Real-Time Head Tracker Using Color, Stereovision and Ellipse Fitting in a Particle Filter *

Bogdan KWOLEK
Rzeszów University of Technology
W. Pola 2, 35-959 Rzeszów, Poland
e-mail: bkwolek@prz.rzeszow.pl

Received: September 2003

Abstract

This paper proposes the use of a particle filter combined with color, depth information and shape features as an efficient and effective way to deal with tracking a head on the basis of image stream coming from a mobile stereovision camera. The head is modeled in the 2D image domain by an ellipse. The color distribution within interior of the ellipse is represented by a color histogram. The color histogram is dynamically updated over time. The length of the ellipse's minor axis is determined on the basis of depth information. The particles representing the candidate ellipses are weighted in each time step in respect of intensity gradient near the edge of the ellipse and matching score of the color histograms representing the interior of an ellipse surrounding the tracked object and currently analyzed one. The proposed algorithm can track a head reliably in case of temporal occlusions as well as varying illumination conditions by dealing with multiple hypotheses for the pose. Experimental results obtained on long image sequences show the feasibility of our approach to perform tracking a head undergoing complex changes of shape and appearance against a varying background. The tracker has been evaluated in experiments consisting in face tracking with a real mobile agent.

Key words: face tracking, color distribution, mobile camera, action recognition.

Introduction

Visual tracking is an important problem in tasks consisting in human action recognition, robot teleoperation as well as human computer interaction and is relevant in surveillance and vision based interfaces. One of the purposes of visual tracking is to

*This work has been supported by the Polish Committee for Scientific Research (KBN) within the project 4T11C01224

estimate the states of objects of interest from an image sequence. However, cluttered backgrounds, unknown and changing lighting conditions and multiple moving objects make the vision based tracking tasks challenging. Many different trackers for various tasks have been developed in recent years and particular interests and research activities have increased significantly in vision based methods. Some vision based systems allow determination of a body position and real-time tracking head and hands. Pfinder (Wren *et al.*, 1997) uses a multi-class statistical model of color and shape to obtain a blob representation of the tracked silhouette in a wide spectrum of viewing conditions. A system developed recently by (McKenna *et al.*, 2000) performs tracking at three levels of abstraction, namely regions, people and groups. A significant attention is paid to color information, particularly to cope with shadows and disambiguating occlusions during tracking pedestrians (walking people). In the techniques known as CamShift (Bradski, 1998) and MeanShift (Comaniciu *et al.*, 2000) the current frame is searched for a region in a variable-size window, whose color content matches best a reference model. The searching process proceeds iteratively starting from the final location in the previous frame. However, the tracking process would break down very fast when a moving window, surrounding the object, contains background colors which are similar to those from a model. Isard and Blake introduced particle filtering for visual tracking in an algorithm which is known as CONDENSATION (Isard and Blake, 1996). The approach presented in work (Dellaert *et al.*, 1999) uses the CONDENSATION algorithm for mobile robot pose estimation. Particle filtering is now a popular solution to problems relying on visual tracking.

The objective of this research is to detect and track the head of a person to perform person following with a mobile agent which is equipped with an on-board camera. The initial position of the head to be tracked is determined by means of face detection. We consider scenarios where a stereo camera is mounted on a mobile agent and our aim is to track an object undergoing complex changes of shape and appearance. There are many difficulties in solving such tasks and challenge lies in the fact that a background may not be static. Moreover, the appearance of the object of interest changes continuously due to non-rigid human motion and change in viewpoints. We consider the problem of head tracking by taking advantage of color together with shape as well as depth information which are combined with the particle filter. The aim of the random search in the particle filter is to find the best-fit head ellipse. The particles representing the candidate ellipses are verified in respect of intensity gradient near the edge of the ellipse and matching score of the color histograms representing the interior of an ellipse surrounding the tracked object and currently analyzed one. During samples weighting in which candidate ellipses are considered one after the other, the projected ellipse size into image is dependent on the depth information. The color histogram and parameters of the ellipse are dynamically updated over time to consider inter-frame changes of head appearance and to discriminate in the next iteration between updated representation of the tracked head and representation of individual candidates at sample positions.

The contribution of our work lies in the employment of particle filters combined with mentioned above cues to robustly solve a difficult and a useful problem of head tracking in color images. The tracker has been evaluated in experiments consisting in face tracking with a stereovision camera mounted on a real mobile agent.

The outline of this paper is as follows. In the next section we briefly describe particle filtering. The usage of color cue, shape information and stereovision in a particle filter is explained in section 3. In section 4 we report results which were obtained in experiments. Finally, some conclusions follow in the last section.

Particle filtering

In this section we formulate the visual tracking problem in a probabilistic framework. Among the tracking methods, the ones based on particle filters have attracted much attention recently and have proved as robust solutions to reduce the computational cost by searching only those regions of the image where the object is predicted to be. The key idea underlying all particle filters is to approximate the probability distribution by a weighted sample collection.

The state of the tracked object at time t is denoted $\mathbf{x}_t$ and its history is $X_t = \{\mathbf{x}_1, ..., \mathbf{x}_t\}$. Similarly the set of image features at time t is $\mathbf{z}_t$ with history $Z_t = \{\mathbf{z}_1, ..., \mathbf{z}_t\}$. The evolution of the state forms a temporal Markov chain so that the new state is conditioned directly on the immediately preceding state and independent of the earlier state, $p(\mathbf{x}_t \mid X_{t-1}) = p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$. Observations $\mathbf{z}_t$ are assumed to be independent, both mutually and with respect to the dynamical process, $p(Z_{t-1}, \mathbf{x}_t \mid X_{t-1}) = p(\mathbf{x}_t \mid X_{t-1}) \prod_{i=1}^{i=t-1} p(\mathbf{z}_i \mid \mathbf{x}_i)$. The observation process is defined by the conditional density $p(\mathbf{z}_t \mid \mathbf{x}_t)$. Given a continuous-valued Markov chain with independent observations, the conditional state density $p(\mathbf{x}_t \mid Z_t)$ represents all information about the state at time t that is deducible from the entire data-stream up to that time.

We can use Bayes' rule to determine the *a posteriori* density $p(\mathbf{x}_t \mid Z_t) = p(\mathbf{x}_t \mid \mathbf{z}_t, Z_{t-1})$ from the *a priori* density $p(\mathbf{x}_t \mid Z_{t-1})$ in the following manner

$$\begin{aligned} p(\mathbf{x}_t \mid Z_t) &= \frac{p(\mathbf{z}_t \mid \mathbf{x}_t, Z_{t-1}) p(\mathbf{x}_t \mid Z_{t-1})}{p(\mathbf{z}_t \mid Z_{t-1})} \\ &= k_t p(\mathbf{z}_t \mid \mathbf{x}_t, Z_{t-1}) p(\mathbf{x}_t \mid Z_{t-1}) \\ &= k_t p(\mathbf{z}_t \mid \mathbf{x}_t) p(\mathbf{x}_t \mid Z_{t-1}) \end{aligned}$$

where k_t is a normalization factor that is independent of $\mathbf{x}$ and

$$p(\mathbf{x}_t \mid Z_{t-1}) = \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid Z_{t-1}) dx_{t-1}$$

This equation is used to propagate the probability distribution via the transition density $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$. The density function $p(\mathbf{x}_t \mid Z_{t-1})$ depends on the immediately preceding distribution $p(\mathbf{x}_{t-1} \mid Z_{t-1})$, but not on any function prior to $t-1$, so it describes a Markov process. Multiplication by the observation density $p(\mathbf{z}_t \mid \mathbf{x}_t)$ in the equation for *a priori* density $p(\mathbf{x}_t \mid Z_{t-1})$ applies the reactive effect expected from observations. The observation density $p(\mathbf{z}_t \mid \mathbf{x}_t)$ defines the likelihood that a state $\mathbf{x}_t$ causes the measurement $\mathbf{z}_t$. The complete tracking scheme, known as the recursive Bayesian filter first calculates the *a priori* density $p(\mathbf{x}_t \mid Z_{t-1})$ using the system model and then evaluates *a posteriori* density $p(\mathbf{x}_t \mid Z_t)$ given the new measurement, $p(\mathbf{x}_{t-1} \mid Z_{t-1}) \xrightarrow{dynamics} p(\mathbf{x}_t \mid Z_{t-1}) \xrightarrow{measurement} p(\mathbf{x}_t \mid Z_t)$.

The density $p(\mathbf{x}_t \mid Z_t)$ can be very complicated in form and can have multiple peaks. The need to track more than one of these peaks results from the fact that the largest peak for any given frame may not always correspond to the right peak. The random search which is known as particle filtering has proven useful in such considerable algorithmic difficulties and allows us to extract one or another expectation. One of the attractions of sampled representations of probability distributions is that some calculations can be easily realized.

Taking a sample representation of $p(\mathbf{x}_t \mid Z_t)$, we have at each step t a set $S_t = \{(\mathbf{s}_t^{(n)}, \pi_t^{(n)}) \mid n = 1...N\}$ of N possibly distinct samples, each with associated weight. The sample weight represents the likelihood of a particular sample being the true location of the target and is calculated by determining on the basis of depth information the ellipse's minor axis and then by computing the gradient along ellipse's boundary as well as matching score of histograms representing the interior of ellipses which bound (i) the tracked object and (ii) currently considered one. Such a sample set composes a discrete approximation of a probability distribution. The prediction step of Bayesian filtering is realized by drawing with replacement N samples from the set computed in the previous iteration, using the weights $\pi_{t-1}^{(n)}$ as the probability of drawing a sample, and by propagating their state forward in time according to the prediction model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$. This corresponds to sampling from the transition density. The new set would predominantly consist of samples that appeared in previous iteration with large weights. In the correction step, a measurement density $p(\mathbf{z}_t \mid \mathbf{x}_t)$ is used to weight the samples obtained in the prediction step, $\pi_t^{(n)} = p(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}^{(n)})$. The complete scheme of the sampling procedure outlined above can be summarized in the following pseudo-code:

$S_t = \emptyset$
for $n = 0$ to N do
  select k with probability $\pi_{t-1}^{(n)} / \sum_{i=1}^{N} \pi_{t-1}^{(i)}$
  propagate $\mathbf{s}_t^{(n)} = A\mathbf{s}_{t-1}^{(k)} + w$
  calculate non-normalized weight $\pi_t^{(n)} = p(\mathbf{z}_t \mid \mathbf{s}_t^{(n)})$
  add $\mathbf{s}_t^{(n)}$ to S_t
endfor

The component A in the propagation model is deterministic and w is a multivariate Gaussian random variable. As the number of samples increases, the precision with which the samples approximate the pdf increases. The mean state can be estimated at each time step as $E[S_t] = \sum_{n=1}^{N} \mathbf{s}_t^{(n)} \pi_t^{(n)}$, where $\pi_t^{(n)}$ are normalized to sum to 1.

Representation of the Target Appearance

The shape of the head is one of the most easily recognizable human parts and can be reasonably well approximated by an ellipse. In work (Birchfield, 1998) a vertically oriented ellipse has been used to model the projection of a head in the image plane. The intensity gradient near the edge of the ellipse and a color histogram representing the interior were used to handle the parameters of the ellipse over time. Additionally, this method assumes that all pixels in the search area are equally important. The discussed tracking method does not work when the object being tracked temporarily disappears from the camera view or changes shape significantly between frames. In the method proposed here, an ellipse based head likelihood model, consisting of gradient along the head boundary as well as a matching score between color histograms as a representation of the interior of (i) an ellipse surrounding the tracked object and (ii) a currently considered ellipse, together with depth information is utilized to find the weights of particles during tracking. Particle locations where the weights have large values are then considered to be the most likely locations of the object of interest. The particle set improves consistency of tracking by handling multiple peaks representing hypotheses in the distribution.

Although the use of color discrimination is connected with some fundamental problems such as the lack of robustness in varying illumination conditions, color is perceived as a very useful discrimination cue because of its computational efficiency and robustness against changes in target orientations. The human skin color filtering has proven to be effective in several settings and has been successfully applied in most of the face trackers relying primarily on color (Hunke and Waibel, 1994) (Fieguth and Terzopoulos, 1997) (Schwerdt and Crowley, 2000) (Sobottka and Pitas, 1996) or on color in conjunction with other relevant information (Darrell *et al.*, 1998) (Hampapur *et al.*, 2003). Color information is particularly useful to support a detection of faces in image sequences because of robustness towards changes in orientation and scaling of an appearance of object being in movement. The efficiency of color segmentation techniques is especially worth to emphasize when a considered object is occluded during tracking or is in shadow.

In our approach we use color histogram matching techniques to obtain information about possible location of the tracked target. The main idea of such an approach is to compute a color distribution in form of the color histogram from the ellipse's interior and to compare it with the computed in the same manner histogram representing the tracked object in the previous iteration. The better a histogram representing the ellipse's interior at specific particle position matches the reference histogram from previous iteration, the higher the probability that the tracked target at considered candidate position is. The outcome of the histogram matching that is combined with gradient information is used to provide information about expected target location and is utilized during weighting particles.

In the context of head tracking on the basis of images from a mobile camera the features which are invariant under head orientations are particularly useful. In general, histograms are invariant to translation and rotation of the object and they vary slowly with the change of angle of view and with the change in scale. A histogram is obtained by quantizing the ellipse's interior colors into K bins and counting the number of times each discrete color occurs. Due to the statistical nature, a color histogram can only reflect the content of images in a limited way and thus the contents of the interior of the ellipses taken at small distances apart are strongly correlated. If the number of bins K is to high, the histogram is noisy. If K is too low, density structure of the image representing the ellipse's interior is smoothed. Histogram based techniques are effective only when K can be kept relatively low and where sufficient data amounts are available. The reduction of bins makes a comparison between the histogram representing the tracked head and the histogram of candidate head faster. Additionally, such a compact representation is tolerant to noise that can result from imperfect ellipse-approximation of a highly deformable structure and curved surface of a face causing significant variations of the observed colors. The particle filter works well when the conditional densities $p(\mathbf{z}_t \mid \mathbf{s}_t)$ are reasonably flat.

It can be demonstrated that with a change of lighting conditions the major translation of skin color distribution is along the lightness axis of the RGB color space. Skin colors acquired from a static person tend to form tight clusters in several color spaces while color acquired from moving ones form widen clusters due to different reflecting surfaces. To make the histogram representation of the tracked head less sensitive to lighting conditions the HSV color space has been chosen and the V component has been represented by 4 bins while the HS components obtained the 8-bins representation.

In order to compare histograms we have implemented the histogram intersection technique (Swain and Ballard, 1991). For a given pair of histograms I and M, each containing j values, the intersection of the histograms is defined as follows: $H = \sum_{u=1}^{K} min(I^{(u)}, M^{(u)})$. The terms $I^{(u)}$, $M^{(u)}$ represent the number of pixels inside the *u*-th bucket of the candidate histogram and the histogram representing the tracked head, respectively, whereas K 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: $H_\cap = H / \sum_{u=1}^{K} I^{(u)}$.

The length of the minor axis of a considered ellipse is determined on the basis of depth information. Taking into account the length of the minor axis resulting from the depth information we also considered smaller and larger projection scale of the ellipse and therefore a larger as well as smaller minor axis about one pixel have been taken into account as well. The length of the minor axis has been maintained by performing the local search to maximize the goodness of the following match: $w^* = \arg\max_{w_i \in W} \{G(w_i) + H_\cap(w_i)\}$, where G and $H_\cap$ are normalized scores based on intensity gradients and color histogram intersection. Particularly, if the length of minor axis of the considered ellipse was different from the length of minor axis of the reference ellipse representing the tracked head, in order to provide j values in the histogram I a histogram normalization with respect to ellipse's area has been realized. The search space W comprises the ellipse's length obtained on the basis of depth information as well as smaller/larger minor axes about one pixel.

The discussed method of target representation has a construction phase and a run phase. In the construction phase which is realized off-line the elliptical upright outlines as well as masks containing interior pixels have been prepared and stored for the future use. We have assumed that a reference ellipse position is located in a central point in a candidate region. Such a candidate area considers all expected head locations which can occur in the next time step. We have then fixed a search strategy allowing us to compute histogram iteratively, i.e. considering adjoining ellipses when processing from top to bottom and from left to right and then from right to left, etc. For each location in the assumed candidate area we constructed a list of positions which should be substituted (added and removed) in the current histogram to determine the histogram at the next location in the utilized search strategy. As a result, for each possible length of the minor axis we obtained a fast strategy to match histograms representing hypothetical head locations. In an on-line phase this strategy allows us to compute the likelihood of each candidate head location and store this information in a two dimensional table, which can be easily accessed by samples. The outlined above algorithm for particle based tracking can be summarized as follows:

1. For the current head location determine the candidate area
2. On the basis of the most recent observation, for each possible head candidate compute a likelihood and store it in a likelihood table
3. Draw N samples
4. Assign weights using the likelihood table
5. Extract the best-fit ellipse
6. Normalize weights
7. Resample with replacement to obtain samples with equal weights
8. Normalize histogram according to area of the ellipse
9. Adapt histogram

The histogram representing the tracked head has been adapted over time. This makes possible to track not only a face profile which has been shot during initialization of the tracker but in addition different profiles of the face as well as the head can be tracked. The actualization of the histogram has been realized on the basis of the equation $M_t^{(u)} = (1-\alpha)M_{t-1}^{(u)} + \alpha I_t^{(u)}$, where I_t represents the histogram of the best-fit ellipse interior, whereas $u = 1...K$. The samples are propagated on the basis of a dynamic model $\mathbf{s}_t = A\mathbf{s}_{t-1} + w_t$, where A denotes a deterministic component describing a constant velocity movement and w_t is a multivariate Gaussian random variable. The diffusion component represents uncertainty in prediction and therefore provides a way of performing a local search about a state. The weight of each hypothetical head region $\pi_t^{(n)}$ is dependent on normalized intensity gradients and color histogram intersection which correspond to ellipses obtained in the local search in the space W.

Tracking on the basis of moving camera

A kind of human-machine interaction which is useful in practice and can be very serviceable in testing a robustness of a tracking algorithm is person following with mobile robot. In work (Meier and Ade, 1999) the CONDENSATION based algorithm is utilized to keep track of multiple objects with a moving robot. The tracking experiments described in this section were carried out with a mobile robot Pioneer 2DX (*ActivMedia Robotics*, 2001) equipped with commercial binocular Megapixel Stereo Head. The dense stereo maps are extracted in that system thanks to small area correspondences between image pairs (Konolige, 1997) and therefore poor results in regions of little texture are often provided. The depth map covering a face region is usually dense because a human face is rich in details and texture, see Fig. 1. Thus this stereovision system used as a separate source of information aids the process of approximating the tracked head with an ellipse.

a)

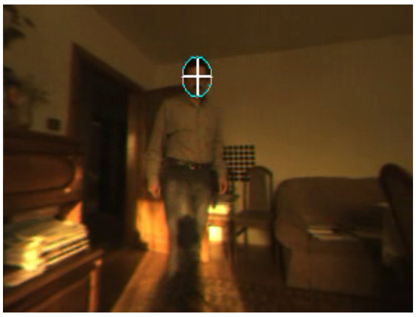
b)

Figure 1: Depth images. a) frame #1, b) frame #600

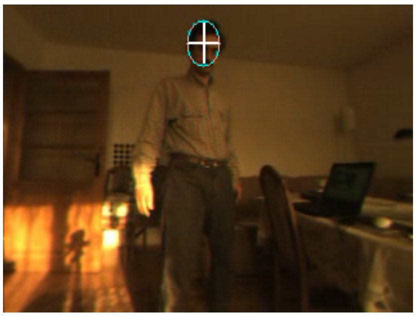
A typical laptop computer equipped with 2 GHz Pentium IV is utilized to run the prepared visual tracker operating at 320x240 images. The position of the tracked face in the image plane as well as person's distance to the camera are written asynchronously in block of common memory which can be easily accessed by Saphira client. Saphira is an integrated sensing and control system architecture based on a client server-model whereby the robot supplies a set of basic functions that can be used to interact with it (*ActivMedia Robotics, 2001)*. Every 100 milliseconds the robot server sends a message packet containing information on the velocity of the vehicle as well as sensor readings to the client. During tracking, the control module keeps the user face within the camera field of view by coordinating the rotation of the robot with the location of the tracked face in the image plane. The aim of the robot orientation controller is to keep the position of the tracked face at specific position in the image. The linear velocity has been dependent on person's distance to the camera. In experiments consisting in person following a distance 1.6 m has been assumed as the reference value that the linear velocity controller should maintain. To eliminate needless robot rotations as well as forward and backward movements we have applied a simple logic providing necessary insensitivity zone. The PD controllers have been implemented in the Saphira-interpreted Colbert language (*ActivMedia Robotics, 2001)*. The tracking algorithm was implemented in C/C++ and runs at frame rates of 8-10 Hz depending on image complexity. We have undertaken experiments consisting in following a person facing the robot within walking distance without the tracked face loss. Experiments consisting in realization of only a rotation of mobile robot which can be seen as analogous to experiments with a pan-camera have also been conducted. In such experiments a user moved about a room, walked back and forth as well as around the mobile robot. Our experiment findings show that thanks to stereovision the tracked head is properly approximated by the ellipse and in consequence, sudden changes of the minor axis length as well as ellipse's jumps are eliminated. Figure 2 indicates selected frames from the discussed scenario, see also Fig. 1.

The depth map covering the face region is usually dense and this together with skin-color and symmetry information as well as eyes-template assorted with the depth has allowed us to apply the eigenfaces method (Turk and Pentland, 1991) and to detect the presence of the vertical and frontal-view faces in the scene very reliably (Kwolek, 2002) and thus to initialize the tracker automatically. Thanks to the head position it is possible to recognize some static commands on the basis of geometrical relations of the face and hands (Kwolek, 2003*a*) and to interact with mobile robot during person following. In work (Kwolek, 2003*b*) 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 best-fit head ellipse is determined in the local search 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. Using the discussed system we have realized experiments in which the robot has followed a person at distances which beyond 100 m without the person loss. In this approach the best-fit head ellipse is extracted on the basis of sample collection. By dealing with multiple hypotheses this approach can track a head reliably even in cases of temporal occlusions and varying illumination conditions.

a)



b)



c)

d)

Figure 2: Face tracking relying only upon a rotation of the moving camera. a) frame #1, b) frame #32, c) frame #300, d) frame #600

Conclusions

We have presented a vision module that robustly tracks and detects a human face. By employing color, stereovision as well as elliptical shape features the proposed method can track a head in case of dynamic background. The combination of above-mentioned cues and particle filter seems to have a considerable perspective of applications in robotics and surveillance. To show the correct work of the system, we have conducted several experiments in naturally occurring in laboratory circumstances. In particular, the tracking module enables the robot to follow a person.

References

ActivMedia Robotics (2001). Pioneer 2.

Birchfield, S. (1998). Elliptical head tracking using intensity gradients and color histograms. In: *IEEE Conf. on Computer Vision and Pattern Recognition*. pp. 232–237.

Bradski, G. R. (1998). Computer vision face tracking as a component of a perceptual user interface. In: *In Workshop on Applications of Computer Vision*. pp. 214–219.

Comaniciu, D., V. Ramesh and P. Meer (2000). Real-time tracking of non-rigid objects using mean shift. In: *Proc. Int. Conf. Comp. Vision Pattern Recognition*. pp. 142–149.

Darrell, T., G. Gordon, M. Harville and J. Woodfill (1998). Integrated person tracking using stereo, color, and pattern detection. In: *Proc. of the Conf. on Computer Vision and Pattern Recognition*. pp. 601–609.

Dellaert, F., W. Burgard, D. Fox and S. Thrun (1999). Using the condensation algorithm for robust, vision-based mobile robot localization. In: *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition*. pp. 588–594.

Fieguth, P. and D. Terzopoulos (1997). Color-based tracking of heads and other mobile objects at video frame rates. In: *Proc. of the IEEE Conf. Comp. Vision Pattern Recognition*. pp. 21–27.

Hampapur, A., S. Pankanti, A. Senior, Y. Tian, L. Brown and R. Boole (2003). Face cataloger: Multi-scale imaging for relating identity to location. In: *IEEE Conf. on Advanced Video and Signal Based Surveillance, IEEE Comp. Society Press*. pp. 13–20.

Hunke, M. and A. Waibel (1994). Face locating and tracking for human-computer interaction. In: *Proc. of the 28th Asilomar Conf. on Signals, Systems and Computers*. pp. 1277–1281.

Isard, M. and A. Blake (1996). Contour tracking by stochastic propagation of conditional density. In: *European Conf. on Computer Vision, Cambridge*. pp. 343–356.

Konolige, K. (1997). Small vision system: Hardware and implementation. In: *Proc. of Int. Symposium on Robotics Research*. pp. 111–116.

Kwolek, B. (2002). Pointing arm-posture recognizing using stereo vision system. *(Best Presentation Award the Int. Conf. on Computer Vision and Graphics ICCVG 2002) Archives of Theoretical and Applied Computer Science* **14**(4), 307–319.

Kwolek, B. (2003*a*). Face tracking system based on color, stereovision and elliptical shape features. In: *Proc. of the IEEE Conf. on Advanced Video and Signal Based Surveillance, IEEE Comp. Society Press*. pp. 21–26.

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

McKenna, S. J., S. Jabri, Z. Duric and H. Wechsler (2000). Tracking interacting people. In: *Proc. of the Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition*. pp. 348–353.

Meier, E. B. and F. Ade (1999). Using the condensation algorithm to implement tracking for mobile robots. In: *Proc. of the Third European Workshop on Advanced Mobile Robots, Eurobot*. pp. 73–80.

Schwerdt, K. and J. L. Crowley (2000). Robust face tracking using color. In: *Proc. of 4th Int. Conf. on Automatic Face and Gesture Recognition, IEEE Comp. Society Press*. pp. 90–95.

Sobottka, K. and I. Pitas (1996). Segmentation and tracking of faces in color images. In: *Proc. of the Second Int. Conf. on Automatic Face and Gesture Recognition*. pp. 236–241.

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

Turk, M. A. and A. P. Pentland (1991). Face recognition using eigenfaces. In: *Proc. of Conf. on Computer Vision and Pattern Recognition*. pp. 586–591.

Wren, C., A. Azarbayejani, T. Darrell and A. Pentland (1997). Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 780–785.

B. Kwolek received the M.Sc. degree in electrical engineering from Rzeszów University of Technology in 1988. In 1998 he obtained his Ph.D. degree from Technical University in Cracow in the field of computer science. His dissertation focused on parallel algorithms of image processing and analysis for real-time visual tracking. Now he is with the Computer and Control Engineering Chair at the Rzeszów University of Technology. His areas of interest lie in image processing and recognition, artificial intelligence and autonomous mobile agents.