Real time monitor system based on dual-camera cooperative fusion

Jilin Hu, Shiqiang Hu*, Zhuojin Sun
School of Aeronautics and Astronautics
Shanghai Jiao Tong University
Shanghai, China
hujilin1229@sjtu.edu.cn, sqhu@sjtu.edu.cn

Abstract-One of the main aims for modern video surveillance is to capture clear imagery of Region of Interest (ROI) in large-scale. However, it is hard to achieve the goal by zooming when the ROI is far from the surveillance system. A station-active dual camera (SADC) system is designed to acquire the detailed and overall information of object at distance. The target can be detected and tracked in the wide Field Of View (FOV) of the static camera, based on the method of modified codebook and Tracking-Learning-Detection (TLD) separately. Then the control part of the system calculates the correct position, which directs the Pan-Tilt-Zoom (PTZ) camera to gaze at the target and get a high resolution image of the detected candidate. In the system, some results are obtained in the experiments of indoor and outdoor setting, which is applicable in real time surveillance. From the comparison, it is found that the TLD have an advantage over codebook on robustness to illumination variation and the scene of multi motion objects.

Keywords-TLD; Codebook; high resolution image; fusion of stationary and PTZ camera; video surveillance

I. INTRODUCTION

As a result of increasing emphasis on modern society security, the video surveillance is widely applied in public places, such as banks, airports, stores and parking lots. Therefore, a growing interest is devoted to getting a high resolution image of the region of interest (ROI).

Extensive researches have been conducted in video surveillance and a static-active dual camera (SADC) system [1-6] has been introduced. By the aid of fusing stationary and steerable camera, the effective view distance could be increased through the use of wide angle camera. Then the high-definition image of the detected candidate can be acquired from the Pan-Tilt-Zoom (PTZ) camera.

In [2], PTZ camera zoomed in the target to acquire the detailed information. However, the features of the target in the PTZ camera haven’t been analyzed in fellow-up process. In the paper [3, 4], it assumes that the position of the object has been obtained. Then the steerable camera receives the command to rotate and reaches the right position. An omni-directional camera has been adopted to provide the wide FOV in [5] which mainly sets up a human face tracking system to increase the effective view at distance and provide a high-definition human face by the aid of fusing the omni-directional and PTZ cameras. The main purpose of [6] is an automation traffic sign recognition system, accomplished by the hybrid camera system combined with appropriate algorithm, while the steerable camera does not zoom in the traffic sign, thus the precision of recognition is decreasing.

In this paper, a monitor system, consisting object detection, object tracking and the zoom controlling of the PTZ camera, is proposed. In this system, the detailed information of the object can be acquired from the PTZ camera when the target is detected in the wide FOV. The main clue can be interpreted as: Firstly, the object candidate is detected and tracked in the wide Field Of View (FOV) of the static camera, based on the method of modified codebook and Tracking-Learning-Detection (TLD). Then, by the method of looking up table (LUT), the control part of the system is adjusted to the correct position. Therefore, a high resolution image of the detected candidate was obtained by the Pan-Tilt-Zoom (PTZ) camera.

In this paper, the SADC system combined with appropriate algorithm is tested in real time video surveillance. The paper is organized as follows. In section II the calibration used in the stationary camera is presented. Section III presents the method of TLD and the modified codebook algorithm. In section IV, the
cooperative control strategies are explained. The experiments and conclusions are described in section V and VI, respectively.

II. Calibration

The goal of PTZ camera calibration is to figure out the relationship between the stationary and PTZ cameras, which the static camera directs the PTZ camera to focus on the selected target. For this purpose, the pose of the PTZ camera for target fixation is maintained by continuously estimating its required panning and tilting angles. And here, the method of Looking Up Table (LUT) [8, 9] is adopted. By using the calibrated table, matrix operation can be replaced for finding the object’s position in the object images. Subsequently, the calibration procedure will be elaborated in this section.

Establishing the correspondence of image points is the main purpose of image registration. In our system, two cameras are installed in a fixed position, so that some properties of the geometry, the point ordering, the homographic transformation, etc., are well defined in advance. Through the iterative approach from a coarse to fine and a refinement stage, the matching pairs generated in a coarser stage would satisfy the new constraints in the fine stage. In this way, a new and small candidate region is generated to find the correct matching pairs iteratively.

A. The Principle of Pinhole Camera

The ideal geometry of the pinhole camera is displayed in Fig. 2, where u-axis and v-axis parallel to the X-axis and Y-axis respectively, the optical axis aligned with the Z-axis and f represents the focal length of this camera.

![Fig. 2 the geometry of a pinhole camera](image)

The homogeneous coordinates of a 3-D point relative to the camera frame and homogeneous coordinates of its image expressed in terms of pixels is depicted in (1).

\[
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix} = 
\begin{bmatrix}
  s_x f & 0 & u_0 \\
  0 & s_y f & v_0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  z
\end{bmatrix}
\] (1)

Where u and v are actual image coordinates in pixels, (x, y, z) are units coordinates of scene point and sx, sy are the scales which transform from units coordinate to pixel coordinate of x and y axes respectively. (u0, v0) are the coordinates (in pixels) of the principal point relative to the image reference frame, or where the Z-axis intersects the image plane.

However, the scene point’s coordinate can’t be acquired for the missing of z when the object maps to the image. Therefore, the theory of establishing dual-camera to acquire 3-D coordinate has been put forward. Here, the result of the dual camera system is listed in (2).

\[
\begin{bmatrix}
  x_2 \\
  y_2 \\
  z_2
\end{bmatrix} = R_{12} \cdot \begin{bmatrix}
  x_1 \\
  y_1 \\
  z_1
\end{bmatrix} + T_{12}
\] (2)

![Fig. 3 The Principle of dual-camera](image)

The parameters in (2), as demonstrated in Fig. 3, where \((x_1, y_1, z_1)\) is the homogeneous units coordinates of point P relative to the left camera frame and \((x_2, y_2, z_2)\) is relative to the right camera frame. \(R_{12}, T_{12}\) are the rotate matrix and the translation vector between the two cameras respectively.

In Fig. 3, point P is the scene object point, point \(P_l\) and \(P_r\) are the point P map onto the left and right image plane respectively. \(O_1\) denotes the focal point of the left camera and \(O_2\) denotes the right. The plane passing through the point P, \(O_1\) and \(O_2\) is called epipolar plane. The intersections of the line \(O_1O_2\) and image plane are called epipolar points, which in Fig. 3 are named \(e_1\) and \(e_2\).

B. The Analysis of Position Changing

In the analysis above, it is assumed that the object be static. The pairs of matching points between the two images from respective camera do exist. What is applied in this paper, however, is not just the static state. The system proposed here is to monitor arbitrary motion status of the selected object, whether it be moving or standing still.

Thus, when the selected object moves in the line of the focal point and the original position, the coordinates relative to the static camera’s image reference frame remains unchanged. However, the coordinates of the object relative to the PTZ camera frame is altered. Whether the corresponding relation can still be satisfied or not under this condition will be discussed in this part.

Fig. 4 depicts the geometry model of the image structure of dual-camera. The principle of this system can be clearly observed. The object is at the point of P, whose units coordinates
relative to the static camera frame is \((x, y, z)\). And its pixel coordinates relative to the static camera’s image reference frame is \(P_s(u, v)\). Here relative to the static camera’s image is \(P_s\). It is assumed that \(P_s\) is principal point relative to the image reference frame. In other words, the paned and titled angle \((\alpha, \beta)\) of the PTZ camera is the matching point pairs of \(P_s(u, v)\).

When the object moves from \(P=(x, y, z)\) to \(P'=(x+\Delta x, y+\Delta y, z+\Delta z)\), the pixel coordinates relative to the static camera frame is still \(P_s(u, v)\) which satisfies (3), where \(f_s\) is the focal length of static camera. \(s_x, s_y\) are the scales which transform from units coordinate to pixel coordinate of \(x\) and \(y\) axes respectively.

\[
(z + \Delta z) \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{s_x f_s}{u_0} & 0 & u_0 \\ 0 & \frac{s_y f_s}{v_0} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ z + \Delta z \end{bmatrix}
\] (3)

Finally, the relationship between \(\Delta x, \Delta y\) and \(\Delta z\) is displayed in (5), (6). 

\[
\Delta x = \frac{\Delta z \cdot (u-u_0)}{f_s}
\]

\[
\Delta y = \frac{\Delta z \cdot (v-v_0)}{f_s}
\]

It can be presented in matrix form as shown in (7).

\[
[\Delta x \ \Delta y \ \Delta z]^T = \Delta z \cdot \begin{bmatrix} (u-u_0) \\ (v-v_0) \\ 1 \end{bmatrix}^T
\] (7)

The homogeneous coordinates of point \(P'\) relative to the PTZ camera frame is \(P''\). \(R_{sp}, T_{sp}\) are assumed as the rotate matrix and translation vector respectively. The expression can be seen in (8).

\[
P'' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_{sp} \cdot P' + T_{sp} = R_{sp} \cdot \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ z + \Delta z \end{bmatrix}
\] (8)

Then the coordinates of point \(P'\) relative to the PTZ camera frame can be calculated as in (9). 

\[
\begin{bmatrix} f_s & 0 & u_0 \\ sf_s & v_0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_{sp} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + T_{sp}
\]

\[
= \begin{bmatrix} s_x f_s & 0 & u_0 \\ 0 & s_y f_s & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ v \end{bmatrix} + T_{sp}
\] (9)

Apply (4) into (6), then it can be shown as in (10).

\[
\left\{ \begin{array}{l}
\frac{u_0}{v_0} \Delta z + \left[ \frac{s_x f_s}{u_0} \ 0 \ \frac{s_y f_s}{v_0} \right] \cdot \begin{bmatrix} u' \\ v' \end{bmatrix} \\
\frac{v_0}{u_0} \Delta z + \left[ \frac{s_x f_s}{u_0} \ 0 \ \frac{s_y f_s}{v_0} \right] \cdot \begin{bmatrix} u' \\ v' \end{bmatrix}
\end{array} \right.
\]

(10)

Where, the variables are all constant besides \(\frac{\Delta z}{z}\). Therefore, the coordinates of point \(P'\) relative to the static camera’s image reference frame is still \(P_s(u, v)\), the variance of coordinates relative to the static camera frame has proportional relation with \(\Delta z\) only, see in (7), so as the variance of coordinates relative to the PTZ camera’s image reference frame see in (10).

Comparing with the distance between the target of interest and the camera, \(\Delta z\), the change of the object’s coordinate in \(z\) direction, is so small that it can be neglected. Therefore, the one-to-one mapping relation between the coordinate of stationary camera, \(P_s(u, v)\) and the absolute position \((\alpha, \beta)\) of the PTZ camera when its center is on the object could be established in (11).

\[
[u \ v]^T \leftrightarrow (\alpha, \beta)
\]

C. Set up the PTZ table

The goal of PTZ table construction is to trigger and rotate the object camera to locate an object at the center. In the other words, assuming that we have known the target’s position of \(M(x, y)\) in the static camera, so we panned a degree of \(\alpha\) and titled a degree of \(\beta\) to make the center of the PTZ camera at the same position of the object. Through this way, we got the corresponding point pairs of the two cameras and recorded it. Here is the procedure of construction of the PTZ table [10].
• Selecting a region (550x200) in the wide angle camera to be ROI, and in this area the PTZ camera can receive a command to steer by the LUT.

• Then we separated the width and height of the region in a selected interval and we got a grid. Thus, the coordinates of the pixel at the intersection point of the grid can be listed as follows: \( M_{11}(x_1, y_1), M_{12}(x_1, y_2), M_{21}(x_2, y_1), M_{22}(x_2, y_2) \), etc.

• For a given point \( M_{11} \), we panned a degree of \( \alpha \) and tilted a degree of \( \beta \) to make the center of the PTZ camera at the same position of the object. And recorded the corresponding pair points i.e. \( L[M_{11}(x_1, y_1)] = (\alpha, \beta)_{11} \).

• For the rest intersection of the grid, repeat the third operation above, and store all the pair points in the formulate of \( L[M_{ij}(x_i, y_j)] = (\alpha, \beta)_{ij} \).

• For the other non-intersections of the grid, making use of the minimum distance four neighbor points (e.g. \( M_{11}, M_{12}, M_{21}, M_{22} \)), which can be computed through the Double linear interpolation formula (see in (12)).

\[
(\alpha, \beta)_s = \frac{1}{(x_2-x)(y_2-y)} [L(M_{11})(x_2-x)(y_2-y) + L(M_{12})(x_2-x)(y_1-y) + L(M_{21})(x_1-x)(y_2-y) + L(M_{22})(x_1-x)(y-y_1)]
\]  

(12)

• Collected the data from the third step to fifth step, we get a LUT for anywhere in the ROI, the PTZ camera will adapt its position to the target’s point.

III. THE DETECTION METHOD

The destination of this system is that the wide angle camera can capture an image of the wide FOV, and for a certain region, the PTZ camera can get a high resolution image. So the wide angle camera is applied to find the position of the target in the detection step. In this paper, the background algorithm of modified Codebook and the method of TLD [7] are applied in the stationary camera to detect and track the object, and both of the results are compared in section VI.

A. Modified Codebook

The codebook (CB) background subtraction algorithm [11] applied in this paper is intended to sample values over long times, without making parametric assumptions. The main features of the algorithm are:

1) an adaptive and compact background model that can capture structural background motion over a long period of time under limited memory. This allows us to encode moving backgrounds or multiple changing backgrounds;

2) the capability of coping with local and global illumination changes;

3) unconstrained training that allows moving foreground objects in the scene during the initial training period;

4) layered modeling and detection allowing us to have multiple layers of background representing different background layers.

The flowchart of using modified codebook algorithm in this system is illustrated in Fig. 5.

B. The Method of TLD

The TLD method, whose learning part is analyzed in [12], was designed for long-term tracking of arbitrary objects in unconstrained environments. The target is tracked and simultaneously learned in order to build a detector that supports the tracker once it fails. The detector is built upon the information from the first frame as well as the information provided by the tracker.

The block diagram of TLD is illustrated in Fig. 6. Every frame is processed by a tracker and a detector and their outputs are passed to an integrator which estimates the object location. The object location as well as the detections and the trajectory are analyzed by a learning block which generates training data for the detector (update of the tracker is not considered in this implementation). The flowchart of using TLD in this system is shown in Fig. 7.
IV. COOPERATIVE CONTROL STRATEGY

A. Zoom Control

One of the crucial intentions in this system is to obtain high resolution images from PTZ camera. Consequently its zoom can be controlled by directly setting the desired FOV for the PTZ camera. This was determined as follows:

\[ γ = \min \left( \frac{W}{W_m \mu}, \frac{H}{H_m \mu} \right) \]  

(13)

Where, W and H are the width and height of the images captured by both of the cameras. \( \mu \) is the size ratio of the object in the PTZ camera and the same object in the same position of the static camera. \( W_m \) and \( H_m \) are the width and height of the target in the stationary camera.

B. The Control Strategy of Different Status

The desired zoom for the object to fulfilling the FOV of the PTZ camera is displayed in (13). Under this condition, if the center of PTZ camera’s ground plane is at the object’s center, a high definition image can definitely be obtained. However, it is easy for the object to move outside of the view. Besides, there is some latency existing in the control unit of the PTZ camera, which increases the difficulty in keeping the candidate object in the FOV with a high zoom. Therefore a zoom control strategy according to the move status is proposed in this paper. This strategy will be discussed in the following two parts.

- **Motion**
  
  The paned and tilted position of the PTZ camera is frequently adjusted due to the motion of the candidate object. Therefore, 0.25\( \gamma \) is selected as the zoom of the PTZ camera when the target is in motion i.e. the object occupies about one-quarter of the view.

- **Static**

  When the object is static, the PTZ camera has enough time to implement the command of paned and titled angle to target at the object and zoom in to get high resolution images. In this mode, the zoom selected is 0.75\( \gamma \) which the candidate object occupies about three-quarter of the view.

C. The Overall Cooperative Control Strategy

At first, the candidate object is manually selected in the frame of stationary camera. Each resultant object is tracked across frames, and estimation of the object’s area, width, height, bounding box are made. Once such parameters are detected, the PTZ camera will be called to zoom in the target, and high resolution images can be captured.

The description of the control strategy can be shown as follows:

1) The PTZ camera tracks the target according to the location command received from the wide angle camera, and zooms in the target whose zoom is about 0.25 \( \gamma \).

2) The algorithm used to decide move statue is listed below.

   a) The coordinates of the latest two bounding box (\( p_1, p_2 \)) could be obtained in the detection algorithm (TLD and Codebook). For each \( p_i=(x_i, y_i, w_i, h_i) \), \((x_o, y_o)\) is the coordinate of the bounding box’s top left point. \( w_i \) and \( h_i \) are the width and height of the bounding box, respectively.

   b) For each frame, \( p_i \) of current frame and the last frame were compared. For the current frame, the parameter defined as \( p_i \), and \( p_{i-1} \) for the later one.

   c) If \( |x_1-x_2|<0.05w_2, |y_1-y_2|<0.05h_2, |w_1-w_2|<0.05w_2, |h_1-h_2|<0.05h_2 \), the four conditions were satisfied. Then the object is static. Otherwise, the object is in motion.

   d) When the object is static, the PTZ camera zooms in the target, with the zoom to be 0.75 \( \gamma \).

3) Repeat the step from the first step.

V. EXPERIMENTS

The configuration of this experiment is depicted in Fig. 1. This configuration equipped with a stationary camera (camera 1) which is a SONY 453P digital video camera with a FUJINON FY28V8A focal lens adding onto it and a second camera (camera 2) which is a Sony EVI-D70 PTZ camera. Camera 2 has an 18X optical zoom and wide pan and tilt ranges (340 degree pan, 120 degree tilt).

These two cameras are communicated over a serial communication on the platform of HP xw6400 Workstation. With the help of VS2008 and OpenCV, some results have been gotten.

The experiment is conducted in the indoor and outdoor setting respectively. Consequently, the experiment results are presented in this section orderly.

A. Indoor application

The system was tested in an offline setting. Here, a user interface was needed to manually select an object of interest. Once an object was selected, the steerable camera would adjust
pan, tilt, and zoom settings in real time to keep the object at the center of the image when it is moving. The zoom was adjusted so that the longest dimension of the bounding box surrounding the object fills the frame. Fig. 8 depicted two sample corresponding images by the two methods in an indoor setting.

Fig. 8 images captured by the two methods in an indoor setting: a. image taken from PTZ camera of using codebook; b. image taken from stationary camera of using codebook; c. image taken from PTZ camera of using TLD; d. image taken from stationary camera of using TLD.

Based on Fig. 8, the two sets of images are both desirable, however, by observing the video of both methods, a conclusion can be drawn that the method of Codebook is sensitive to the illumination variation while the TLD is robust to it. In other words, when the illumination changes dramatically, the Codebook method becomes invalid while the TLD can still be working.

B. Outdoor Application

The system has also been tested in an outdoor setting. The cameras were put in the front door of our school to monitor the courtyard before it.

A 19s video has been recorded using the method of Codebook in this system. Fig. 9 depicted the two extremes of this method, where the shortage of Codebook is figured out.

The experiment is done on a sunny afternoon, the illumination is so strong that the weak robustness of the codebook to illumination, mentioned in last part, can be ignored. The problem appeared in Fig. 9(c) is due to the tree behind the person is shaking fiercely. Therefore, the change is detected by the codebook and the position command is sent to the PTZ camera, which lead to this result.

Owing to the cloudy weather when we test the method of TLD, the ability of resisting to the low illumination of the system is tested in this setting. Under the low illumination, the color of the person’s clothing in this setting is similar to trees in the background. Favorable results are displayed in Fig. 10.

Fig. 9 images captured by the Codebook in an outdoor setting: a. ideal image from PTZ camera; b. image taken from the stationary camera corresponding to the image of (a); c. wrong detected image taken from PTZ camera; d. the image from the stationary camera corresponding to the image of (c)

Fig. 10 images captured by the TLD in an outdoor setting: a. image captured from the PTZ camera when the object is far away; b. image captured from the stationary camera corresponding to (a); c. image captured from the PTZ camera when a person is passing in front of the object; d. image corresponding to (c) from the stationary camera

As is shown in the two sets of results, by comparing them, some conclusions can be drawn.

- Some desirable results can be achieved by both of the two methods which are applicable in reality surveillance.
The method of Codebook used in this system seem to be weak robust in multi-motion objects. In other words, when there are multi motion objects appearing in wide FOV and the selected object is standing still, it is usually to miss the target by the Codebook. However, the method of TLD is more robust.

VI. CONCLUSIONS/FUTURE WORK

In this paper, a station-active dual camera (SADC) system is proposed to acquire the detailed and overall information of object at distance. A stationary wide Field Of View (FOV) camera is used to monitor an environment for detecting and tracking objects. Then, by the method of LUT, a Pan-Tilt-Zoom (PTZ) camera is steered to zoom in the target detected in the stationary camera. When the target is static, PTZ camera put it in the center of the FOV and zooms in to acquire high-definition image. While the object continued moving, PTZ camera receives the location of the target from the station camera and tracks it continuously. Finally, the wide angle camera detects and tracks the candidate object while the PTZ camera zooms in the target.

The method of TLD used in our paper has no restriction on the shape of object. The next step we are going to take is adding a detection algorithm into the PTZ camera so that the camera can zoom in the target by itself and adjust more accurately the image in the center of the view. Furthermore, the clearer the image is, the more details of the image can be obtained. In this way, the details can be transformed to features which feedback the detection process in the static camera.

REFERENCES