PERSON TRACKING USING OMNIDIRECTIONAL VIEW

Ing. Igor POTÚČEK, Doctoral Degree Program (1)
Dept. of Computer Graphics, FIT, BUT
E-mail: potucek@fit.vutbr.cz

Supervised by: Dr. Pavel Zemčík

ABSTRACT

This paper presents an approach for tracking parts of human body. A color detection method used is suitable for recognition of hands and heads. Kalman filter provides prediction and identification of trajectory of moving objects in sequence. Presented system will serve for human tracking and gesture recognition.

1 INTRODUCTION

The main purpose of the method is monitoring gestures and movement of people in a meeting room. Such assignment traditionally requires several video cameras. Disadvantage of such approach can be solved by using one multi directional view system. The system is based on standard video camera equipped with hyperbolic mirror that allows capturing a large portion of the space angle - e.g. 120 x 360 degrees. Two options exist to set-up such system. One of them has a mirror on the top holder since we need greater resolution in upper part of the image, where faces and hands occur.

2 TRANFORMATION OMNIDIRECTIONAL VIEW

An image obtained by using the hyperbolic mirror can be transformed into a standard perspective image. It is used for simple transformation that assumes linear pixel distribution along the direction of radius. The coordinates of the panoramic view are P_x and P_y. We must transform these coordinates into multi directional image. The real world elements are projected on a cylinder, radius of which is equal to *d*. The axis of the cylinder is identical with the mirror and camera axis. The horizontal size of the panoramic view is a perimeter of cylinder $Width = 2\pi d$. The coordinates of the video camera image X_M are following:

$$x_M = (d - P_y)\cos\alpha + CenterX \qquad (2.1)$$

$$y_M = (d - P_y)\sin\alpha + CenterY \qquad (2.2)$$

where angle $\alpha = \frac{P_x}{d}$.

The calculated pixels in the camera image do not correspond “one to one” to the pixels of projected image so sub pixel anti-liasing methods should be used. In this case, it is sufficient to use weighted average of neighboring pixels, as the size of the pixels’ output is comparable to its input.

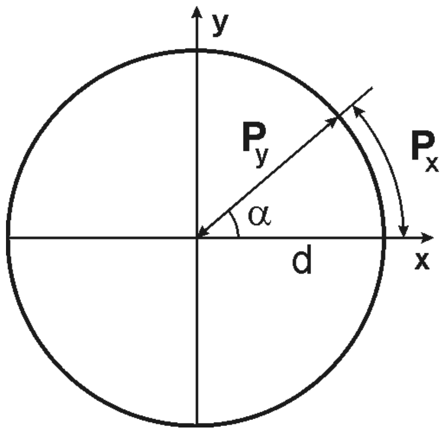

Fig. 1: *Transformation into panoramic view*

It is suitable to trim the part of image that contains the center of the multi directional view. This usually displays part of the camera image anyway.

3 SKIN DETECTION AND SEGMENTATION

Color is the key feature for hands and head detection. It is the method that is most often used for human parts detection. Its main advantage is low computational cost. On a negative side, it is only a partial method because of its low reliability. Appearance of the skin-tone color depends on the lighting conditions. Therefore we used normalized rg-color space that provides good solution to the problem of varying brightness. Normalized rg-color is computed from RGB values.

The *r* and *g* components create 2D color space with the normal probability distribution. Various face color pixels are picked manually and then color class Ω_k is computed. The color class Ω_k is determined by its mean vector μ_k and the covariance matrix K_k by its distribution. We need to compute probability of each pixel in the image by this equation:

$$p(c \mid \Omega_k) = \frac{1}{2\pi\sqrt{\det K_k}} \exp\left(-\frac{1}{2}(c-\mu_k)^T K_k^{-1}(c-\mu_k) \right) \qquad (3.1)$$

Segmentation serves to separate hands and head from the rest of the visual image. The output of skin color detection is a binary map. It informs us whether a given pixel is skin or non-skin and that is based on a parameter threshold. The spatially separated groups of skin pixels are treated as separate objects. A seed algorithm is used for the extraction. The extracted information about shape, size and position of separated objects is obtained from groups of skin pixels. Size threshold is used as a recognition criterion. The size is than computed as number of joint skin pixels. If objects are of smaller size than a given threshold then they are discarded. The area of object is restricted by outer pixels and their center of gravity is computed as the average in x and y-axis chart. This information than serves us to mark the searched object.

4 MOTION CORRESPONDENCE

The tracked trajectory consists of objects found in the frames at certain sequence. It is necessary to determine the correspondence of these objects in order to determine trajectories in the sequence. Our aim is to attach to an object in frame *t* its correspondent object in the frame *t+1*. Let us suppose that we have a prior knowledge of the object's motion in the previous frames. We are then limited to prediction of the next object position from its previous motion. Prediction of object position is therefore based only on the positions in past frames. Thus we can define a boundary in which the searched object occurs. The boundary we use is circular and it is defined by the radius *r*.

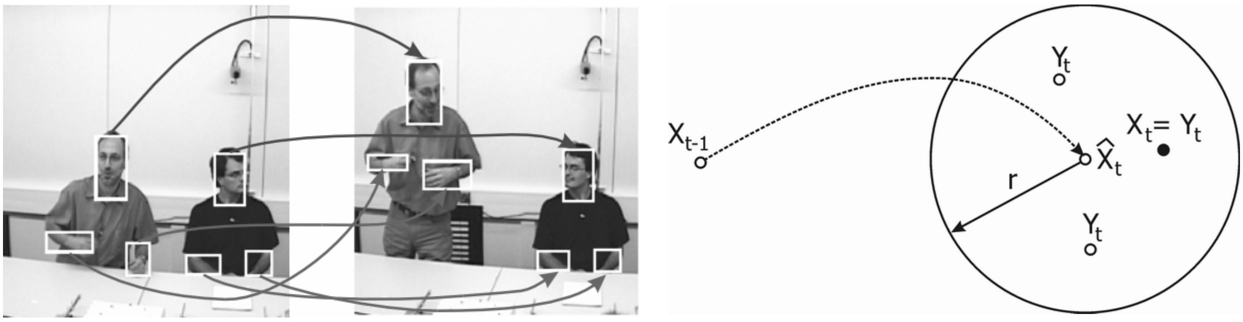

Fig. 2: *Selection of right corresponding object*

A movement vector for tracked objects is summarized in the accepted kinematics model, which best describes its changes with time. Position $\hat{X}_t$ to the object X_{t-1} from the previous frame t-1 is predicted. The objects Y_t are candidates, which could be searched for correspondent object X_t. The choice of the best candidate is conducted according to its distance from the predicted position and a comparison of object's features.

The problem is restricted to 2D projection of real 3D movement. The video sequence contains n frames and each frame contains certain number of tracked objects. This number is tracked in each frame, because new objects are being detected and existing objects may disappear that represents initiation and termination of the trajectory. A uniqueness constraint is adopted; stating that one detected object uniquely matches one featured object. The objects, which overlap each other, are resolved as one detected object and in this case one of the trajectories are terminated.

4.1 POSITION PREDICTION

Position prediction is an important part of correspondence determination when we do not have enough information about features of tracked objects. It can be a suitable tool when tracked objects overlap or disappear. The most often used method is Kalman filtering. Kalman filter is adaptive filter used for modeling states of a discrete dynamic system. Advantage of this filter is its recursive structure, whereas its coefficients modify each step of computation on the basis of available information in such a way, that it affords optimal estimation of a future state. New filter originates by correcting previous filter in each step, without remembering information about previous positions. Two cases can occur:

1. Searched object fits into predicted area.

2. Predicted area is empty.

The object position in next frame is predicted at first. Failing that, we can extrapolate position from available information in previous frames. This extrapolation can be done on only certain number of frames, which is dependent on size of uncertainty. The extrapolation can resolve short overlap or disappearance. If the object appears at extrapolated position, its trajectory continues further.

5 EXPERIMENTAL RESULTS

The Sony DCR-TRV310E digital camera and Neovision hyperbolic glass mirror H3G with mirror holder were used. The captured sequence is unwrapped into normal panoramic image sequence that can be seen in the figure 3. An algorithm with pixel value averaging is sufficient for this task.

Fig. 3: *Original and unwrapped image*

The pixels for skin detection are picked manually from skin regions. Figure 4 shows detected areas that contain skin pixels. Often the results of skin color detection either contain noise, or many colored objects create considerable clutter in skin probability image and thus make skin regions not so clearly distinguishable.

Fig. 4: *Skin color detection*

The next step is the object extraction and for that size of threshold is used. Objects of size smaller then 0.1% of total image are discarded. This criterion has shown good results in object selection and it can separate the noise. Problematic part is the set-up of Kalman filter parameters for prediction. There are too many variables that are interdependent. It is most suitable method for these problems, because it provides good prediction of new position of tracked object when it disappears. The tracked objects are marked by rectangle seen in figure 5.

Fig. 5: *Object tracking*

6 CONCLUSION

This paper describes use of multidirectional system for person’s tracking. The significant advantage of this system is its coverage of whole area - 360 degrees - while using one standard camera. The primary method was chosen the skin pixel detection, which is key feature for hands and head detection. Its main advantage is low computational cost. The methods have to be improved in order to achieve yet better results. The main stages of improvement are:

1. Skin color detection – Implementation of adaptive classification for minimizing of problems with changes in illumination.
2. Implementation of shape detection for improvement of the head recognition by using algorithm of ellipse fitting.

ACKNOWLEDGEMENTS

The paper has been prepared as a part of the solution of EU IST Program project IST-2001-34485.References

REFERENCES

[1] Svoboda, T.: Central Panoramic Cameras Design, Geometry, Egomotion, Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University, September 30, 1999.

[2] Faugeras, O.: Three-Dimensional Computer Vision, A Geometric Viewpoint, The MIT Press, Cambridge, Massachusetts, 1993.

[3] Eiserloh, P.: An Introduction to Kalman Filters and Applications, Assault and Special Projects Naval Air Warfare Center, China Lake, 2002.

[4] Kruppa, H., Bauer, M., Schiele, B., Skin Patch Detection in Real-World Images, Perceptual Computing and Computer Vision Group, ETH Zurich, Schwitzerland, 2002.

[5] Vezhnevets, V.: Method for Localization of Human Faces in Color-Based Face Detectors and Trackers, Department of Computational Mathematics & Cybernetics, Moscow State University, Vorobjovy Gory, Russia, 2001.

[6] Jones, M., Rehg, J.: Statistical Color Models with Application to Skin Detection, Cambridge Research Laboratory, Computer Vision and Pattern Recognition (CVPR99), Ft.Collins, CO, 274-280, June, 1999.