FACE TRACKING USING TWO COOPORATIVE STATIC AND MOVING CAMERAS

P. Amnuaykanjanasin and S. Aramvith[*], T.H. Chalidabhongse[†]

[*]Department of Electrical Engineering
Chulalongkorn University
Bangkok 10330 Thailand
Tel: +66-2218-6909
E-mail: Supavadee.A@chula.ac.th

[†]Faculty of Information Technology
King Mongkut's Institute of Technology
Ladkrabang
Bangkok 10520 Thailand
Tel: +66-2737-2551 Ext.526
E-mail: thanarat@it.kmitl.ac.th

ABSTRACT

In this paper, we present a new stereo approach for tracking human face by using only two cameras in system. One pan-tilt camera is used for tracking person focused on face. One static camera cooperate with pan-tilt camera are used as a stereo system to estimate face 3D position. We propose to update relative position between cameras to reflect camera moving and the change of relative position. Experimental results shows that our proposed system is able to track one person in camera viewing and can estimate the 3D moving path of interesting person.

1. INTRODUCTION

Face detection and tracking are widely interesting research topic and can be applied to many applications such as human surveillance, facial gesture recognition, sign language translation, and human-computer interaction. There have been two main approaches in solving this problem. Early approaches concern with 2D segmentation and tracking; not many features beside 2D shape can be used. This limits the capability of visual modeling from images. Later, many approaches presently more interest in working in 3D which provides more visual cues for detecting and tracking position of the interesting targets.

As the human tracking using visual cues has been interested by many researchers, the problem is extended to ability to collect and analyze the human motion and use it to detect and recognize some interesting events. Temporal and details information of the target is needed. Currently, many researches apply active camera system such as moving and pan-tilt-zoom (PTZ) camera in detecting and tracking interesting target to follow and deliver a fine view of it. Ser-Nam Lim et al. [1] present a wide area surveillance system using multiple co-operative cameras that can be zoomed in to and follow a target. S. Tsuruoka et al. [2] work with two active camera systems trying to understand what is on the lecture board as well as understand the lecturer gesture in distance learning system. S. Bahadori et al. [3] present a surveillance system based on 3D reconstruction of a museum environment. The system signal an alarm when detecting a person located close to certain museum areas.

There exists a few works which have addressed the issue of 3D in PTZ camera systems. Among these, it includes A. Hampapur et al. [4] which propose the person identification system using multi-scale imaging. H. Hongo et al. [5] apply a set of cooperative fixed cameras and PTZ cameras for face and hand gesture recognition. Both systems consist of four cameras, two fixed stereo cameras for depth recovering and two PTZ for 3D tracking of two objects.

Our work is focused on building a system that can do both estimating 3D position of a person's face and tracking its 3D motion using two PTZ cameras.

The organization of this paper is as follows: Section 2 describes the overall architecture of the proposed system. Section 3 presents the stereo matching technique used in this system. Section 4 proposes a method of managing a moving stereo camera. Section 5 shows the process for camera control. The experimental results are shown in Section 6. Section 7 concludes the paper and addresses future works.

2. SYSTEM OVERVIEW

Figure 1 shows the camera setup of the two Sony EVI-D100 cameras for the proposed system. The two cameras have overlapping fields of view and are used for stereo triangulation to estimate 3D position of person. First camera is assigned as a fixed static camera. It is used to detect the person's face and set as a reference face blob to match with candidate face blobs that can detect in the second camera. The second camera is set as a pan-tilt camera that follows the target face while tracking. Both

0-7803-9332-5/05/$20.00 ©2005 IEEE

cameras are calibrated using Bouguet's camera calibration toolbox [8].

The overall architecture shows in Figure 2. In stereo, there are two key steps; stereo matching and triangulation. The first step needs to find correspondence between cameras by match feature points between the left and right images. Our work use human's skin color as a cue for matching. In addition, for static camera, we also use background subtraction to accurately detect the real face blob. The second step, which is the triangulation, is the process of computing to obtain the estimate of 3D position of the person's face. The 3D face position is then used by the active camera control to assign the right pan-tilt parameters to the camera in order to follow the target.

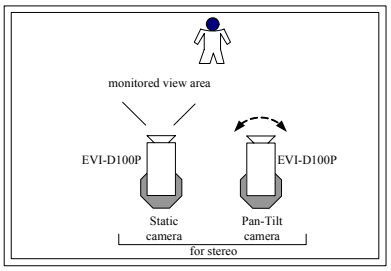


Figure 1. System configuration

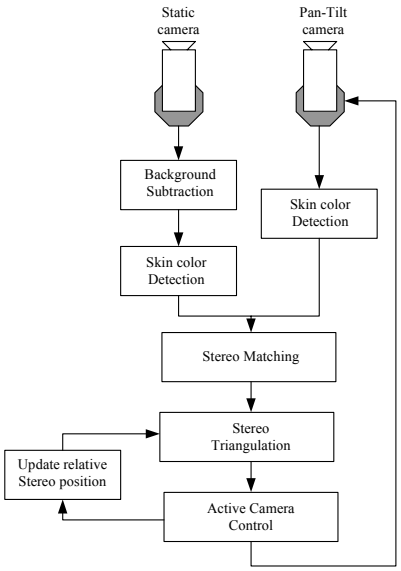


Figure 2. System architecture

3. STEREO MATCHING

To find a corresponding point for 3D position estimation, we use color cue to detect face region from the stereo camera images. Various methods, that are used to detect skin, usually differ in reference color space. Experimental study [6, 7] has suggested that the YCbCr color space is a suitable choice for skin color modeling. The YCbCr color space consists of the brightness component (Y) and color values component (Cb and Cr) so it is capable of detect skin region in changeable brightness environment. We employ method proposed in [6,7] to extract skin color region. The method models skin color with elliptical function on CbCr space. The Y component is discarded to make it invariance to brightness. Figure 3 shows the CbCr color distribution of the human skin color that fits elliptical model.

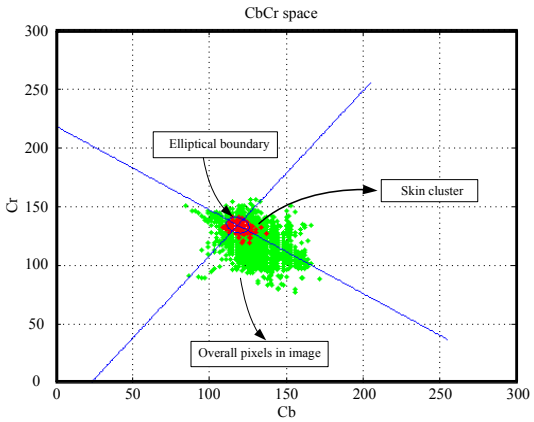


Figure 3. Elliptical model and color distribution

3.1. Face detection in static camera

In static camera, in addition to the skin color, we also apply the background subtraction technique to locate face region. We first subtract image from the background model to discard any background including the skin-color-like background. In this work, we assume that the human's clothes are different from skin color. Figure 4 (a) shows the input image. Figure 4 (b) shows the result of background subtraction and Figure 4 (c) is the final result after applying the skin color detection on Figure 4 (b).

3.2. Blobs candidate search in active camera

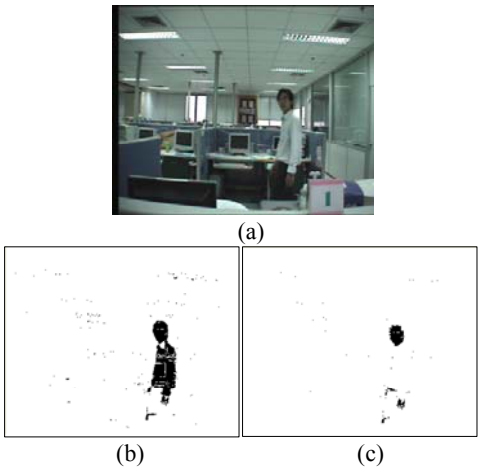
In active camera, the camera's field of view instantly changes by the pan and tilt motion. Basic background subtraction does not work, so we use only skin color detection to detect the face. The result of face detect still include many noises. Thus, to reduce such noises, we apply the affine transform in Eq. (1-2) to calculate the approximate disparity between camera both horizontal and vertical direction when camera pan and tilt.

$$u' = a_0 u + b_0 v + c_0 + du \quad (1)$$

$$v' = a_1 u + b_1 v + c_1 + dv \quad (2)$$

where (u',v') is the estimate position of face blob in active camera. (u,v) is the center of gravity of the face blob in static camera. a_0, a_1, b_0, b_1,c_0 and c_1 are affine parameters that can pre-calculate from both image in the

initial position of two camera before active camera will moved. *du* and *dv* are estimate shift pixels that appropriate to pan and tilt angle of each frame.



(a) (b) (c)

Figure 4. (a) input image (b) foreground result of background subtraction (c) detected face after add skin color detection

4. STEREO SYSTEM MANAGEMENT

The second step in stereo is stereo triangulation. Normally most stereo applications use two static cameras. The relative position between cameras is fixed. This is not the case in our proposed system, since we use one active camera cooperate with the other one static camera. So the relative position between cameras will be changed according to the camera movement. Thus we have to update the cameras' matrix to reflect the camera moving and the change of relative position.

4.1. Update relative matrix between cameras

In the calibrated environment, the two cameras coordinate systems are related by a rotation matrix (***R***) and a translation matrix (***T***). Eq. 3 shows the relation when using the static camera as a reference.

$$X' = R^T X + T \tag{3}$$

X' : active camera coordinate systems
X : static camera coordinate systems

We assume that when active camera changes the field of view, the projection center does not change position so a translation matrix has the same value as in the initial position. A rotation matrix will be changed and need update procedure to find new relative position after active camera moved. Figure 5 shows camera coordinate relation when active camera coordinate changed by pan angle (ϕ) and tilt angle (ψ). Rotation matrix can be changed in 2 cases. First, rotation matrix for pan direction (R_{pan}) is updated by Eq. 4. Second, rotation matrix for tilt direction (R_{tilt}) is updated by Eq. 5. So the new relative position matrix between cameras ($R_{new}(t+1)$)can update as shown in Eq. 6 and ready to calculate the object depth in next time.

$$R_{pan} = \begin{bmatrix} \cos(\phi) & 0 & -\sin(\phi) \\ 0 & 1 & 0 \\ \sin(\phi) & 0 & \cos(\phi) \end{bmatrix} \tag{4}$$

$$R_{tilt} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\psi) & -\sin(\psi) \\ 0 & \sin(\psi) & \cos(\psi) \end{bmatrix} \tag{5}$$

$$R_{new}(t+1) = R(t) \cdot R_{pan} \cdot R_{tilt} \tag{6}$$

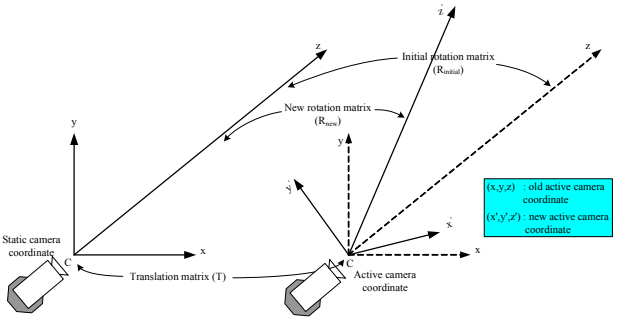


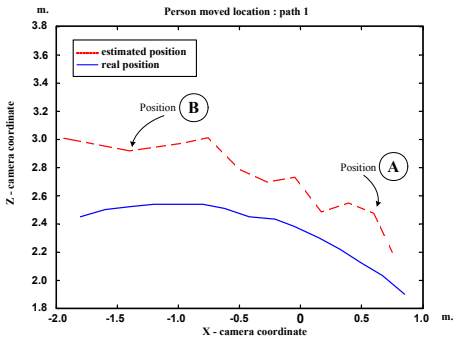
Figure 5. Relative position between cameras when active camera moved

5. ACTIVE CAMERA CONTROL

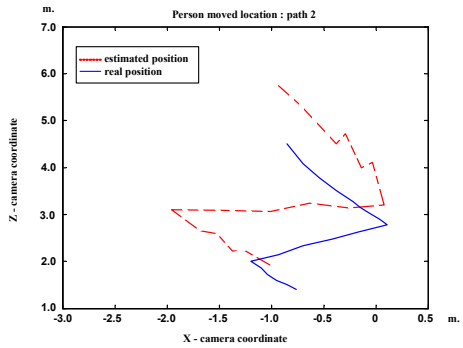
This module is in charge of controlling the pan and tilt parameters of the active camera. The objective of camera control is to maintain the person's face being tracked within center of its camera view. We use 3D object position that get from triangulation for steering the pan-tilt active camera to position the detected face location at the center of the active camera image. The pan and tilt angle are calculated from 3D estimate location

6. EXPERIMENTAL RESULTS

We first modeled the elliptical skin color in CbCr space using a set of 20 person's face image blob. The cameras were then calibrated to obtain camera matrix as well as the relative location between cameras. We marked two trajectory paths on the ground and let the subject walk along. One path was a curve; while the other was a zigzag. Figure 6 shows trajectories of tracked subject (red) versus the groundtruth (blue) plotted on XZ plane (top view). Figure 7 shows the corresponding images of two different frames captured while the subject was moving.



(a)



(b)

Figure 6. Trajectories of tracked moving subject (red) compared to the groundtruth (blue) plotted on XZ plane for both curve path (a) and zigzag path (b).

The 3D Euclidean distance error for the curve path is 0.369 meter while for the zigzag path is 0.836 meter. The results show some location errors however the patterns of the path are quite corresponding to the groundtruth.

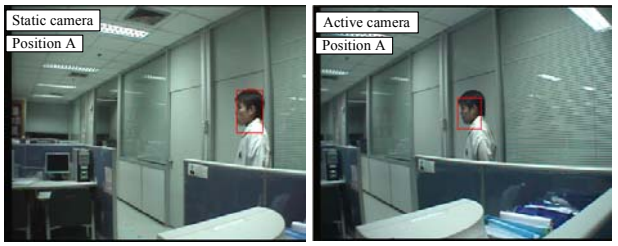
7. CONCLUSIONS AND FUTURE WORK

We have presented a new system that can track a person focus on face and can estimate location of a person by using two stereo cameras, one for static camera and another for active camera. Simulation results shows that our system can track person for fine view and can evaluate the 3D person position that has precise 3D coordinate direction but has some error in the same direction.

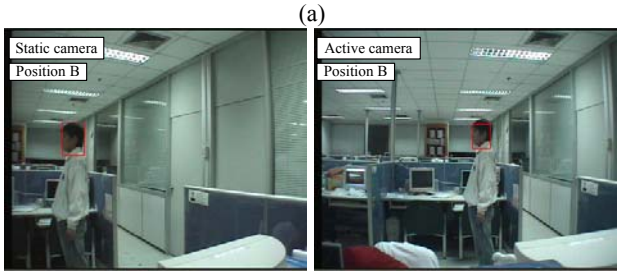
Currently, we are working on camera zoom that can deliver a multi-scale face image and can be useful to many applications such as face gesture recognition. The multiple cues approach will also be pursued to improve the corresponding points matching.

8. ACKNOWLEDGEMENT

This work is supported in part by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University, Thailand.



(a)



(b)

Figure 7. Corresponding images of two different frames (a) and (b) captured while the subject was moving. Left images are taken from static camera. Right images are from the moving camera.

9. REFERENCES

[1] S.-N. Lim, A. Elgammal, and L. S. Davis, "Image-Based Pan-Tilt Camera Control in a Multi-Camera Surveillance", in *Proceedings of IEEE International Conference on Multimedia and Expo*, Maryland, pp. I-645-8, Jul 2003.

[2] S. Tsuruoka, T. Yamaguchi, K. Kato, T. Yoshikawa and T. Sninogi, "A Camera Control Based on Fuzzy Behavior Recognition of Lecturer for Distance Lecture", in *IEEE International Conference on Fuzzy Systems*, Hawaii, pp. 940-943, Jan 2001.

[3] S. Bahadori, and L. Iocchi, "A Stereo Vision System for 3D Reconstruction and Semi-Automatic Surveillance of Museum Areas", in *Workshop on Intelligenza Artificiale per I Beni Culturali*, Pisa, 2003.

[4] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle, "Face Cataloger : Multi-Scale Imaging for Relating Identity to Location", in *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance*, Florida, pp. 13-20, Jul 2003.

[5] H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, "Focus of Attention for Face and Hand Gesture Recognition Using Multiple Cameras", in *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, pp. 156-161, Mar 2000.

[6] N. Soontranon, S. Aramvith, and T. H. Chalidabhongse, "Face and Hand Localization and Tracking for Sign Language Recognition", in *International Symposium on Communication and Information Technologies(ISCIT'04)*, Sapporo, Oct 2004.

[7] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face Detection in Color Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 696-706, May 2002.

[8] Camera Calibration Toolbox for Matlab. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/