Facial Feature Extraction, and Tracking," *Signal Process.: Image Communication*, 12(3):263–281, 1998.

[23] M.J. Swain, D.H. Ballard, "Color Indexing," *Intern. J. Comp. Vis.*, 7(1):11–32, 1991.

[24] K. Toyama, "Prolegomena for Robust Face Tracking," *Microsoft Research Technical Report*, MSR-TR-98-65, 1998.

[25] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, New Jersey, 1998.

[26] J. Yang, A. Waibel, "A Real-Time Face Tracker," *IEEE Work. on Applic. Comp. Vis.*, Sarasota, 142–147, 1996.

[27] G. Wyszechi, W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Second Ed. New York: Wiley, 1982.

42   # 69

99   # 147

177   # 267

309   # 552

582   # 654

684   # 1059

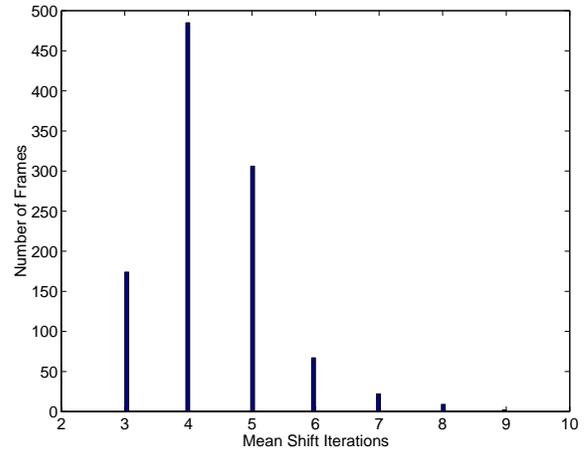Figure 5:   *Dorin* sequence.

Figure 6:   Histogram of mean shift iterations for *Dorin* sequence.   The mean number of iterations is 4.35 per frame.
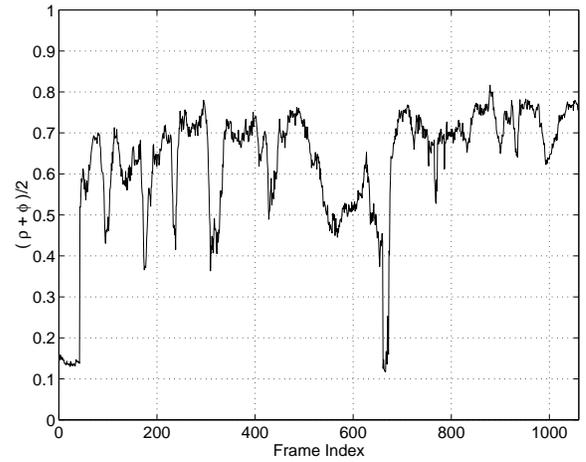

Figure 7:   Combined measure γ computed as the mean of Bhattacharyya coefficient ρ and normalized gradient measure ϕ for *Dorin* sequence.
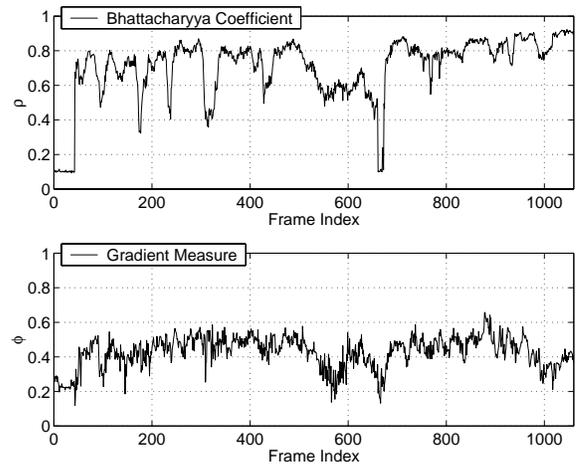

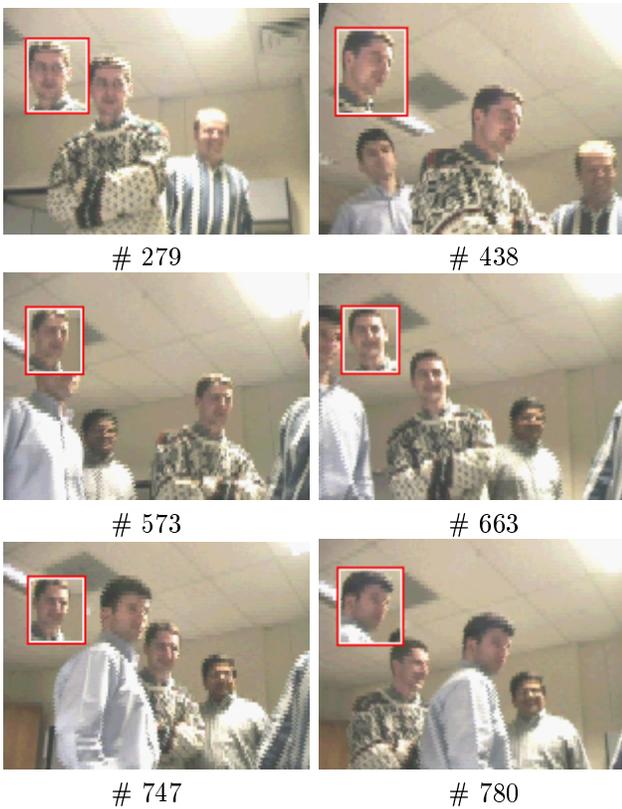Figure 8:   Bhattacharyya coefficient ρ and normalized gradient measure ϕ for *Dorin* sequence.

108   # 174

282   # 846

1041   # 1080

Figure 9:   *Ramesh* sequence.

279   # 438

573   # 663
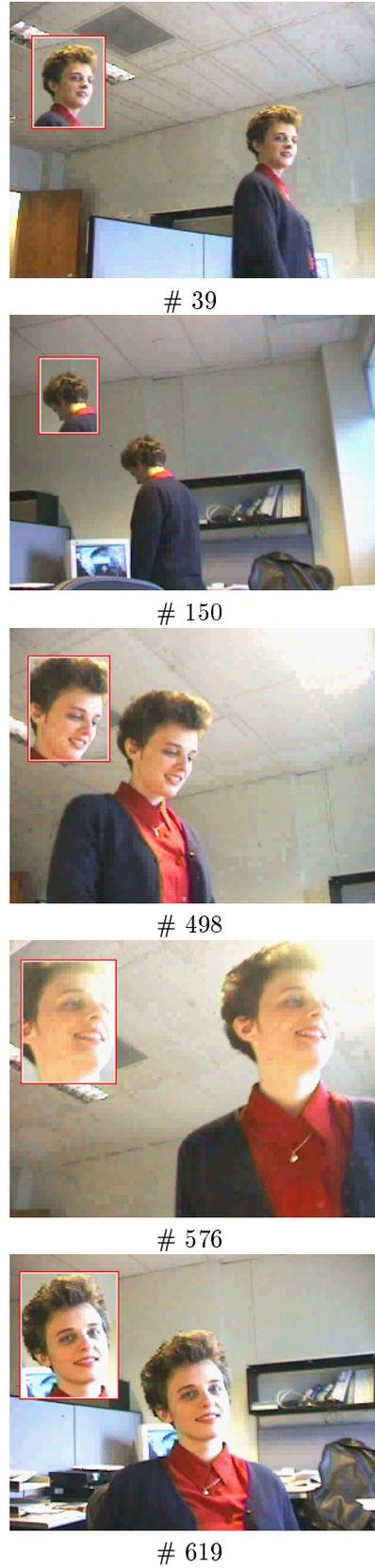
747   # 780

Figure 10:   *People* sequence.

39

150

498

576

619

Figure 11:   *Cristina* sequence.