Two New Surveillance Systems

Ming-Deng Huang (黃明燈), Ling-Hwei Chen (陳玲慧)
E-mail: eden@debut.cis.nctu.edu.tw and lhchen@cis.nctu.edu.tw
Department of Computer and Information Science
National Chiao Tung University
Hsinchu Taiwan 300, Republic of China
Tel: (886)3-5712121 ext. 59271, Fax: (886)3-5721490

Abstract

In this paper, we will propose two surveillance systems. One is an elevator detection system, which can detect subjects' coming in and going out. The other is a human tracking system, which can track pedestrians on the street. The surveillance systems are implemented utilizing two steps: segmentation and tracking. In the segmentation step, a segmentation algorithm proposed by Horprasert et al. [8] is modified to segment persons in mages. Color and brightness information are used to determine a pixel belonging to background or foreground. In the tracking step, based on the segmented result, we can track the moving persons. The spatial information is adopted to solve the intersection condition when two persons meet during the tracking time.

1. Introduction

In the past years, video acquisition hardware has been considerably improved. The streams obtained from video acquisition hardware have good quality, so the area of developing video-surveillance systems has attracted researchers' attention. Many people have proposed some efficient algorithms for surveillance systems. Celenk and Reza [1] described a technique for the analysis of a sequence of time-varying images taken by a stationary camera. The proposed algorithm is divided into three stages; (i) motion detection, (ii) object location, and (iii) trajectory tracing. In the first stage, two consecutive images are compared and a difference image is formed to detect the presence of any motion object. A subimage is then defined in the difference image as the active region in which the motion is analyzed. In the second stage, the subimage is compared with both previous and next frames. The local widows enclosing the motion object detected are determined in each respective image. These windows are then used in the last stage for tracing the trajectory of the motion object. The mass center of the motion object in each frame is computed, and the motion trajectory is obtained simply by joining these points. In this method, if the background is complex, locating objects is difficult.

Yachida et al. [2] presented a system that detects and tracks moving objects from the videotape images. Yachida's method gets meaningful measures of objects' movements, such as linear and angular velocities, from a set of consecutive images. Temporal and spatial differences of gray levels are used to separate moving objects from the stationary ones. However, if the object is missed or not found in subsequent images, then there is no previous information for discovering the missing object in this system.

Olson and Brill [3] proposed a general application for moving object detection and event recognition. Under this application, moving objects are detected utilizing change detection and tracked utilizing the first-order prediction and the nearest-neighbor matching. The system also utilizes background subtraction to find a person using Gaussian distribution of intensity of each pixel. It is designed for indoor surveillance, but it cannot handle small motions of background objects. This application is a single person tracking system.

Wren et al. [4] has developed a real-time tracking system named "Pfinder". Pfinder mainly uses a statistical background model to locate people. It has been evolved for several years and has been used in many applications. However, Pfinder only tracks a single person using a fixed camera.

Rohr [5] proposed an approach for recognizing human's movement in video images. The approach is applied to images with passersby walking parallel to the image plane. The postures of the persons in each frame are estimated by matching gray scale edges in the image with model edges. Intersection problems in a multiple person situation are not solved and considerations regarding real time operation are not made.

To handle the particularities of nonrigid objects, Choi et al. [6] define a target as an individually tracked moving region or as a group of moving regions globally tracked. They define a moving region as a set of primitives (e.g. corners, contour edges, color) and track globally the moving region by combining all its primitives. However, the tracking algorithm must have high quality of video sequences.

Many researchers introduce template or model-based algorithms to track human. Segen and Pingali [7] used another model-based approach to do object tracking. The algorithm identifies feature points in each video frame, and matches feature points across frames to produce feature "paths", then short-lived and partially overlapping feature paths are grouped into longer living trajectories representing motion of individual persons. However, the matching of feature points between the current and former frames is frequently not successful. In addition, they can not perform tracking when two persons intersect.

In this paper, a useful algorithm for surveillance systems is proposed. We track people within a static background.

We choose two different environments, indoor and outdoor, to implement our tracking algorithm. The elevator environment is taken as the indoor testing one. The entering and leaving of persons are monitored. When somebody enters or leaves, the tracking system will send a signal to the remote control center. Then, the guard can know somebody in or out the elevator. On the other hand, we also test our tracking system in the outdoor environment. The video camera is set on the street to track the pedestrians. When a pedestrian enters the view of our camera, we locate the pedestrian and start to track his trajectory. The remainder of the paper is described as follows. In Section 2, we will present a modified object segmentation method. In Section 3, we will propose two surveillance systems. They are an elevator detection system and a human tracking system. Finally, we will give some experiments in Section 4 to show that the proposed method can perform well in the two proposed systems. Our conclusion will also be presented in Section 5.

2. Object Segmentation

The proposed surveillance method consists of the segmentation and tracking steps. This section focuses on the segmentation step. The segmentation method adopted in this paper comes from Horprasert's [8] idea.

Color and brightness informations are used to extract the regions of the moving objects. Some post-processings are provided to eliminate noise and shadow and fill some holes within the extracted regions. These post-processings make the moving object tracking more convenient.

In the segmentation step, we mainly use the background subtraction skill. In the following, we will describe the method in details.

2.1. Background Modeling

In this section, we will define two measurements: color difference and brightness difference. These two measurements can help us to perform segmentation. In the paper, the average of a sequence of background images is viewed as the reference frame.

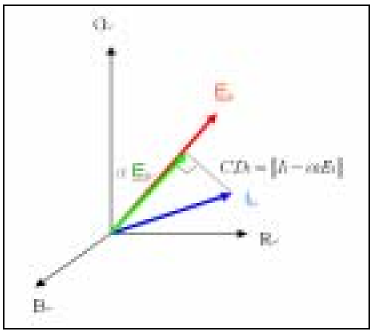

Fig. 1 Color model.

Fig. 1 shows the RGB color model, I_i represents the color of a given i-th pixel in an input image and E_i represents the color value of the corresponding pixel in the reference image. CD_i is the color difference between E_i and I_i. $\alpha_i E_i$ is a vector which is the projection of I_i on E_i, and α_i is a scale value. Then, the color difference can be evaluated by $CD_i = \| I_i - \alpha_i E_i \|$.

In the segmentation step, we find that not all of the pixels in input frames have color information. That is, some pixels are gray scale ones. Therefore, we use brightness information to accomplish the segmentation step. So, for any pixel without color information, we compute the brightness distance (BD_i) to accomplish the segmentation step.

Due to the environmental noise, the background will have some variation. To tolerant the variation, we will estimate the variation for each background pixel in advance.

First, a short video with only background scenes is taken. Suppose that there are N (N>1) frames in the short video. The color difference, CD_i of the i-th pixel between the reference image and each background image is calculated. Based on the CD_i values of the i-th pixel in N background images, we select the upper bound of CD_i (τ_{CD_i})as the color difference threshold for the i-th pixel. Similarly, the brightness difference threshold for the i-th pixel (α_{BD_i}) is got in the same way.

2.2. Pixel Classification

In the previous section, we have defined the color and brightness difference. We also find the color and brightness threshold for each pixel. Based on these thresholds, a pixel in an incoming image can be classified into a background or foreground pixel. For each pixel i in an incoming image, evaluate CD_i and BD_i between the incoming image and the reference image first. Then, the following procedure is applied.

Pixel Classification

If $CD_i > \tau_{CDi}$

then pixel i is **the foreground pixel**

Else if $I_R(i) \cong I_G(i) \cong I_B(i)$ and $BD_i > \alpha_{BDi}$

then pixel i is **the foreground pixel**

Else

then pixel i is **the background pixel**

2.3. Post-Processing

The post-processing contains three parts: (1) noise-removing; (2) hole-filling; (3) shadow-eliminating. We now go into more detail about these three parts.

2.3.1 Noise-Removing

After applying pixel classification, many noises will exist. To remove them, a region growing for foreground pixels is conducted. If the pixel number of a region is too few, the region is considered as background.

2.3.2 Hole-Filling

After the noise-removing procedure, some holes inside objects may exist. These holes need to be filled. Region growing skill is also used to fill holes. First, the region growing is used to find background, and the background is filtered out. Then, the region growing is applied to find holes in object regions. Finally, these holes are filled.

2.3.3 Shadow-elimination

Sometimes, we will find that there is a shadow under the moving object. To eliminate the shadow, we will propose a method to remove the shadow and describe it as follows.

First, 1D vertical projection of the moving objects is obtained. We can find that the shadow part will result in low frequency. Based on the fact, a threshold, T, is given, and any foreground point corresponding to its 1D vertical projection component less than T is removed.

3. Tracking

After segmenting the objects, we can start to trace persons. Two scenes, the indoor and outdoor environments are chosen to implement the surveillance method.

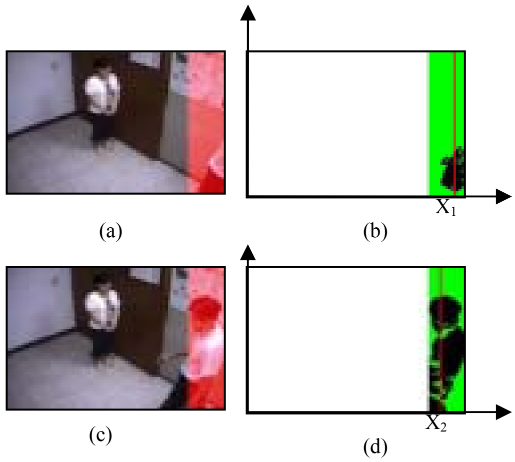

Fig. 2 An example of a person coming in. (a) The first frame captured by a video camera when someone comes into the elevator and the red area is the "Passing area". (b) The segmentation result of the passing area of (a) and the x-coordinate (x_1) of the centroid of the guy coming in. (c) The second frame when someone comes in. (d) The segmentation result of the passing area of (c) and the x-coordinate (x_2) of the centroid of the guy coming in.

3.1 Indoor-tracking ---

An elevator detection system

In this system, a video camera is set up at the left-top of the elevator door to detect persons entering in or going out. We mainly use the spatial information to determine whether somebody comes in or goes out. We find that when someone comes in or goes out of the elevator, he has to be close to the elevator door. Therefore, we can detect the motion in the area near the elevator door.

When somebody is found to enter the ***Passing Area*** (as shown in Fig. 2(a)), we compute his centroid. We only record the x-coordinate of the centroid, like x1 and x2 shown in Figs. 2(b) and 2(d). When the value of (x1 – x2) is positive, we think that there is somebody entering the elevator. On the other hand, if (x1 – x2) is negative, we think that there is somebody leaving the elevator.

3.2 Outdoor tracking ---

A human tracking system

In the outdoor environment, the video camera is set at a corner where moving objects will pass seldom. In the outdoor tracking environment, we focus at most two persons existing in one scene and subjects enter the scene individually. When a pedestrian enters the scene of our camera, we locate the pedestrian and start to track his/her trajectory. A trajectory of a person is composed of the person's centroid in each frame. In order to describe our system, two variables are defined:

(1) ***Current_Blob_Count***: The number of blobs that are obtained from the segmentation result in the *current* frame.

(2) ***Former_Blob_Count***: The number of blobs that are obtained from the segmentation result in the *former* frame.

Base on the relationship between Current_Blob_Count and Former_Blob_Count, the system status is determined and described as follows:

Case 1: Current_Blob_Count > Former_Blob_Count

There are two reasons that will cause this situation. One is that there is a person entering the view of the video camera, and the other is that the two intersecting persons separate. To determine which reason cause the situation, we detect whether there is any person near the boundary of the view in the current frame. If the answer is yes, we consider that a person is entering the view and create a new trajectory to the person. Otherwise, we consider that the two intersecting persons are separating.

If the two intersecting persons separate, we can find out the corresponding person according to their color feature. We extend each trajectory to the centroid of the corresponding person.

Case 2: Current_Blob_Count = Former_Blob_Count

In this situation, for each blob in the former frame, we find the nearest region in the current frame region obtained from segmentation result. Then, we extend the blob's trajectory to the centroid of the found region.

Case 3: Current_Blob_Count < Former_Blob_Count

There are two reasons that will cause this situation. One is that there is a person leaving the view of the video camera, and the other is that original two separating persons intersect. To determine which reason causes this situation, we mainly determine whether there is a person near the boundary of the view of the video camera in the former frame and not in the current frame. If the answer is yes, we consider that there is a person leaving the view of the video camera. Otherwise, we consider that the original two separating persons intersect.

If a person leaves the view of the video camera, we delete the leaved person's trajectory. If two separating persons intersect, only one blob will appear in the scene. In this case, each trajectory is extended to the centroid of the only one blob in the scene.

4. Experimental results

In this Section, the experimental results performed under our proposed algorithm are given. We choose two scenes, the indoor and outdoor environment, to implement the surveillance method.

4.1 The testing result of the elevator surveillance system

Here, a corner of the Automatic Information Processing Laboratory in NCTU is used to simulate an elevator environment. When somebody comes in, the ”in” signal will be sent. When somebody goes out, the ”out” signal will be sent. We have done many tests including someone coming in and leaving, and the experimental result shows that the elevator surveillance system can work well.

4.2 The testing result of a human tracking system

We set up a video camera on someplace, where moving objects seldom pass, to track pedestrians. Figs. 3 show two persons enter the view one after another and leave the view individually. We track their trajectory using red and green lines.

5. Conclusions

In this paper, we have constructed two real-time visual surveillance systems to detect and track persons and monitor their activities in indoor and outdoor environments. The two real visual surveillance systems, the elevator surveillance system and the human tracking system, have been constructed. The elevator surveillance system can monitor persons' entering and leaving the elevator. The human tracking system can track the pedestrians on the street. We have proposed a simple and flexible segmentation algorithm to extract moving persons under complex backgrounds. The spatial information is used to help us to simultaneously track multiple persons even with intersection. The experimental results also show the effectiveness of the proposed method.

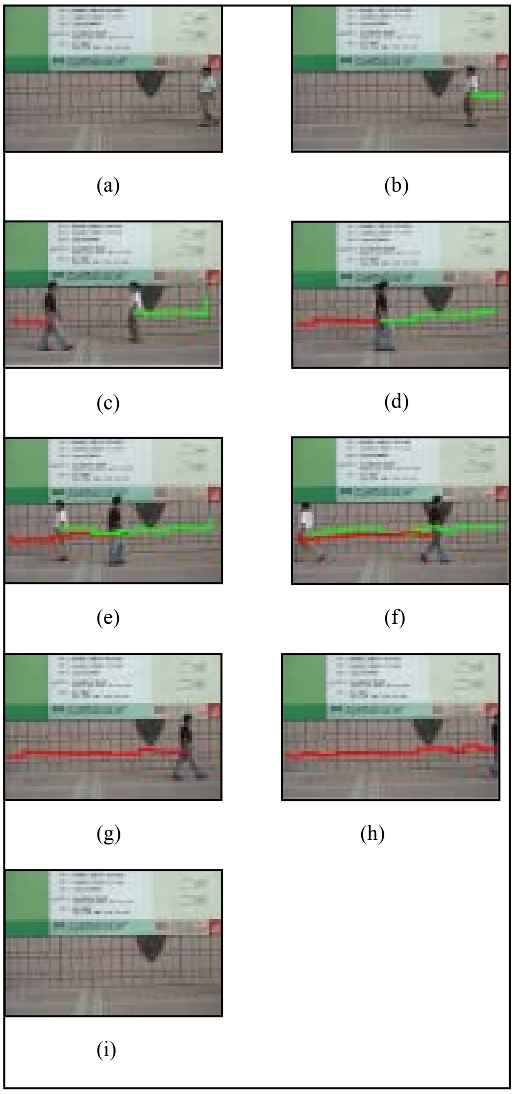

Fig. 3 The tracking result.

REFERENCES

[1] M. Celenk and H. Reza, “Moving Object Tracking Using Local Windows,” Proc. IEEE International Symposium on Intelligent Control, pp.180-185, 1989.

[2] M. Yachida,, M. Asada, and S. Tsuji, “Automatic Analysis of Moving Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.3, no.1, pp.12-19, Jan. 1981.

[3] T. Olson and F. Brill, "Moving Object Detection and Event Recognition Algorithms for Smart Cameras," Proc. DARPA Image Understanding Workshop, pp. 159-175, 1997.

[4] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997.

[5] K. Rohr, "Towards Model Based Recognition of Human Movements in Image Sequences," Computer Vision Graphics and Image Processing: Image Understanding, vol.59, no.1, pp.94-115, January, 1994.

[6] S. Choi, Y. Seo, H. Kim, and K. Hong," Where Are the Ball and Player? Soccer Game Analysis with Color-based Tracking and Image Mosaic," International Conference on Image Analysis, pp. 83-91, 1997.

[7] J. Segen and S. Pingali, "A Camera-Based System for Tracking People in Real Time," Proc. IEEE International Symposium on Pattern Recognition, vol.3, pp.63-67, 1996.

[8] T. Horprasert, D. Harwood, and L.S. Davis , "A Robust Background Subtraction and Shadow Detection ," Proceedings of 4th Asian Conference on Computer Vision, Taipei, Taiwan, January 2000 .