Face Tracking System Based on Color, Stereovision and Elliptical Shape Features

Bogdan Kwolek
Rzeszów University of Technology, 35-959 Rzeszów, Poland
bkwolek@prz.rzeszow.pl

Abstract

In this paper we present a vision system that performs tracking a human face in 3D. To achieve this we combine color and stereo cues to find likely image regions where face may exist. A greedy search algorithm examines for a face candidate focusing the action around the position of the face which was detected in the previous time step. The aim of the search is to find the best-fit head ellipse. The size of the searched ellipse projected into image is scaled depending on the depth information. The final position of the ellipse is determined on the basis of intensity gradient near the edge of the ellipse, depth gradient along the head boundary and matching of the color histograms representing the interior of the actual and the previous ellipse. The color histogram and parameters of the ellipse are dynamically updated over time and compared with previous ones. The frontal view face is detected using PCA to make the tracking more reliable and in particularly to update the color model over time with only face-like skin pixels.

1. Introduction

Robust real-time tracking and segmentation of a moving face in image sequences is a fundamental step in many vision systems including automated visual surveillance, human-machine interface, and very low-bandwidth telecommunications. An advanced surveillance interface may use face tracking and detection techniques to direct the attention of the computer to a human being and maintain the face in the camera's field of view and in consequence reduce the communication bandwidth as well as a memory space by transmitting/storing only the fraction of a video frame which contains the area of interest (the tracked face). Continuously locating the position of the user's face may be necessary or at least helpful for a monitoring system to accomplish tasks such as recognizing the user's face [6]. Service robots are intelligent machines that provide services for human beings. They operate in dynamic and unstructured environment and interact with people using user-friendly interface in natural end efficient way. The mobile agents can aid the surveillance tasks and provide useful information about human activity [8].

This paper presents a simple, fast and robust vision tracking system that operates on an autonomous mobile robot and is capable to detect human head (face) on the basis of color combined with depth information, image segmentation, elliptical approximation of the head (eyes detection and eigenfaces, respectively). The outlined above cues are combined and independent units of the vision module which provide sufficient redundancy can generate results even when some features are temporary not available, e.g. the response of the color module is poor or the eyes can not be detected. The system is person independent and has been tested in complex environments and scenarios.

A lot of work has been performed in the area of human tracking, particularly for video based surveillance applications [3, 7, 8, 23]. In recent years many methods have been proposed to detect human faces in a single image [20] or sequence of images [4, 16, 25] using gray [18] or color [16, 20, 25] images. Often a single technique is applied to extract face features and to isolate them from the rest of the image [25]. The Intel CamShift algorithm [5] was designed to handle precise tracking of facial location on the basis of a nonmoving camera. It is considered as a very good tracking algorithm especially suited for perceptual interface. The CamShift can segment the tracked face within its search window when the hue information is unique and when the observed scene is under constant lighting. The authors of the work [17] demonstrated that color model adaptation is a solution to the problem of varying illumination in tracking. A Gaussian mixture model was employed to represent color distribution and a linear extrapolation was used to adjust the parameters of the model by a set of labeled training data drawn from the new image. However, since the new image is not segmented, this labeled data set is not reliable. In Birchfield's real-time head tracking system [4], the projection of a head in the image plane was modeled by an ellipse. The intensity gradient near the edge of the ellipse and a color histogram representing the interior were used to update the ellipse parameters over time. However, these histograms were static and did not accommodate to varying illumination. Darrell, at al. [7] combine stereo and color via an intensity pattern classification method to track people. The CMU face detector [18] library has been used to discriminate the frontal face from other body parts. A static skin model which was constructed on the basis of the hue color component has been applied to detect patches of skin colored pixels through an image sequence. Another approach to localize the human head on the basis of combination of color segmentation and depth from stereo has been proposed in [14]. The detection of head has been achieved using perspective projection of the parameterized head model and a correlation based search for the head model in the depth map. During detection no interframe correspondences between the region size as well as between the position have been taken into account. In work [21] a real-time system for tracking multiple people has been discussed. The face recognition module is based on Visionics' FaceIt developer kit and multiple static and Pan-Tilt-Zoom cameras. Gaussian estimation is used to generate a look-up table of the prepared in advance color model. The system requires that people in the viewing area are not wearing a shirt or blouse of the same color.

2. System overview

In method proposed here, information about skin colors is used first to find likely image regions where a face may exist. The outcome is a graylevel image, with pixels representing the probability of membership to the skin class. In the second step, pixels belonging to extracted areas are thresholded and further processed by a connected component labeling procedure. However, if an input image contains a background whose color is similar to the color of the true skin region, the labeled segments in the image can be of any arbitrary shape and not necessarily elliptical and therefore the information about the hypothesized face location is not reliable at this stage.

The gravity centers of the detected face candidates are then utilized as seed regions of the region growing procedure which operates on corresponding depth image in order to divide the scene into sections according to their distance to the camera. The output of this stage is an image with extracted depth layers.

Next, taking into account the depth layers and corresponding to them the distances to the camera we can determine for each face candidate the size of the best-fit ellipse approximating the oval shape of human face. Thanks to the being in disposal the depth layers we can connect the neighboring labels in the case of splitting one face into several labels and thus determine the centers of the ellipses more precisely. In the case of over-labeling, the depth information is useful in extraction the pixels from the labeled area which are parts of a face and discarding background ones.

The final position of the ellipse is determined on the basis of the intensity gradient near the edge of the ellipse, the depth gradient along the head boundary, the matching score of color histograms representing the interior of the previous and actual ellipse and the hypothesize-and-test paradigm [4]. Using only face-like pixels from the most likely face area, a new face histogram is computed and the previous model is averaged with it.

Face candidates are further verified in respect of reflectional symmetry and of eyes presence. Afterwards, each of the face candidates is normalized into 16x16 image based on the eye positions and according to the distance between the candidate and the camera, then converted into a point in the 16-dimensional eigenface space. Face identification is evaluated on the basis of the Euclidean distance between two corresponding points in the eigenface space.

The skin model is updated only from ellipse's interior pixels which satisfy the determined in advance chromatic constraint criteria. To cope with situations where the subject turns around and skin feature is not available, the system relies on intensity gradients and the depth information to set the position and size of the ellipse. When the detection of the oval shape on the basis of intensity gradients is impossible the system tries to detect an ellipse representing the tracked head on the basis of stereovision.

3. Image segmentation

In our approach we first locate skin-like regions by performing color image segmentation. Skin color is used for detecting faces in images and particularly in image sequences because the color as a cue is computationally inexpensive, robust towards changes in orientation and scaling of an object. The efficiency of color segmentation techniques is especially worth to emphasize if a considered object is occluded or is in shadow what can be in general significant practical difficulty using edge-based methods. One way to increase tolerance toward intensity changes is to transform a RGB image into a color space whose chromaticity and intensity are separate and use only chrominance for detection. The chrominance is the measure of the lack of whiteness in a color [2]. Projection to normalized color space rg where $r = R/(R+G+B)$, $g = G/(R+G+B)$ reduces thus brightness dependence.

Yang at al. [25] have shown that human skin colors form a relatively compact cluster in a small region in the normalized color space. Moreover, a single Gaussian with full covariance is very often sufficient for modeling the skin color distribution [17, 19] thus avoiding the need for iterative Expectation Maximization EM algorithm which offers a way to fit probabilistic models to the observation data. Some research results show additionally that (1) skin color clusters have more difference in intensity than in color [25]; (2) with the change of lighting conditions the major translation of skin color distribution is along the lightness axis of the RGB color space; (3) skin colors acquired from a static face tend to form tight clusters in the color space while moving ones form widen clusters due to different reflecting surfaces during the movement. Therefore, the normalized color space

rg is considered to be capable of characterizing human faces being in a movement.

For the generation of a skin color model we manually segmented a set of images containing skin regions in typical illumination conditions which appear in our laboratory. It has been observed that the set of possible skin chromaticities in normalized color coordinates forms a shell-shaped structure. Such a structure has been used to determine chromatic constraint criteria of the apparent skin color in typical illumination conditions which appear in our laboratory. The average histograms as non-parametric skin color density models in typical illuminations have been then extracted and stored for the future use. Prior to run-time initialization phase a one of the prepared in advance histograms is chosen according to the illumination conditions. The Gaussian is fitted to the histogram and then is utilized to generate a probability image. The frame to be segmented is transformed into the color space and each pixel with chromaticity $[r_i, g_i]^T$ is assigned the value of the Gaussian and compared further to a likelihood threshold. Only those pixels whose likelihood is above the threshold are classified as belonging to a potential face. Because our goal is a coarse detection of face skin-like region, the chose of the threshold have not a considerable influence on the obtained results and therefore a small threshold is applied in our system.

The extraction process of skin-tone is analogous to the single hypothesis classifier described in [9]. Single hypothesis classifier deals with problems in which one class is well defined while others are not. Let $\xi = [r_i, g_i]^T$ denote the feature vector that consists of color components of a pixel and ω_s denote the skin class. Thus the probability that pixel belongs to class ω_s can be expressed as

$$p(\xi|\omega_s) = \frac{1}{2\pi\sqrt{det(\Sigma_s)}} e^{-\frac{1}{2}(\xi-\mu_s)^T \Sigma_s^{-1} (\xi-\mu_s)} \qquad (1)$$

where μ_s is the mean color vector and Σ_s is the covariance matrix of considered class. The better the pixel matches the color model, the higher the probability and response of such a color filter are. The parametric model which we use to extract skin-like pixels provides a generalization about small amounts of training data and tends to smoothen a training set distribution. After an image is thresholded, morphological closing operator is performed to fill holes in detected regions. The binary image extracted in such a way is further processed by connected component labeling procedure, resulting in labeled blobs of flesh tone. The labeled regions are used to calculate areas and gravity centers of the detected candidates of face.

Color segmentation techniques are very robust, but they may typically produce large false positives, and thus, need to be combined with other techniques in order to give a solid basis for reliable face detection. The outlined above color segmentation module provides the stereo module with an initial region of interest. Stereovision gives in particular information about the real distance between object and the camera. The attendance of shadows at subject, lighting changes has little effects on obtained results. Moreover, assuming the constancy of position from frame to frame, partially occluded objects can be detected in sequence of images, which can be a meaning problem for other techniques. However, the distance between the camera and the face to be isolated in the depth image is limited to certain range due to stereovision geometry. If there are not enough textures, some parts of the scene are not included in the depth image.

A commercially available stereo system from SRI has been chosen to compute dense depth images. This stereo system produces 320x240 range images at 15 fps on Pentium III. The stereo depth information is extracted thanks to small area correspondences between image pairs [11] and therefore it gives poor results in regions of little texture. But the depth map covering a face region is usually dense because the human face is rich in details and therefore the applied stereo system as separate source of information aids the process of face detection.

The centers of gravity of detected face candidates are used to determine seed values in a region-growing procedure which operates on depth images. Our aim is to extract the tracked face and therefore the labeled pictures of skin candidates allow us to extract the person distance layer via the region growing in the depth image. The person distance layer which is obtained in this manner is then used in further processing of labeled regions to give initial estimates of the centers of ellipses as approximations of the oval shape of the face. Particularly, if a candidate region has some darken regions it can be represented through several separate blobs via the labeling operation. Thanks to the being in disposal the depth layers we can connect the labels which are in proximity on each other and simultaneously share similar distance to the camera. In the case of over-labeling the depth information is used to discard from the labeled blobs the background pixels. Extracted in such a way face candidates are further verified using elliptical shape features and color histogram matching.

4. Searching for the face

To distinguish between the hands and the face, a face identification method is utilized, employing elliptical shape features and color histograms. The aim of the searching for a face is to find the location of the projection of the head in the image plane whose intensity gradient near the object's boundary, depth gradient near the object's boundary and a color histogram representing the object's interior best match the values of the model. The search for the best-fit-ellipse is accomplished via a hypothesize-and-test procedure [4]. The head is modeled as a vertical ellipse with an assumed fixed aspect ratio equal to 1.2. The position (x, y) and size d (length of the minor axis) of the head in an image are tracked from frame to frame based on velocity and size constancy. A greedy search algorithm for correspondences reviews the image with aim to locate a closest unmatched region focusing the action around the position of the face which was detected in the previous time step. The state of the ellipse is then maintained by performing the local search to maximize the goodness of the following match:

$$s^* = \arg\max_{s_i \in S}\{\overline{\phi}_g(s_i) + \overline{\phi}_c(s_i) + \overline{\phi}_d(s_i)\} \tag{2}$$

where $\overline{\phi}_g$, $\overline{\phi}_c$ and $\overline{\phi}_d$ are the normalized matching scores based on intensity gradients, color histograms and depth gradients, respectively. The search space S is the set of all states within some range of predicted state if the gravity center of examined candidate of face is close to predicted position. If not, the size of the ellipse projected on the image is scaled depending on the depth information. In order to avoid the searching for a face with a false face candidate the feature is applied in equation (2) only when its match score is above an empirically determined threshold.

The gradient magnitude has been obtained on the basis of the Sobel mask [10]. The filtering with Gaussian mask preceded the gradient extraction. Color histogram is one of the most important techniques for pattern matching because of its efficiency and effectiveness. However, due to the statistical nature, color histogram can only reflect the content of images in a limited way. We have implemented a histogram intersection technique [22] thanks to which only the similar face candidates in an image sequence are used to estimate the face pose. For a given pair of histograms I and M, each containing n values, the intersection of the histograms is defined as follows: $H(I, M) = \sum_{i=1}^{N} min(I_i, M_i)$. The terms I_i, M_i represent the number of pixels inside the i-th bucket of the candidate and the model histogram, respectively, whereas N the total number of buckets. The result of the intersection of two histograms is the number of pixels that have the same color in both histograms. To obtain a match value between zero and one the intersection is normalized and the match value is determined as follows: $H_\cap(I, M) = H/\sum_{i=1}^{N} I_i$.

In the searching for an ellipse with maximal intersection score there is an overlapping between adjacent states of the ellipses resulting in common pixels. Such a redundancy is utilized in fast matching since for each new considered ellipse, its color histogram can be extracted via subtracting common pixels and adding the new ones [4].

5. Real-time face detection

The goal of face detection is to determine whether or not there is any human face in the image, and, if present return its location. Face detection is a difficult task because certain common but significant features, such as glasses or a moustache, can either be present or absent on a considered face candidate. Because of three-dimensional facial structures a change in lighting conditions, a movement of a face or a camera can cast or remove significant shadows from a tracked face. Face detection is a time-consuming operation due to the lack of constraint on the size and the location of tracked face in a sequence of images.

The digital images are spatially redundant. The Karhunen-Loeve transformation which is also know under the name Hotelling transform or principal component analysis being based on statistical properties ensures the extraction of data with decreasing statistical significance (smaller and smaller eigenvalues). In the methods that are based on eigenfaces the feature space is reduced using principal components [24]. The scaled query windows of the input image are projected into the classification subspace and the closer the distance to the projected training image is, the greater the probability that such window corresponds to the trained human face is. Therefore, in order to apply the eigenfaces approach we must specify a window of proper size which will be examined in terms of the face presence. Although this algorithm is suitable for detecting the desired feature, its computational cost is considerable. Therefore we utilize the results from the previous stage to reduce the number of points that have to be checked in the input image.

Before being processed, a face in an input image should first be located and registered in a frame which size is equal to the size used in the training collection. Our algorithm of frontal view face detection requires that the window pattern to be classified be 16x16 pixels in size and appropriately masked. All windows patterns of different dimensions must first be re-scaled to this size and masked. The starting point for the face detection is the position of the ellipse which was obtained as described in previous section. Thanks to the being in disposal information about the distance of the examined candidate of face to the camera, the reflected symmetry is checked in an ellipse covering only the considered face. If the symmetry is detected a simple binary matching technique [10] suitable for real-time computation is used to effectively detect and locate eyes in the frontal view of a human face. One of the prepared off-line templates is chosen according to the distance between the examined face candidate and the camera. It has been proved that human eyes distinguish themselves from the whole rest of a face even if a user wears glasses similarly to the author. Once the eyes are detected we resize the potential facial region by using bicubic interpolation technique. The proper determination of the window localization has a considerable influence on efficiency of detection. If we were to attempt face detection on window patterns which varied in lighting, alignment, scale and rotation, the dimensionality of the task would grow significantly. In consequence PCA could fail at classifying because the problem which is now a low dimensional one has become high dimensional as an effect of variations in lighting, alignment, scale and rotation.

The eigenfaces algorithm operates on gray images and a collection of training faces [24]. We have prepared a collection of training faces that consists of 64 images. While preprocessing, the average image for this collection is computed. Next, this image is subtracted from each training image and placed in the matrix. This matrix is used to create the covariance matrix. The eigenvalues and eigenvectors of such a covariance matrix are then determined. The first 16 normalized eigenvectors, sorted by decreasing eigenvalue represent subspace in which classification at run-time phase is performed. The eigenfaces are normalized eigenvectors which are the principal components of the face space and they reflect the statistical properties of facial appearance. To improve the performance the eigenfaces method in the presence of lighting variation we remove the largest two principal components.

6. Experiments

A mobile robot Pioneer 2DX [15] that was designed to move across a relatively flat surface was used in experiments and tests of prepared software. The robot was equipped with SRI's Megapixel Stereo Head. A typical laptop computer equipped with 2 GHz Pentium IV is utilized to run the prepared software. The position estimate of the tracked face as well as person distance to the camera are written asynchronously in block of common memory which can be easily accessed by Saphira client. Saphira is an integrated sensing and control system architecture designed to operate with a robot server [12]. The robot server sends a message packet to the client every 100 milliseconds, containing information on the velocity of the vehicle, sensor readings, and etc. If the person is located, the vision system keeps his/her face within the camera field of view by coordinating the rotation of the robot with the tracked centroid location. The aim of the robot orientation controller is keep the position of the tracked person at specific position in the image. The linear velocity has been dependent on person's distance to the camera. A distance 1.7 m has been assumed as the reference value that the linear velocity controller should keep during experiments consisting in person following. To eliminate unnecessary rotations as well as forward and backward robot movements we have applied a simple logic. The PD controllers have been implemented in the Colbert language [12] .

The image processing and recognizing algorithm runs at 320x240 image resolution at frame rates of 8-10 Hz depending of image complexity. It was implemented using C/C++ language. We have realized experiments in which the robot has followed a person at distances which beyond 100 m without the person loss. At the beginning of experiments the robot motors are switched off and the visual system tries to locate the person to be tracked. After the person's layer extraction, the detection of the face presence, the motors are switched on and system is turned into tracking mode. When in sequence of several frames the tracked face is not found the motors are off. In such a situation the robot tries to localize skin blobs or dominant areas of movement. To localize areas with dominant movements the robot is still for a period of time needed to acquire three consecutive frames and then the optical-flow field approach [1] is applied.

The performance of tracking that is based on a moving camera usually depends on the environment, the behavior of the mobile robot and the lighting conditions. To coarsely estimate the effectiveness of our approach we included into our system the CamShift algorithm as a separate tracking module. We compared the algorithms in experiments consisting in tracking the face of the person moving around the mobile robot (the linear velocity of the robot was set to zero). The CamShift did not re-initialize and respond well to changes in lighting, or to rapid repositioning of the person mainly due to delays introduced by robot. The performance of the CamShift was also poor when small face images have been processed. One of the examples where the CamShift failed and our method performed successfully is shown in fig. 1. Experiments showed that the person layer is useful in several situations and can be used to re-localize the tracked person when color module fails.

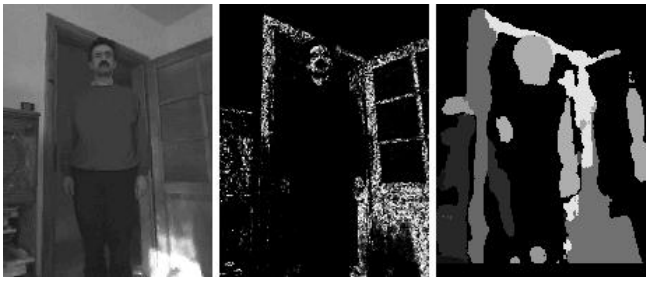

Figure 1. Input image, probability image, depth layers

7. Conclusions

We have presented a system that robustly tracks and detects a human face. The color segmentation plays an important part in our system because there is an outlook that in a short period of time will be accessible cameras that will be oriented to automatic skin-tone detection. Stereovision has proved to be very useful and robust cue in experiments consisting in face tracking and detecting during person following with mobile robot. The combination of skin-tone color segmentation and stereovision seems to have a considerable perspective of applications in robotics and surveillance. A human-machine interaction consisting in signaling commands by a person facing a mobile agent and observing the actions performed in response is natural especially at a learning stage. One of the most significant and natural arm postures is pointing towards an object of interest. In our system the direction of pointing is determined on the basis of 3D positions of the face (eyes) and the hand signaling the command. Thanks to the face position it is possible to recognize some static commands on the basis of geometrical relations of the face and hands [13]. The detection of a presence of the vertical and frontal-view faces in the scene is fast and reliable because the depth map covering the face region is usually dense and this together with color cues allows us to utilize symmetry information as well as the eyes-template before the usage of computationally expensive the eigenfaces method. The elaborated method of face detection in sequence of images is fast and reliable. The color adaptation over time has been proven to have significant influence on obtained results, particularly during the following a person with the mobile robot. Therefore the adaptation of flesh tone pixels takes place by utilizing only pixels from the face area. The knowledge of possible skin pixels that were obtained by the robot camera in different illumination conditions is used to select only face-like skin pixels for updating the color model. The system automatically initializes without user intervention, and can re-initialize when the tracking is lost.

The presented system runs at 320x240 image resolution at frame rates of 8-10 Hz on 2 GHz Pentium IV laptop which has been installed on Pioneer 2 DX mobile robot. The vision system enables the robot to follow a person with velocity not exceeding 30 cm per second. To show the correct work of the system, we have conducted several experiments in naturally occurring in laboratory circumstances. One of the limitations of the current implementation is too large processing time.

8. Acknowledgement

This work has been supported by the Polish State Committee for Scientific Research (project 4T11C01224). The author would like to thank the anonymous reviewers for their helpful comments.

References

[1] P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman. Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Trans. on Robotics and Automation*, 9(2):152–165, 1993.

[2] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, Inc., 1982.

[3] D. Beymer and K. Konolige. Real-time tracking of multiple people using continuous detection. In *IEEE Int. Conf. on Computer Vision*, Corfu, Greece, 1999.

[4] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *IEEE Conf. on Computer Vision and Patt. Rec.*, pages 232–237, Santa Barbara, CA, 1998.

[5] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, 1998.

[6] R. Chellappa, S. Zhou, and B. Li. Bayesian methods for face recognition from video. In *Int. Conf. on Acoustics Speech and Signal Processing*, Orlando, Florida, 2002.

[7] T. Darrell, G. Gordon, M. Harville, and J. Woodfill. Integrated person tracking using stereo, color, and pattern detection. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, pages 601–609, Santa Barbara, 1998.

[8] L. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Hariatoglu, and M. Black. Visual surveillance of human activity. In *Asian Conf. on Computer Vision*, pages 267–274, 1998.

[9] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Acad. Press, 1990. sec. ed.

[10] R. C. Gonzalez. *Digital Image Processing*. Addison Wesley Publ., London, 1977.

[11] K. Konolige. Small vision system: Hardware and implementation. In *Proc. of Int. Symposium on Robotics Research*, pages 111–116, Hayama, Japan, 1997.

[12] D. Kortenkamp, R. P. Bonasso, and R. Murphy. *Artificial Intelligence and Mobile Robots-Case Studies of Successful Robot Systems*. The MIT Press, Massachusetts, 1998.

[13] B. Kwolek. Visual system for tracking and interpreting selected human actions. *Journal of WSCG*, 11(2):274–281, 2003.

[14] F. Moreno, J. Andrade-Cetto, and A. Sanfeliu. Localization of human faces fusing color segmentation and depth from stereo. In *8th IEEE Int. Conf. on Emerging Technologies and Factory Automation*, pages 527–535, 2002.

[15] Pioneer 2. *ActivMedia Robotics*, 2001.

[16] E. Polat, M. Yeasin, and R. Sharma. Tracking body parts of multiple people. In *IEEE Workshop on Multi-Object Tracking*, pages 35–42, Vancouver, 2001.

[17] Y. Raja, S. J. McKenna, and S. Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, 17(3-4):225–232, 1999.

[18] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 203–207. IEEE Comp. Society Press, 1996.

[19] A. Sirohey and A. Rosenfeld. Eye detection in face image using linear and nonlinear filters. *Pattern Recognition*, 34:1367–1391, 2001.

[20] K. Sobottka and I. Pitas. Face localization and facial feature extraction based on shape and color information. In *Proc. of IEEE Int. Conf. on Image Processing*, volume III, pages 483–486, Lausanne, 1996.

[21] S. Stillman, R. Tanawongsuwan, and I. Essa. A system for tracking and recognizing multiple people with multiple cameras. In *Proc. of Int. Conf. on Audio-and Video-Based Person Authentication*, pages 96–101, Washington, 1999.

[22] M. J. Swain and D. H. Ballard. Color indexing. *Int. Journal of Computer Vision*, 7(1):11–32, 1991.

[23] R. Tanawongsuwan, A. Stoytchew, and I. Essa. Robust tracking of people by a mobile robotic agent. Technical report, Georgia Tech University, Feb 1999.

[24] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. of Conf. on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[25] J. Yang and A. Waibel. A real-time face tracker. In *Proc. of 3rd IEEE Workshop on Applications of Computer Vision*, pages 142–147, Sarasota, Florida, 1996.