Real-Time Head Tracking System With an Active Camera*

Tao Yang, Quan Pan, Jing Li, Yongmei Cheng, Chunhui Zhao

College of Automatic Control
Northwestern Polytechnical University
Xi'an, 710072 China

yangtaonwpu@msn.com; quanpan@nwpu.edu.cn

***Abstract* – Human head detection played an important role in applications such as video surveillance, perceptual user interface, face recognition and people tracking. In this system, an improved moving object detection algorithm was used to start the head detection and tracking process, which was based on the gradient and the color information. Firstly, we decreased the image size through a sampling method, then we detected skin regions over the entire image, and generated a searching space from them, after that an ellipse model was used to detect the head in the searching space, and the detection result was reflected to the original image, finally the final result above was being used to control the active camera. The evaluation of this system's performance was shown in the last part of the article. Experiment result shows that this system runs in real-time on a standard PC, being robust to changing background, partial occlusion, clutter, face scale variations, rotations in depth, and fast changes in subject or camera position.**

Index Terms - Face detection, head tracking, people tracking, video surveillance.

I. INTRODUCTION

Head detection is currently one of the most active research topics in the domain of computer vision. This strong interest is driven by a wide spectrum of promising applications in many areas such as anti terrorism, video surveillance [1], virtual reality, perceptual user interface and so on. It provides input to high-level processing such as recognition and identification, or is used to initialize the analysis and classification of human activities [2].

Various methods have been proposed for head detection and tracking in the past several years. Many of these deal with detecting faces in a scene. Once a face has been detected, the head region is tracked using features such as the geometry characters of person's head [3,4], the color information of skin [5,6] and so on. Although they performs well in some certain situations such as static background, no disturbing and sheltering, small image size etc. They still suffer when the environment has a cluttered and changing background; in addition, with the increase of the size of the image, it is hard for them to be real-time.

In this paper, we present a robust, reliable head detection and tracking system under a complex environment. The system detects the people through analyzing the state of some detection lines in the image over a period of time, once there is somebody in the image, a searching algorithm is active to find the exact position of the person's head. Here, we integrate several different visual modules, each using a different criterion and each focusing on the different characters of the target. As we all know, the gradient describes the outline of the object in the image, which the color can not represent at all. They are orthogonal to each other so that when one module fails the other one can come to its aid, finally, the system controls the rotating direction and time of the camera by the detection result.

The remainder of this paper is organized as follows. Section 2 briefly introduces the outline of the system. Section 3 describes the moving object detection method we used in this system. Section 4 explains the head detection and tracking algorithm. Section 5 introduces on the camera control model. Section 6 focuses on the performance evaluation of this system, and finally Section 7 and 8 contain the experimental results and discussion of future extensions.

II. OUTLINE OF THE SYSTEM

A block diagram of our system is shown in Fig. 1. Firstly, a moving object detection algorithm is used to find whether there is a target in the view field of the camera or not (Sec.3). After that a head segmentation and detection algorithm, which is based on skin color and head shape model is active to detect the person's head in the image (Sec.4). Finally, the active camera is controlled by the servocontrol system whose input signal is the detection result of the target (Sec.5).

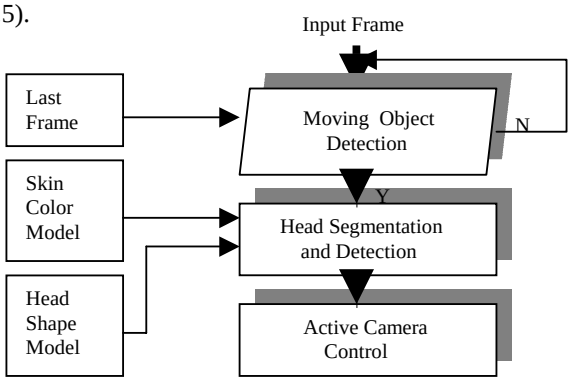


Fig. 1 The block diagram of the system.

* This work is supported by the science and technical research program of ShaanXi province, #2003k06-G15.

III. MOVING OBJECT DETECTION

Background subtraction, the process of subtracting the current image from a reference image, is a simple and fast way to obtain the moving object in the foreground region and has been employed by many surveillance systems [1]. However, algorithms that use a stationary reference image are sensitive to the environment changing such as the changes of illumination, the shadow and even the changes of the background, what's more, those algorithms above have to store and subtract the whole image to find the moving objects which is not only with heavy compute burden, but also easy to be influenced by the noises from both the environment and the camera itself.

In order to find the moving object in real time, we manually set several detection lines in the image, whose positions are decided by our needs. By contrast to those stationary reference image methods, we use a dynamic one, each time we update the whole reference image with the last frame, and only subtract the detection lines between the current frame and the reference image. In this way can we greatly decrease the compute burden and improve the robust of the head detection and tracking system. We set the detection lines under the following assumptions.

- People have to pass the detection line to enter into the camera's field of view.
- The sampling speed camera is high enough that the environment changes between two related frames can be neglected.
- The number of the frames in which people is passing the detection line is more than one frame.

We define $G_i(x, y, k)$ to represent the subtraction result of the pixels on the ith detection line L_i at the time k, here (x, y) is the position of the pixels. The mean and variance of $G_i(x, y, k)$ are $\overline{G_i}(k)$ and $R_i(k)$, then the state of the ith detection line is

$$S_i(k) = \begin{cases} 0 & R_i \le Th \\ 1 & R_i > Th \end{cases} \quad (1)$$

where

$$R_i(k) = \frac{1}{M} \sum_{x,y \in L_i} (G_i(x, y, k) - \overline{G_i}(k))^2 \quad (2)$$

Here Th is the threshold and M is the number of pixels on the detection line. Thinking about the speed of the CCD camera can catch the image sequence at 25frames/Sec, we can take it for granted that the third assumption above comes into existence. Then the moving object detection result is represented as P.

$$P = \begin{cases} 1 & S_i(k+m) \equiv 1, m = 0,1,2 \\ 0 & otherwise \end{cases} \quad (3)$$

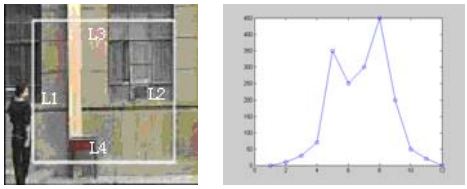


Fig. 2 Left: Image of our moving object detection subsystem. Right: The variance of the state in the detection line.

Fig.2(left)shows four detection lines in an image, and Fig.2(right) illustrates the state S_1 changes of the detection line L_1 when a person passing by it, the variance will be greatly changed and can be easily detected and calculated.
In the detection method above , all that we need to do is to see the state changes of the detection lines, no matter how complex the environment it is. In addition, the low compute cost improved the performance of the real time tracking system and provide a good foundation for the next head segmentation and detection step.

IV. HEAD SEGMENTATION AND DETECTION

After finding a moving object, the system actives the head detection subsystem. As we well know, the robust, compute time and surveillance area are very important factors for a tracking system, however they are always interacted with each other. Often a big surveillance area means large image size and more compute time; a small one will limit the system's application fields. In this system, we solve the problem by the following steps.

- Decrease the image size through down sampling.
- Skin color segmentation and create the searching space.
- Using ellipse modal to find the human head in the searching space.
- Reflect the result to the original image.

The original image defined by $I_o(x, y)$, $x = 1..H$, $y = 1...W$,the decreased image is $I_d(x', y')$

$$\begin{aligned} I_d(x', y') &= REDUCE(I_o(x, y)) \\ &= I_o(x / r, y / r), r \ge 1, x = 1..H, y = 1..W \\ &= I_d(x', y'), x' = 1..H / r, y' = 1..W / r \end{aligned} \quad (4)$$

Here r indicates the minification, x / r , y / r are integral numbers. After we find the head position $(\hat{x}', \hat{y}')$ in I_d, we can use (5) to get the real position $(\hat{x}, \hat{y})$

$$\hat{x} = \hat{x}' \cdot r \qquad \hat{y} = \hat{y}' \cdot r \quad (5)$$

Because decreasing the image size in certain scope does not change the shape of the person's head and its color, the pretreatment process above will do little influences on our detection step. In addition, it can greatly decrease the compute time and promise the system run in real-time. When we finish the pretreatment process and begin to find the person's head in

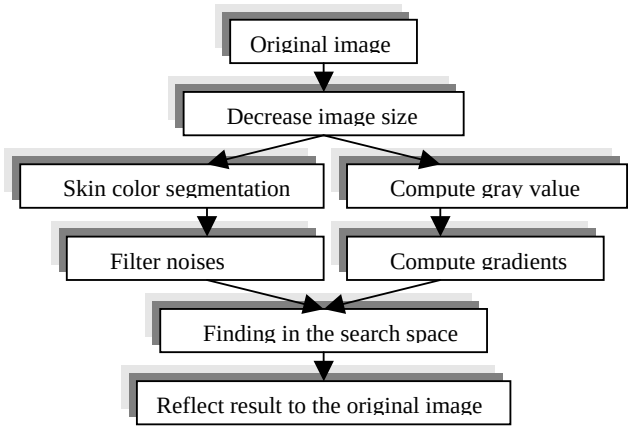


Fig. 3 Flow chat of the head segmentation and detection step.

the image, we have to consider many factors that may influence the performance of our detection algorithm, for instance, complex background, the changing illumination, target shadow, equipment noises and camera moving. In this situation, obviously it is not suitable to use only one character to detect the head, thus we combine both color and gradient information to make the final decision. As we have mentioned above, the gradient and color represent different characters of the target, and they are orthogonal to each other so that when one module fails the other one can come to its aid.

In this article we adopt the YCbCr space to detect the skin color since it is widely used in video compression standards (e.g., MPEG and JPEG) and it can separate the luminance and chrominance. R.L.Hsu[7] proposes a skin detection algorithm for color images. Based on a nonlinear color transformation, their method detects skin regions over the entire image. In the system, R. L. Hsu's color modal is being used. After the segmentation, we get the searching space I_s .

$$I_s = \{(x', y') \mid I_d(x', y') \in skincolor\} \qquad (6)$$

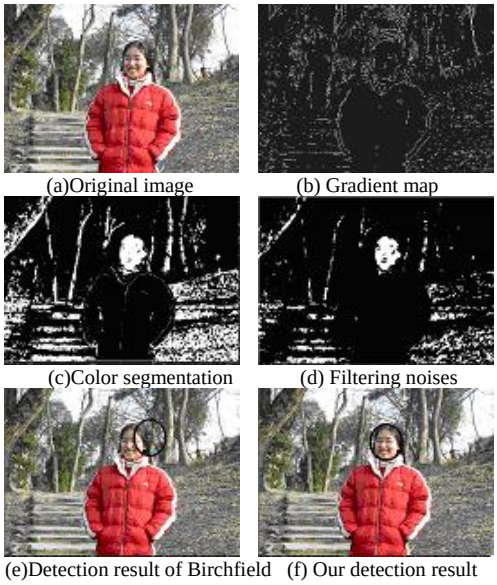
Then an elliptic modal [3,8] is used to find the person's head. The basic idea of the ellipse modal is to calculate the mean gradient of the ellipse's outline (7), we move the center of the ellipse in the searching space I_s to find the $S_{\max}$.

$$S_j = \frac{1}{N}\sum_{i=1}^{N}|g_i| \qquad (7)$$

$$S_{\max} = \max(S_j) \qquad (8)$$

Here N is the number of the pixels on the ellipse's outline, g_i is the gradient on the ith pixels. Fig. 3 shows the flow chat of the head segmentation and detection step. Fig. 4 shows the experiment results of the head segmentation and detection algorithm.

V. Active Camera Control


(a)Original image (b) Gradient map
(c)Color segmentation (d) Filtering noises
(e)Detection result of Birchfield (f) Our detection result
Fig. 4 Head segmentation and detection result.

The adequate control of the active camera is an important part of our system. After we get the coordinate of the person's head in the image, a servocontrol system is being used to control the camera turning. In the servocontrol system, the expectation is the coordinate of each image center, and the error is the subtraction of expectation and the detection result. The control signal is the rotation direction of the camera d (up,down,left and right), and the rotation time t (the rotation speed of the camera is static in our system), Fig. 5 shows the flow chat of the control system, where (x_0, y_0) is the coordinate of the image center, and $(\hat{x}, \hat{y})$ is the head detection result. Because the electromotor which enables the camera turning is a non-linear system, it is difficult for us to set up a precise modal, we build the relationship between the rotation time and the changing pixels in the image through many experiments and we use the polynomial to evaluate the all outputs. Fig.6 shows the polynomial approximate the experimental data precisely.

Then we use the function to control the camera turning, Experiment results shows that the performance of this control modal is satisfied

VI. Performance Evaluation

As we all know, robust, compute time and precision are most import factors of a common tracking system. We think .the precision in head tracking should include the precision of angle and the similar shape. Here we define a new index area matching $\phi(k)$ to evaluate both of the angle and shape precision.

$$\phi(k) = 2W_o(k)/(W_e(k) + W_f(k)) \qquad (9)$$

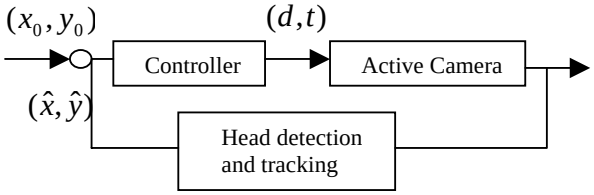


Fig. 5 The flow chat of the control system.

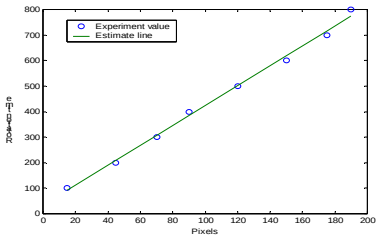


Fig. 6 Polynomial Evaluation.

Here $W_f(k)$ is the area of the human head, $W_e(k)$ is the area of the ellipse, $W_o(k)$ is the overlapped area of the both above. Only when the $W_o(k) = W_e(k) = W_f(k)$, the $\phi(k)$ reaches it's up threshold. If the $\phi(k)$ is less than 0.3, we take it as false detection. F equals to the total number detection divided by the number of false detection.

TABLE I
THE PERFORMANCE EVALUATION

Performance Index	Sign
Image size	Sp
False detection	F
Area matching	$\phi(k) \in [0,1]$
Compute time	$T(ms)$
Robust	R

VII. EXPERIMENTS AND RESULTS

Our system has been implemented on standard PC hardware (Pentium III at 733MHz) with VC6.0. The image size is 376×348 pixels. To evaluate the suitability and performance of the real-time head tracking system, we have tested it on 8 different people. Among the test candidates, there were variations in sex, skin color, position of hairline, hair style, amount of hair and head shape. The system was tested in different indoor and outdoor environments with a lot of variations in lighting conditions and background clutter. What's more, the test candidates were specially asked to do actions to try to make the tracker fail while remaining in camera's field of view and to simulate partial and complete occlusions using hands or external objects.

The sequence presented in Fig. 7 shows the active camera is tracking the person's head during walking (frame 6,18,24) and turning (frame 37,52,397,412). The subject tries to escape the tracker by performing fast lateral movements (frame 280 and 282)or hand waiving (frame 330 and 332).Observe the blurring that accompanies these movements, without affecting the tracker. Next, the candidate tries to hide behind a newspaper, but only when the head is completely occluded the tracker fails(frame 568,569,570). However, once the occlusion is terminated, the head is immediately recovered.Fig.8 shows samples from another tracking sequence demonstrating the detection and tracking of the subject in outdoor environment .The people is walking (frame 26,29,117,162,245)and turning (frame42,43,51,294) in different distance.

Finally, the sequence from Fig.9 shows the capability of the tracker to handle scale changes. Note that the person's hair style and head shape is different with the one in Fig. 6 and Fig. 7.Table 2 is the computing times under different minifications. Table 3 shows the performance valuation results under different conditions above.

TABLE II
COMPUTING TIMES

Performance	Head tracking system			
Sp	376×348 (Pixels)			
Disturbing	No		Yes	
Distance (m)	1~5	6~10	1~5	6~10
F	5%	7%	10%	15%
$\phi(k)$	0.85	0.89	0.69	0.75
T (ms)	21.5	21.5	21.5	21.5
Robust	Satisfied			

TABLE III
PERFORMANCE EVALUATION

r	Plus times	multiply times
1.0	8.4379×10^8	2.8297×10^8
0.5	68895782	24925325
0.25	5625365	2195519
0.125	629779	278780

VIII. CONCLUSION

We have presented a real-time system for the detection and tracking of human head with an active camera. The core component of the system is based on a new head detection algorithm that involves low computational cost. Experimental results demonstrated the robust of the system to scale variations, fast subject movements and camera saccades, partial occlusion, out-of-plane rotations, and complex background.

There are still rooms to improve the proposed approach. In particular, our future research will be focused on addressing the following issues. Firstly, when the background color is similar to the skin color and its gradient is high, How to keep on tracking the object in real-time is still a challenge. Also, if there is more than one person in the view field of the camera, it is hard for our method to decide which one is the target.

To solve those problems above, we are now focusing on using particle filter [9] to track multiple people with multiple cameras.

REFERENCES

[1] Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, and Hasegawa, "A System for Video Surveillance and Monitoring: VSAM Final Report," *Technical report CMU-RI-TR-00-12*, Robotics Institute, Carnegie Mellon University, May, 2000.

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

[3] S. Birchfield, "An Elliptical Head Tracker," *In 31st Asilomar Conference on Signals, Systems, and Computers*,pp.1710-1714, Nov 1997.

[4] M.Pardas, E.Sayrol, "A new approach to tracking with active contours," *Proc. Int. Conf. on Image Processing*, Vol.2, pp.259-262,2000.

[5] S.J.Mckenna, Y.Raja and S.Gong, "Object tracking using adaptive colour mixture models," *Lecture Notes in Computer Science*,1998.

[6] J.Ahlberg, "Using the active appearance algorithm for face and facial feature tracking,"*Proc.2nd Int. Workshop Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems* , pp.68-72,2001.

[7] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain ,"Face detection in color images,"*IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706 .

[8] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," *in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Santa Barbara, CA*, pp. 232-237,1998.

[9] A.Doucet.N.Freitas,N.Gordon, "Sequentional Monte-Carlo Methods in Practice," *Springer*,2001

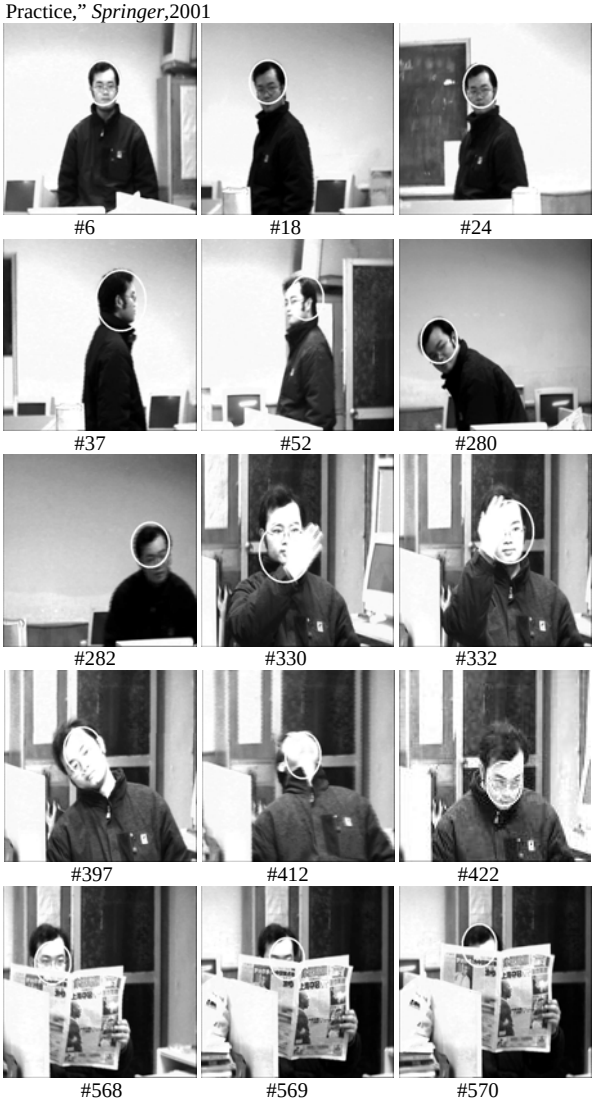


Fig. 7 People sequence I.

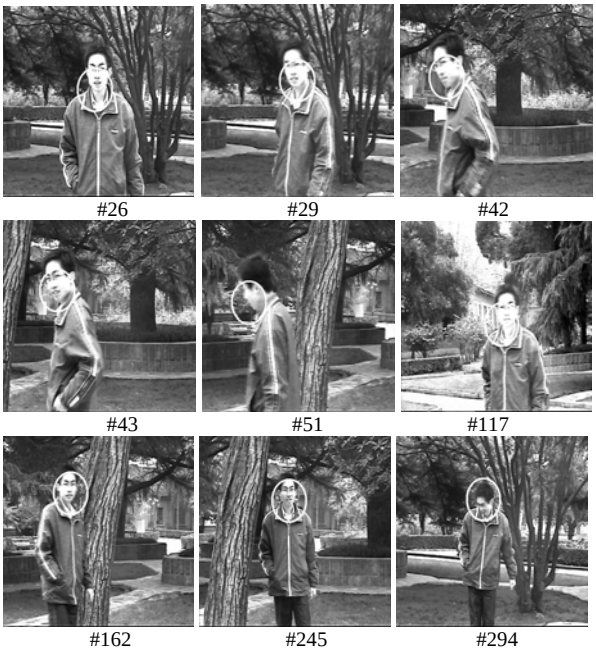


Fig. 8 People sequence II.

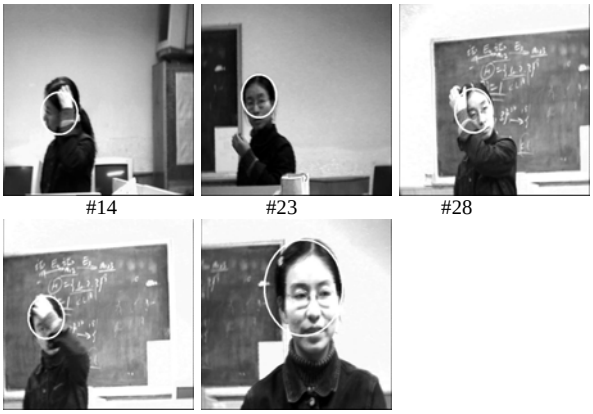


Fig. 9 People sequence III.