Pedestrians Tracking in a Camera Network

Rahman Yousefzadeh
Department of Computer and IT Engineering
Shahrood University of technology, Shahrood, Iran
rah.you@ieee.org

Received: 2011/01/24 ;Accepted: 2011/03/11 Pages: 97-104

Abstract

With the increase of the number of cameras installed across a video surveillance network, the ability of security staffs to attentively scan all the video feeds actually decreases. Therefore, the need for an intelligent system that operates as a tracking system is vital for security personnel to do their jobs well. Tracking people as they move through a camera network with non-overlapping field of view is a challenging issue. It requires re-identification of objects when they leaves one field of view and appears at another scene later. The appearance of the people is changed significantly among the cameras due to illumination changes, parameters of camera, viewing angle, and deformable geometry of people. Additionally the observations of people are often broadly separated in time and space and common proximity techniques can not be used to constrain possible correspondences. In this paper a new feature was proposed to represent the appearance of people, which is capable of dealing with the common illumination changes occurring in indoor environment. To produce the proposed feature, first the image was transformed to YCbCr color space. Then diagonal vectors of color co-occurrence matrix of each individual were extracted and normalized. Experimental results from a real surveillance scene showed the efficiency of the proposed method.

Keywords: *Tracking, disjoint camera views, co-occurrence matrix, YCbCr color space, Jensen function*

1. Introduction

Object tracking is the task of approximating the trajectory of an object as it moves in a certain area. In most cases, covering the complete surveillance area is not feasible by a single camera due to limitations of cameras field of view and the structure of scene. Therefore, surveilling a wide area requires a system with the ability to track objects across multiple cameras. Moreover, it is usually not feasible to thoroughly cover large areas with cameras having overlapping views due to economical and computational constraints. Thus, in realistic scenarios, the system should be able to handle multiple cameras with non-overlapping fields of view. A feature that is commonly used for tracking is the appearance of the objects. Color is a significant feature of a person's appearance that is influenced by scene illumination, camera parameters, viewing angle, surface material properties and geometry of people. Thus, the color distribution of an object can be different when viewed from two different cameras. However, the object surface distinct properties may remain unchanged as it moves across cameras.

Various papers in the literature have addressed the problem of pedestrian tracking across multiple, possibly disjoint cameras. Rahimi et al [1] have introduced the use of spatial-temporal constraints to consistently track objects between disjoint views by explicitly estimating the spatial relationship between several cameras.

Other researchers have explored how to address color and intensity changes across disjoint views. Porikli et al. [2] attenuates color differences by utilizing a similarity metric that is somewhat invariant to changes in global illumination. Javed et al. [3] have used a low dimensional subspace method for a brightness transfer function via principal component analysis (PCA) on a set of known intensity mappings obtained from object observation samples, under some linearity assumptions of the transfer function and independence of the color channels. To increase accuracy, their matching scheme is then combined with additional space-time cues from known camera topologies. However, the authors' have assumed that objects are planar, radiance diffuse and illumination uniform throughout the whole field of view do not generally hold in real applications.

Jeong and Jayne have developed a method that operates on chromaticity samples to increase the temporal stability of the transfer function between any camera pair[4]. In [5] Madden et al. have introduced a feature named Major Color Spectrum Histogram Representation (MCSHR) to describe the object's main colors. They have used an intensity re-mapping of the object's R, G and B components to mitigate the problem of varying illumination across disjoint views. In [6] Prosser et al have proposed a Cumulative Brightness Transfer Function (CBTF) for mapping color between cameras located at different physical sites.

This paper proposed a new method to establish correspondence between observations from disjoint camera views in typical indoor environment. A new representation of people appearance was introduced that is capable of dealing with the typical illumination changes. This representation was constructed from the diagonal elements of co-occurrence matrix in YCbCr color space.

The next section introduces our proposed method in this research. In Section III experimental results were explained which validated the proposed method. Finally Section IV drew conclusions of this paper.

2. Proposed Method

Any tracking system needs to initially detect the moving objects in the scene. A number of moving object detection approaches have been reviewed in [7].

In this research, background modeling was used to detect the objects. Background was modeled by Gaussian Mixture model proposed in [11] and people segmented form background following by a number of post-processes to eliminate the effect of disturbing factor such as noise and shadow. The moving objects were then tracked using the Kalman filter, which is capable of tracking people at presence of noise and occlusion [12].

In this approach, individual cameras in the network detect and track the pedestrians in their own view. The results of single camera tracking are stored in a set of observations. Each observation denotes the track and the appearance model of different people in each camera. The problem of pedestrian tracking in camera network is then reduced to find which of the observations in the system belong to the same object [8].

The field of view of each camera was divided into entrance scenes, where people enter /exit the environment, or normal scenes. The processing of entrance scenes and normal scenes was somewhat different. Tracking was performed in both types of scenes, but feature extraction, matching and inference were only performed in the entrance scenes.

Color histogram is a feature that is widely used to model the appearance of the people and is utilized in various multi tracking systems to determine correspondence between observations. However color of an object is a function of parameters such as scene illumination, viewing angle and camera intrinsic parameters. Hence color associated with a person may change significantly in different scenes.

Color is typically represented as a histogram, which is a quantity that captures global distribution of color in an image. One of the main disadvantages of the histogram-based methodologies is that the spatial distribution and local variations in color are ignored.

Texture is another low-level feature that is usually used to represent the appearance of an object. Gray scale Co-occurrence Matrix (GCM) is a popular method for texture extraction in the spatial domain [9]. A GCM stores the number of pixel neighborhoods in an image with a certain gray scale combination. GCM is a square matrix that gcm(i, j) is the number of times the gray level of a pixel p equals i and gray level of its neighbor equals j. GCM matrix is generated for neighborhood directions, that are, 0, 45, 90 and 135 degrees. In the GCM method for texture extraction, color information is completely lost as pixel gray levels are measured.

To combine spatial information with the color of image pixels, the proposed method used Color Co-Occurrence Matrix (CCM). Instead of using Gray scale values, CCM works on each channels of color image.

To mitigate the effect of illumination changes in different scenes, image pixels were transferred to YCbCr color space.

This color space is more robust to such changes because brightness (illumination) values are stored in a separate component. YCbCr represents color as brightness and two color difference data, while RGB represents color as red, green and blue. In YCbCr, the Y is the brightness (luma), Cb is blue minus luma (B-Y) and Cr is red minus luma (R-Y).

To convert from RGB to YCbCr the following equation was used [10]:

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ 0.439 & -0.386 & 0.071 \\ -0.148 & -0.291 & 0.439 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

In the next step, CCM was computed for each channel of the image, i.e. three color co-occurrence matrices were produced for Y, Cb and Cr channels of image. Figure 1.b depicted the CCM for Y channel of the image shown in Figure 1.a.

As it is shown in Figure 1.b non-zero values of CCM are located near the main diagonal of the matrix. This structure is due to the nature of human appearance. In most cases, dresses consist of limited colors and have uniform color distributions. Our proposed feature was generated by selecting the main diagonal elements of CCM and ignored others. After generating diagonal vectors, in the final step two post processes were applied to increase the robustness of the feature. First post-process normalized the extracted feature vector in order to decrease the effect of different sizes. The second post-process converted the feature vector to cumulative distribution in order to mitigate the effects of shift imposed by illumination changes. After this step, three vectors existed for color co-occurrence matrices of each channel.

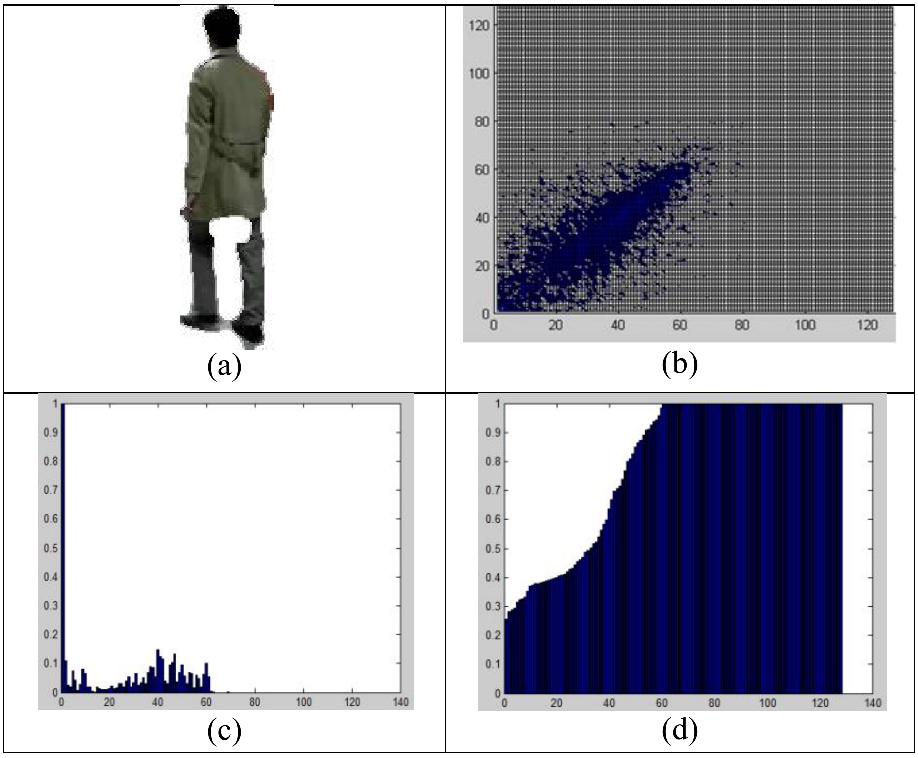

Figure 1. (a) image of a person (b) co-occurrence matrix of Y channael (c) Diagonal vector of b (d) Cumulative distribution of c

To measure the similarity of extracted features between two objects, Jensen similarity measure was utilized on correspondence vectors of each object (i.e. Y vector of first object with Y vector of second object, same for Cb and Cr vectors).

The Jensen function is a measure to compute the similarity of two vectors P and Q, of k elements as defined below:

$$j(p,q) = \frac{1}{2}\sum_{i=1}^{k}\{A_i - B_i\}$$

$$A_i = P_i^{'} \log_2 P_i^{'} + Q_i^{'} \log_2 Q_i^{'},\ B_i = (P_i^{'} + Q_i^{'}) \log_2 (P_i^{'} + Q_i^{'})$$

$$P^{'} = p_i \Big/ \sum_{i=1}^{k} p_i ,\ Q^{'} = Q_i \Big/ \sum_{i=1}^{k} q_i$$

Output of the Jensen function is between 0 and 1. The output of the Jensen function is zero if two sequences P and Q are exactly the same.

Overall similarity is computed by multiplying all three similarity values.

3. Experimental Results

In this section, results of the proposed method were considered. This method was tested on videos captured by five disjoint cameras installed in the Laboratory buildings at the Shahrood University of Technology. Cameras were installed in three different floors of the building. Several complexities were assumed in these videos. a number of disturbing parameters such as low quality of videos, shadow, noise and different illumination conditions were in each view. Also size of the people could change dramatically. Figure 2 showed a little part of tracking sequence of a person in the proposed method in four cameras.

To validate the accuracy of the proposed method, matching procedure was performed with color histogram and our method in different color spaces including RGB, Normalized RGB, YCbCr and HSV. Figure 3 depicted the performance of the proposed method compared with color histogram in different color spaces. The proposed method outperformed the color histogram in those color spaces with separate brightness value.

In other test the effect of similarity measure was also considered in the accuracy of object matching procedure. Euclidian, Kullback-Leilber, Jensen and Bahhatecheryya methods were used to compute similarity of our proposed method. As shown in Figure 4, Jensen function led to better results in our database.

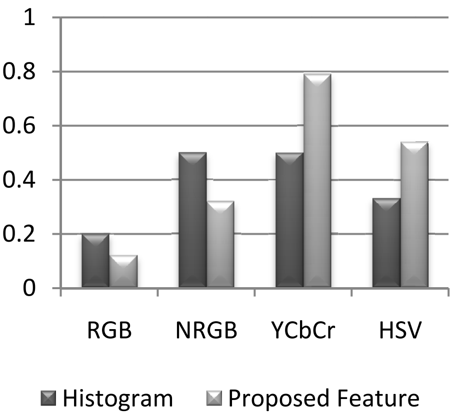

Figure 2. Accuracy of the proposed method comared with color histogram

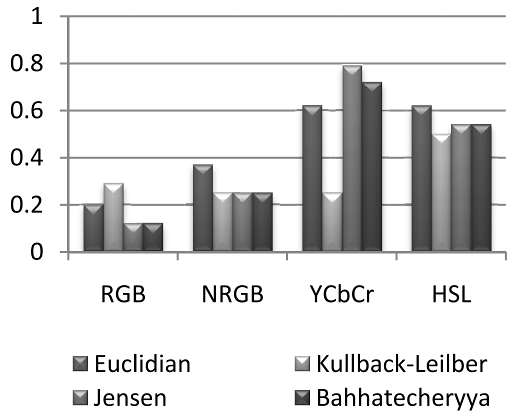

Figure 3. Accuracy of object matching with proposed method in different color spaces

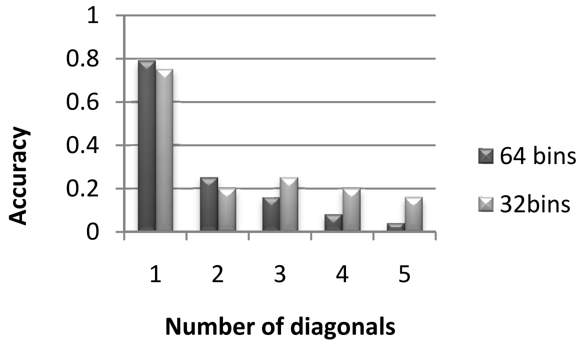

Figure 4. Impact of the number of diagonals. Performance of the matching procedure is fallen dramatically bu increasing the number of diagonals

As stated above, only the diagonal elements of co-occurrence matrix were included in the proposed feature. To check the effect of the number of diagonal elements, the proposed method was tested with 1 to 5 diagonals of co-occurrence matrix. As depicted in figure 6, performance of the matching procedure was fallen with increasing the number of diagonals.

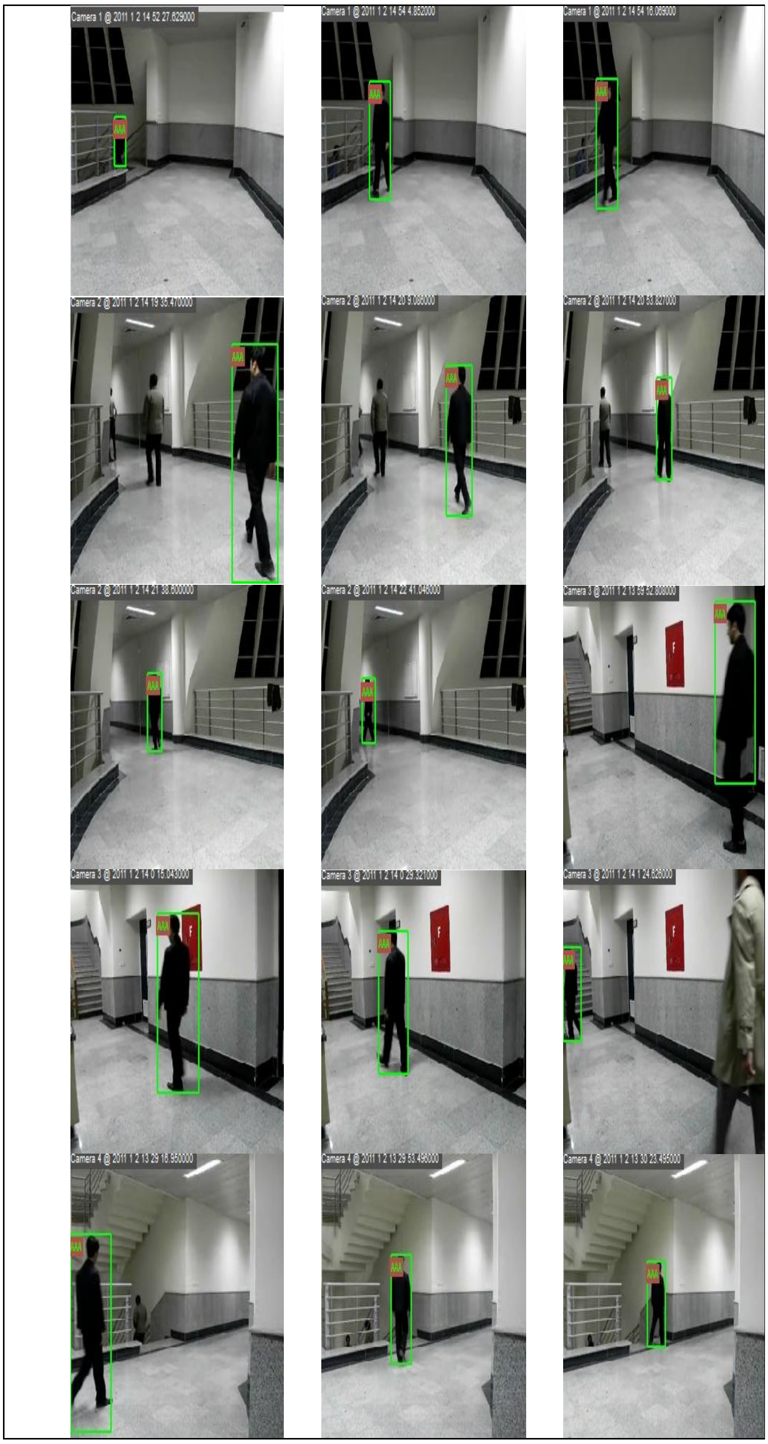

Figure 5. A sample result of the tracking with proposed method in 4 cameras

4. Conclusions

In this paper, a method was proposed for tracking people through disjoint camera views based on Color Co-Occurrence matrix of person's appearance representation. Tracking in non-overlapping cameras is a challenging issue because of varying appearance of objects due to illumination conditions between different scenes. Our proposed method was working on YCbCr color space that separated the brightness information of the color. In addition this method combined spatial information as well as color information of image to provide more robust feature than color histogram. An advantage of the proposed system was that it combined spatial information with the color of image pixels. Experimental results demonstrated the better accuracy of proposed method compared with color histogram.

References

1. A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous calibration and tracking with a network of non-overlapping sensors," Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., pp. 187-194.
2. F. Porikli and a Divakaran, "Multi-camera calibration, object tracking and query generation," 2003 International Conference on Multimedia and Expo. ICME '03.Proceedings (Cat.No.03TH8698), 2003, p. I-653.
3. O. Javed and K. Shafique, "Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05), pp. 26-33.
4. K. Jeong and C. Jaynes, "Object matching in disjoint cameras using a color transfer approach," Machine Vision and Applications, vol. 19, May. 2007, pp. 443-455.
5. C. Madden, "Tracking People Across Disjoint Camera Views," 2009.
6. B. Prosser, S. Gong, T. Xiang, and Q. Mary, "Multi-camera matching using bi-directional cumulative brightness transfer functions," Proceedings of the British Machine Vision Conference, Citeseer, 2008.
7. A. Yilmaz, O. Javed, and M. Shah, "Object Tracking : A Survey," ACM Computing surveys, vol. 38, 2006.
8. O. Javed and M. Shah, Automated Multi-Camera Surveillance Algorithms and Practice, Springer, 2008.
9. A. Vadivel, S. Sural, and a Majumdar, "An Integrated Color and Intensity Co-occurrence Matrix," Pattern Recognition Letters, vol. 28, Jun. 2007, pp. 974-983.
10. Gonzalez, R.C. and Woods, R.E., "Digital Image Processing", Prentice Hall, (2004).
11. C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., IEEE, 2002, pp. 246-252.
12. G. Welch and G. Bishop, "An introduction to the Kalman filter," University of North Carolina at Chapel Hill, Chapel Hill, NC, 1995.