Multiple Camera-based Correspondences of People using Camera Networks and Context Information

Hyun-uk Chae. Suk-Ju Kang
Kang-Hyun Jo

Graduate School of Electrical Engineering, University of Ulsan,
Ulsan, Korea, (e-mail: {hwchae, sjkang, jkh2009}@islab.ulsan.ac.kr)

Abstract: This paper proposes a method to correspondence a same person from different views. To design a system with multi vision sensors, we should to solve the correspondence problem. For obtain a correspondence, a proposed method has three steps. The first step is to detect moving objects by background subtraction from multiple background model. The background model is generated from image sequence and it updates by itself in each process. The temporal difference is used jointly to compensate a limitation of multiple background model. The detected silhouettes are divided as individual by labeling when more than two people are detected in an image. The second step is to generate blobs from detected silhouette. The silhouette is segmented as blobs by a criterion with color range and context information of human body. A person is composed of a set of blobs. The blob set is represented as a model by estimation of Gaussian mixture model (GMM). In the final step, a GMM of detected person from a camera is matched with other GMMs from different cameras to identify a same person.

Keywords: Camera networks, Multiple cameras, Gaussian mixture model, Maximum likelihood.

1. INTRODUCTION

In recently, the automated systems improve the necessity by itself in a structured space such as robot town, smart environment, intelligent space and surveillance system [Hashimoto, 2002 and Javed et al, 2008]. Those systems are generally composed of robot, several sensors and computers. The systems need technologies which are process of sensing information, fusion of each sensor, network communication between robot, sensors and computer, and so on for designing the systems. This paper regards a focus as designing an object tracking system based on multi-vision sensors as a part of process of sensing information in the above network system.

The multiple cameras are generally used to overcome the restriction of FOV (Field of View) when a single camera is used. The camera calibration is executed in the multiple camera systems for explaining the three dimensional space. The calibration of multiple cameras is a process of getting intrinsic and extrinsic parameters of each camera, i.e., we are able to know inter-relationship between cameras. Such calibrated systems usually make an assumption that whole cameras are calibrated in advance. However, if we need to add cameras, it is difficult to calibrate the whole cameras immediately. Therefore, we address the problem of corresponding same people under different views in this paper. For solving the problem, a proposed method has three steps as shown in Fig. 1. The first step is a detection of moving objects by a background subtraction. The multiple background model and temporal difference are generated for detecting a motion and they update by themselves in each frame. The number of used image for generating the background model is selected empirically. The detected regions then are divided as a person by labelling when more than two people are detected in an image.

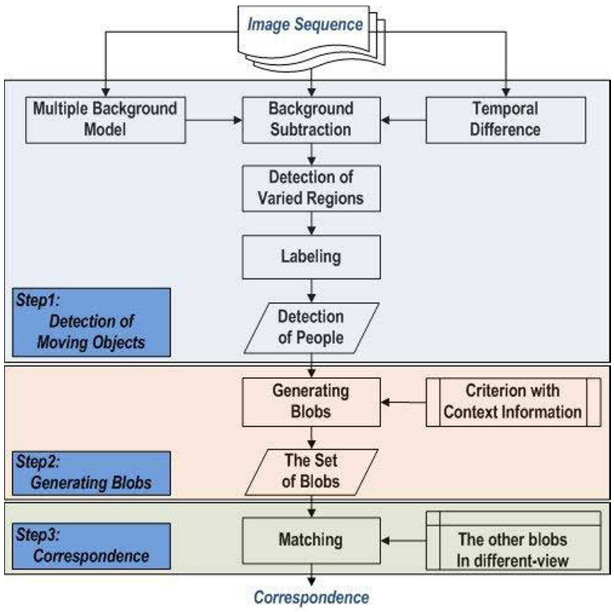

Fig. 1. Flow diagram of proposed method.

The second step is to generate blobs from detected silhouettes. Each silhouette is segmented by a criterion with color range and context information of human body. The generated blobs are described as Gaussian probability distribution. A person is composed of a set of blobs. The set of blobs is generated as a model by estimation of Gaussian mixture model (GMM). In the final step, an estimated GMM from a set of blobs in an image are matched with other GMMs from different views using maximum likelihood estimation. We then identify a same person in the different views. Accordingly, we are able to replace the calibration of additional cameras from obtained motion information of objects.

The organization of this paper is as follows: The detection of moving object is described in section 2. The generating color blobs are explained in section 3. In section 4, we present the method to obtain correspondence of objects. The experimental results and conclusion are explained in section 5 and 6, respectively.

2. MOTION DETECTION

A method for detection of moving people is presented in this section. The moving objects are detected by background subtraction from multiple background model. A temporal difference is also used to reduce a noise from variance of scene condition. The multiple background model and temporal difference are generated from image sequence and they update themselves in each frame. A labelling method then is used for dividing each person into regions when more than two people are detected in an image. Small regions are removed as noise. In this process, there is an assumption that moving object is only person. The process of motion detection is described in Fig. 2.

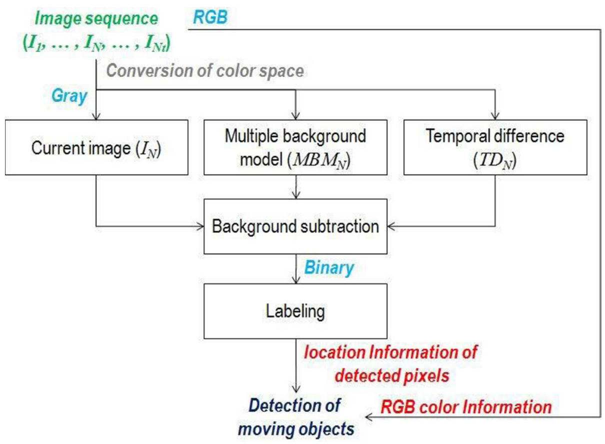

Fig. 2. The process of motion detection

2.1 Generating background model

The background subtraction is to segment moving objects from difference between background model and current image. It is simple and efficient method in fixed camera system. The generation of background model is important task to detect accurate motion. We use multiple background model which has several backgrounds because the pixel intensity is frequently changed over time. The input image sequence is converted RGB into grayscale. The process of generating multiple background model has three steps which are online clustering, removing clusters with small weight, and generating background models. The more detail explanation is described in [Kim et al, 2007 and Seo et al, 2008]. The number of frames for generating the multiple background model is selected empirically (N=100). Then a temporal difference (TD) is obtained by difference between current image (I_N) and previous image (I_{N-1}) as in Eq. (1). δ_{TD} is threshold value and selected empirically.

$$TD_N = \begin{cases} 255, & |I_N - I_{N-1}| \geq \delta_{TD} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

2.2 Background subtraction and labeling

The moving object is detected by intersection of results of background subtraction and temporal difference. The result of background subtraction depends on generation of exactly background model and selection of threshold value. For more exactly results, we use background model of two types and the threshold value is selected by iterative method. The labeling is executed for segmenting each person among candidate moving objects and removing noise. The detected regions are scanned by mask to search connected components in the detected region. The connected regions are labelled (Label = 1, 2, ⋯ , n) only when the number of pixel in one label is larger than 200. Fig. 3 describes the results of motion detection. The first row is the results of background subtraction. The second row is the results of labelled objects.

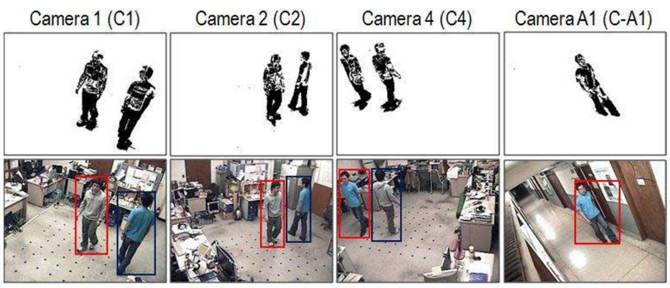

Fig. 3. The results of motion detection

3. GENERATING BLOBS

We present the method to generate blobs from result of detecting motion in this section. The color space in the detected region is converted RGB (Red-Green-Blue) into HSI (Hue-Saturation-Intensity). The region is separated by a criterion as blobs which have Gaussian probability distribution. An object has a set of blobs and it is represented as a model by estimation of GMM.

3.1 Color segmentation

Before generating blobs, the color space is converted RGB into HSI. The HSI color space is also widely used as RGB and YUV color space. Furthermore, this color model is robust illumination variance than RGB color model. The range of each channel value is usually as follows: hue is from 0° to 360°, saturation and intensity are 0 to 1. The hue value is undefined when the saturation is zero and the saturation value is defined only when the intensity is nonzero. In this paper, the degree of hue (0°~360°), saturation and intensity