2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Optimum Region-of-Interest Acquisition for Intelligent Surveillance System using Multiple Active Cameras

Young-Ouk Kim*, Chang-Woo Park*, and Ha-Gyeong Sung *, Chang-Han Park+, Jae-Chan Namkung+
*Korea Electronics Technology Institute
203-103 BD 192, Yakdae-Dong, Winmi-Gu, Puchon-Si, Kyunggi-Do,Korea
+Department of Computer Engineering, Kwangwoon University, Seoul, Korea
Email: kimyo@keti.re.kr

Abstract-- In this paper, we present real-time, accurate face region detection and tracking technique for an intelligent surveillance system. It is very important to obtain the high-resolution images, which enables accurate identification of an object-of-interest. Conventional surveillance or security systems, however, usually provide poor image quality because they use one or more fixed cameras and keep recording scenes without any clue. We implemented a real-time surveillance system that tracks a moving person using four pan-tilt-zoom (PTZ) cameras. While tracking, the region-of-interest (ROI) can be obtained by using a low-pass filter and background subtraction. Color information in the ROI is updated to extract features for optimal tracking and zooming. The experiment with real human faces showed highly acceptable results in the sense of both accuracy and computational efficiency.

1. INTRODUCTION

Recently, intelligent surveillance systems gain more attraction than simple CCTV systems, especially for complicated security environment. The major purpose of these systems is to monitor an intruder. More specifically, accurate identification of each intruder is more important than simply monitoring what they are doing. Most existing surveillance systems simply keep recording the fixed viewing area, and some others adopt the tracking technique for wider coverage [1][2]. Although panning and tilting the camera can extend the viewing area, only a few automatic zoom control techniques for acquiring the optimum ROI has been proposed.

In this paper, we present efficient, real-time implementation of a 4-channel automatic zoom in/out module for high-resolution acquisition of face regions. Although object tracking is an active research topic in computer vision, its practical implementation is still under development due to high computational complexity and difficulty in analysis of false detection.

Optimum zooming in/out control plays an important role in tracking performance [3] and at the same time it provides highly accurate identification of an intruder. To realize this function, we first detect and track the face of moving object in front of four PTZ cameras, and extract several features for tracking and optimum zooming scale. Existing real-time tracking techniques include: CAMSHIFT [4], condensation [5] and adaptive Kalman filtering. But these algorithms can't rack the object when the target moves far away from the camera.

Because faced chroma distribution information is factor that does importance most, and relative computing speed is fast in a face tracking technology, a lot of algorisms were invented. Yang and Weibel [6] proposed a real time face tracking using on normalized color. Yao and Gao [7] present face tracking based on skin chroma and lip chroma transform. Huang and Chen [8] built a statistical color model and deformable template for tracking of multiple faces. Problem in tracking that use the face color distribution is that is hard to solve occlusion in plural face and background that have similar color distribution. So complicated methods like as model based tracking and deformable-template matching was proposed. However these methods are hardly achieving real time performance.

In this paper, we present multiple techniques using color distribution of face for locating the face and simple ellipse template matching for mentioned occlusion problem.

2. SYSTEM OVERVIEW

The architecture of the proposed surveillance system is shown in Figure 1. We used four PTZ cameras for tracking and recording the moving object and additional four fixed cameras for recording the wide-angle view.

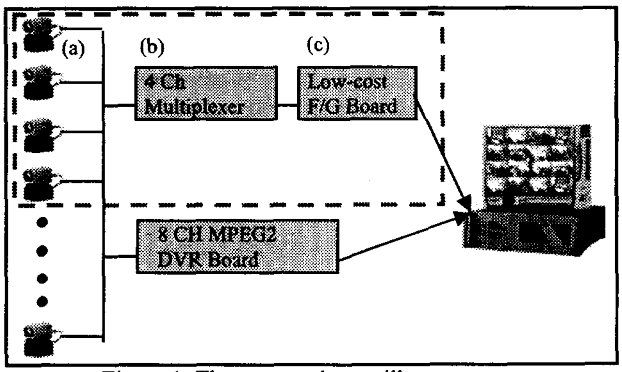

Figure 1. The proposed surveillance system

Figure 1(a) represents a PTZ camera with 25x zooming ratio. Figure 1(b) represents a 4-channel multiplexer. Since most surveillance systems have a recording device with various kinds of image compression, we deploy another frame grabber to acquire raw image data as shown in Figure 1(c). In this system, we use Microsoft DirectShow® to minimize redundant computations for real-time image processing.

DirectShow® can seamlessly integrate the modules to play back or capture a particular media type [9].

3. AUTOMATIC ZOOMING WITH FACE TRACKING

For accurate identification of an intruder, the optimum zooming ratio must be automatically generated in the system. This optimum zooming ration can be obtained only by a robust tracking algorithm. Features used by most tracking algorithms include: (i) color, (ii) motion, and (iii) contour. These algorithms may fail when the target object becomes extremely small in the viewing region since color, motion, and contour information are subject to be unstable [10].

In this paper, we adopt the low-pass filter based technique, proposed in [11], to detect the candidate area of a moving object. After detecting the moving object, we segment the face area from the background based on the HSV color system. We can then extract the appropriate zooming ratio and features for tracking based on the fault analysis of four inputs at the same time. Figure 2 shows the flowchart of the proposed algorithm.

3.1. Adaptive motion detection

In a tracking algorithm, automatic ROI detection is very important to meet the perceptual requirement.

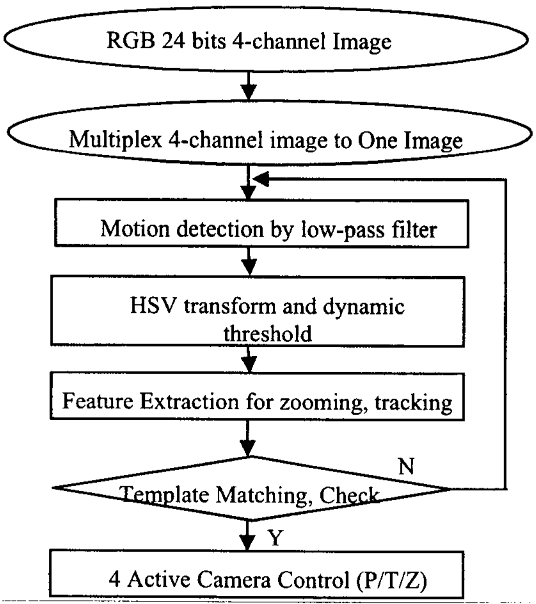

Figure 2. PTZ camera control algorithm

This processing, in general, consumes large amount of system resources because of computationally expensive processing, such as: color correlation, blob detection, region growing, prediction, and contour modeling [4].

We were able to detect the reasonably accurate candidate region using the Gaussian low-pass filter. The candidate area of a moving object is obtained as

$$\hat{I}_{nm} = I_{ng} - I_{mg}, \qquad (1)$$

where, I_{ng} and I_{mg} respectively represent the Gaussian filtered n and m-th image frames, which are converted to the normalized RGB color coordinate system.

This method can be applied as is useful when destination object (face) as well as face candidate area detection disappears in tracking initializing or tracking failed.

Figure 3. Candidate moving area detection

Figure 3 shows the result of candidate moving area detection: top left image represents I_5, top right image I_{23}, bottom left $I_{5g} - I_{8g}$, and bottom right $I_{23g} - I_{25g}$.

3.2. Skin color segmentation from background

Color information of a moving object is one of the most important features. However color changes due to illumination change and reflected light. The HSV color model is a well known to be robust since it is affected by light changes less than any other color models. In the proposed surveillance system, the skin color of moving object changes according to the distance between the object and camera even if the light condition is fixed.

Figure 4 presents experimental results of the skin color changes according to the distance between the cameras and the moving object. In this figure the horizontal axis represents the hue value of face and the vertical axis represents the number of pixels having the same hue value. As shown in this figure, we see the distribution of hue values changes according to the distance from camera even for the same face.

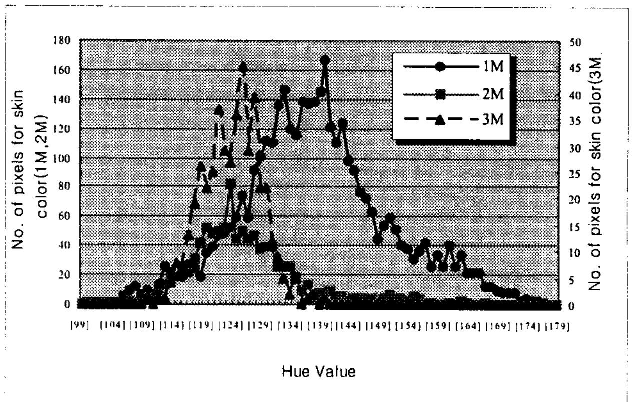

Figure 4. Skin color histograms in three different distances

We can extract maximum, low-threshold, and high-threshold values of a face using the previously defined candidate area. These three variables can efficiently segment the face region form the background. Figure 5 presents hue distribution of a face within the ROI with 3 values.

Using the three values, $f(x)_{Max}$, $f(x_i)_{Low\text{-}th}$, and $f(x_i)_{Hi\text{-}th}$, we can segment the face region within the ROI from the background. The hue index of $f(x_i)_{Max}$ can be iteratively calculated and the other variables $f(x_i)_{Low\text{-}th}$, and $f(x_i)_{Hi\text{-}th}$, can be formulated as

$$f(x_i)_{Low\text{-}th} : f'(x_i) \bullet f'(x_{i-1}) \leq 0 \quad (f(x_i) < f(x)_{Max}), \quad (2)$$

$$f(x_i)_{Hi\text{-}th} : f'(x_j) \bullet f'(x_{j+1}) \leq 0 \quad (f(x_j) > f(x)_{Max}). \quad (3)$$

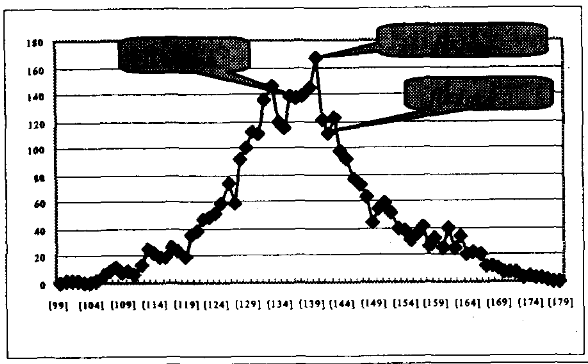

Figure 5. Hue distribution within the ROI

Figure 6 respectively shows the original input image, the corresponding HSV image, and the segmented face region from the background.

Figure 6. Skin color segmentation result

3.3. Feature extraction for zooming and tracking

In this paper, we select three features for automatic zooming and face tracking. The first feature is the mean location of hue values, x_c, y_c which are located between $f(x_i)_{Low\text{-}th}$, and $f(x_i)_{Hi\text{-}th}$, within the detected ROI.

$$x_c = \frac{\sum_x H(x,y)}{E_H}, y_c = \frac{\sum_y H(x,y)}{E_H}, \quad (4)$$

where $H(x,y)$ represents the pixel location of an effective hue value and E_H the number of selected pixels having effective hue values. The second feature is area of detected ROI, and the third is effective pixel ratio within the detected ROI. The mean location x_c and y_c decide the direction of moving object and the second feature A_{ROI} determines the optimum zooming ratio and the third feature R_{ROI} is used for fault detection of zooming and tracking. The second and the third features can be formulated as

$$A_{ROI} = Wid_{ROI} \times Hi_{ROI}, \text{ and } R_{ROI} = \frac{E_H}{A_{ROI}}. \quad (5)$$

The automatic zooming is performed using the A_{ROI} feature. There are two experimentally selected limiting values for automatic zooming, that is, *Tele* and *Wide*. If A_{ROI} is greater than *Wide*, the zoom lens turns wide for zooming down, and vice versa. Figure 7 presents experimental results of the proposed face tracking algorithms using only pan/tilt function. And the result of 4-channel automatic zoom with face tracking is shown in Figure 8.

Figure 7. Single face tracking

Figure 8. Automatic zooming with face tracking

In (5), the effective pixel ratio indicates an error probability of zooming and tracking. If this value is smaller than the pre specified value, we must detect a new candidate area of moving object having with the last $f(x_i)_{Low\text{-}th}$ value, and $f(x_i)_{Hi\text{-}th}$ value. In this time dynamic change of ROI is useful. This process is presented in Figure 9.

Figure 9. Dynamic change of ROI

3.4. Simple template matching for occlusion problem

In order not to obtain the abnormal increase of face rectangular region by another face, an ellipse fitting is performed every 30 frame. Utilizing the generalized Hough transform in the edge image of the rectangular region searched based on color distribution can make the ellipse fitting of face area.

The implicit equation with four parameters to be determined of ellipse p, q, a, b is

$$p^2(x-a)^2 + q^2(y-b)^2 = p^2q^2 \qquad (6)$$

and to reduce the size of parameter space and computational burden, the major and minor axis of the ellipse are restricted by using the ratio of them as

$$p = 1.3q \qquad (7)$$

and hence, the resultant parameters to be searched are reduced as :

The Hough transform procedure is shown below.

1 Quantize the parameter space appropriately.
2 Assume that each cell in the parameter space is an accumulator. Initialize all cells to zero.
3 For each point in the image space, increment by 1 each of the accumulator that satisfy the equation (1).
4 Search the maximum in the accumulator array.

The ellipse fitting followed by region search using color distribution can make the detection robust. The following figures show the ellipse fitting in the edged image.

Figure 10. Occlusion between two faces (top), Sobel edge within the ROI (down-left), Ellipse fitting (down-right)

4. EXPERIMENTAL RESULTS

Performance of the proposed algorithm is evaluated to measure the processing time for a set of algorithms. The tested image frames of 320×240 resolution were used. Table 1 summarizes the processing time of the algorithm using different PC platforms. For real-time tracking, at least 15 frames per second (FPS) is required, and the algorithm showed acceptable speed even with the lowest grade PC (Pentium 3, 667MHz).

Step \ CPU	P3-0.6GHz	P3-1.2GHz	P4-1.7GHz
Motion detection*	(10.22)	(8.38)	(7.03)
HSV transform	36.96	28.55	24.86
Dynamic threshold	6.72	5.19	4.52
Feature extraction	4.70	3.63	3.16
Fault analysis Template Match*	2.02 (50.2)	1.56 (35.2)	1.36 (31.4)
Camera interface	3.36	2.60	2.26
Total Time (ms) **Speed (fps)**	**67.20 (ms)** **14.88 (fps)**	**51.90 (ms)** **19.27 (fps)**	**45.20 (ms)** **22.12 (fps)**

*** (This is not executed every frame)**

Table 1. Result of algorithm speed

5. CONCLUSION

In this work we presented the real-time, the optimum ROI detection technique, especially for face tracking purpose in an intelligent surveillance system. Since accurate identification of a human face is more important than just tracking a moving object, efficient method to detect the face region and the corresponding high-resolution acquisition are needed.

The proposed intelligent surveillance system with built-in automatic zooming and tracking algorithms can efficiently detect high-resolution face images and stably track the face. The major contribution of this work is the development of real-time, robust algorithms for automatic zooming and tracking, and an intelligent surveillance system architecture using multiple PTZ cameras with seamless interface.

6. REFERENCES

[1] L. S. Davis, I. Haritaoglu, and D. Harwood, "W^4: real-time surveillance of people and their activities," *Proc. 2000 Int. Conf. Pattern Recognition*, vol. 4, 2000.

[2] R. T. Collins, O. Amidi, and T. Kanade., "An active camera system for acquiring multi-view video," *Proc. 2002 Int. Conf. Image Processing*, 2002.

[3] X. Cindy, F. Collange, F. Jurie, P. Martinet, "Object tracking with a pan-tilt-zoom camera: application to car driving assistance" *Proc. 2001 Int. Conf. Robotics, Automation,* 2001.

[4] G. R. Bradski, "Computer vision Face Tracking For use in a Perceptual User Interface," *Intel Tech. Journal*, Q2, 1998.

[5] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking", *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5-28, 1998.

[6] J. Yang and A. WaiBel, "A Real-Time Face Tracker", *Proceedings of WACV'96*, pp. 142-147, 1996.

[7] H. Yao and W. Gao, "Face locating and tracking method based on chroma transform in color images", *Signal Processing Proc.2000* , vol 2. pp. 1367-1371, 2000.

[8] F. Huang and T. Chen, "Tracking of multiple faces for human-computer interfaces and virtual environments" *IEEE International Conference on Multimedia and Expo 2000* pp. 1563-1566, vol.3.

[9] M. Linetsky, *Programming Microsoft Direct Show*, Wordware Publishing, Inc. 2002.

[10] P. Fieguth and D. Terzopoulos, "Color-based tracking of heads and other mobile objects at video frame rates," *Proc. 1997 Computer Vision, Pattern Recognition*, pp. 21-27, 1997.

[11] B. Menser and M. Wien, "Automatic face detection and tracking for H.263 compatible region-of-interest coding," *Proc. SPIE*, vol. 3974, 2000.