Pedestrian Tracking through Camera Network with Disjoint Views

Rahman Yousefzadeh and Hamid Hassanpour

Abstract—**Tracking people as they move through a camera network with non-overlapping field of view is a challenging issue due to the significant changes of their appearance among the cameras. The varying appearance of a pedestrian is the consequence of illumination changes, parameters of camera, viewing angle, and deformable geometry of people. In addition the observations of individuals are often widely separated in time and space in such systems and common proximity techniques cannot be used to constrain possible correspondences. In this paper a new feature is proposed to represent the appearance of people, which is capable of dealing with the typical illumination changes occurring in indoor environment. To construct the proposed feature, co-occurrence matrix of image computed in YCbCr color space. Then diagonal vectors of the co-occurrence matrix is extracted and normalized. Experimental results from a real surveillance scene show the efficiency of the proposed method.**

Index Terms—**Tracking, disjoint camera views, co-occurrence matrix, YCbCr color space, jensen function.**

I. 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 fairly 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] 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' assumptions 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 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.* introduced a feature named Major Color Spectrum Histogram Representation (MCSHR) to describe the object's main colors. They use an intensity re-mapping of the object's R, G and Bcomponents to mitigate the problem of varying illumination across disjoint views. In [6] Prosser *et al.* proposed a Cumulative Brightness Transfer Function (CBTF) for mapping color between cameras located at different physical sites.

In this paper we propose a new method to establish correspondence between observations from disjoint camera views in typical indoor environment. A new representation of people appearance is introduced that is capable of dealing with the typical illumination changes. This representation is 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 are explained which validate the proposed method. Finally Section IV draws conclusions of this paper.

II. 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 is used to detect the objects. The moving objects are then tracked using the Kalman filter, which is capable of tracking people at presence of noise and occlusion.

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].

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 that have 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 colorof image pixels, the proposed method uses 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 are 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 YCbCrthe following equation is used [10]:

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ 0.439 & -0.386 & 0.071 \\ -0.148 & -0.291 & 0.439 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

In the next step, CCM is computed for each channels of the image, i.e. three color co-occurrence matrices are produced for Y, Cb and Cr channels of image. Fig. 1 (b) depicts the CCM for Y channel of the image shown in Fig. 1(a).

As it shown in Fig. 1 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 is generated by selecting the main diagonal elements of CCM and ignores others. After generating diagonal vectors, in the final step two post processes are applied to increase the robustness of the feature. First post-process normalizedthe extracted feature vector in order to decrease the effect of different sizes. The second post-process converts the feature vector to cumulative

distribution in order to mitigate the effects of shift imposed by illumination changes.
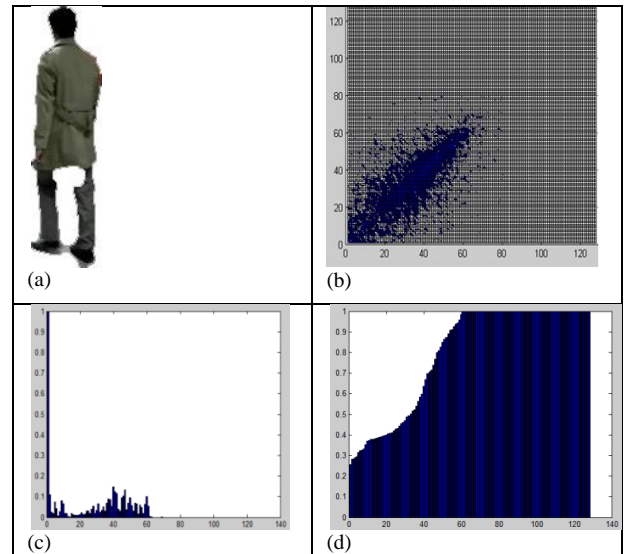

Fig. 1. (a) Image of a person. (b) Co-occurrence matrix of Y channael. (c) Diagonal vector of b. (d) Cumulative distribution of c.

After this step, three vectors are exists for color co-occurrence matrices of each channel.

To measure the similarity of extracted features between two objects, Jensen similarity measure is 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, ofk elements as defined below:

$$j(p,q) = \frac{1}{2} \sum_{i=1}^{k} \{(P_i' \log_2 P_i' + Q_i' \log_2 Q_i') - ((P_i' + Q_i') \log_2 (P_i' + Q_i'))\} \quad (2)$$

$$P' = \frac{p_i}{\sum_{i=1}^{k} p_i}, Q' = \frac{Q_i}{\sum_{i=1}^{k} q_i} \quad (3)$$

Output of the Jensen function is between 0 and 1. The output of the Jensen function will be equal to zero if two sequences P and Q are exactly the same. Overall similarity is computed by multiplying all three similarity values.

III. EXPERIMENTAL RESULTS

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

To validate the accuracy of the proposed method, matching procedure is performed with color histogram and our method in different color spaces including RGB, Normalized RGB, YCbCr and HSV. Fig. 3 depicted the performance of the proposed method compared with color histogram in different

color spaces. The proposed method outperforms the color histogram in those color spaces with separate brightness value.
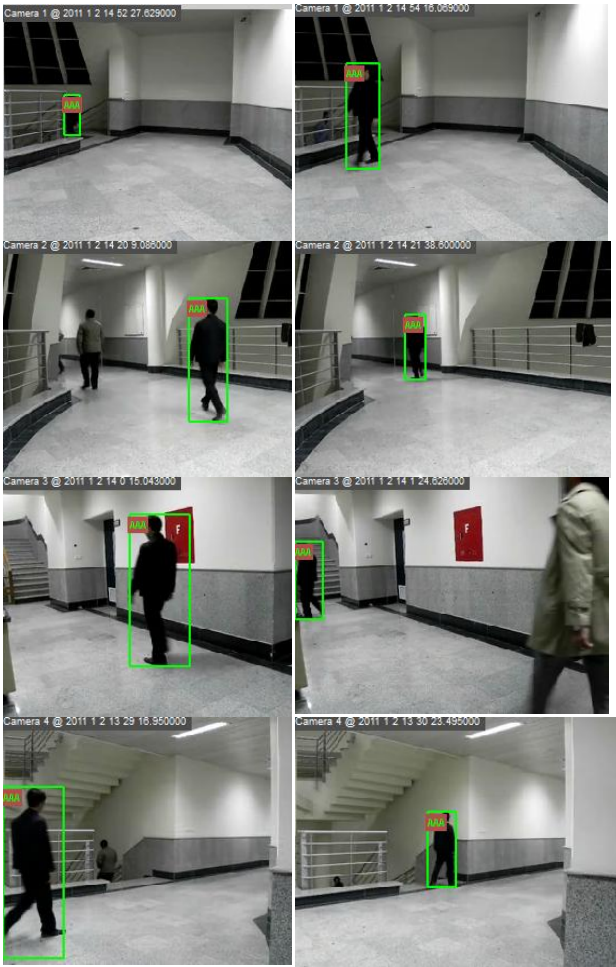

Fig. 2. A sample result of the tracking with proposed method in 4 cameras.
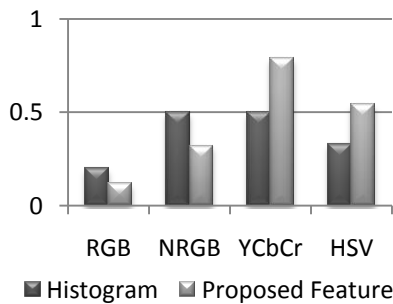

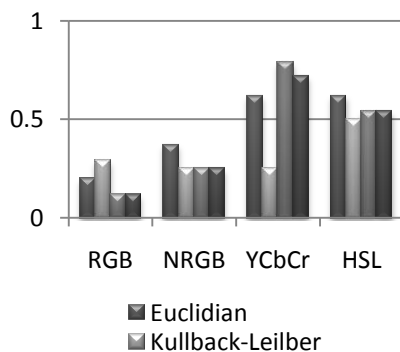Fig. 3. Accuracy of the proposed method compared with color histogram.

Fig. 4. Accuracy of object matching with proposed method in different color spaces.

In other test we also consider the effect of similarity measure in the accuracy of object matching procedure. Euclidian, Kullback-Leilber, Jensen and Bahhatecheryya methods are used to compute similarity of our proposed method. As shown in Fig. 4, Jensen function leads to better results in our database.

IV. CONCLUSIONS

In this paper, we have proposed a method 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 is working on YCbCr color space that is separate the brightness information of the color. In addition this method combines spatial information as well as color information of image to provide more robust feature than color histogram.

REFERENCES

[1] A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous calibration and tracking with a network of non-overlapping sensors," in *Proc. IEEE Computer Society Conf. 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," in *Proc. 2003 International Conf. on Multimedia and Expo. ICME '03 (Cat.No.03TH8698)*, 2003, pp. I-653.

[3] O. Javed and K. Shafique, "Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras," in *Proc. 2005 IEEE Computer Society Conf. 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, pp. 443-455, May 2007.

[5] C. Madden, "Tracking people across disjoint camera views," Ph.D. dissertation, Faculty of Information Technology, University of technology, Australia, 2009.

[6] B. Prosser, S. Gong, T. Xiang, and Q. Mary, "Multi-camera matching using bi-directional cumulative brightness transfer functions," in *Proc. British Machine Vision Conf.*, 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, pp. 974-983, Jun. 2007.

[10] C. A. Poynton, *A Technical Introduction to Digital Video*, John Wiley & Sons, Inc., 1996, pp. 175.

Rahman Yousefzadeh received the B.S. degree in computer engineering from Zanjan University, Iran in 2008, and the M.S. degree in computer engineering from Shahrood University of Technology, Shahrood, Iran in 2011. His research interests include computer vision, healthcare robotics, and ambient intelligence. Yousefzadeh is selected as a distinguished student researcher of the year in Shahrood University of Technology, Shahrood, Iran in 2009. Also he is the member of IEEE since 2008.

Hamid Hassanpour received the B.S. degree in computer engineering from Iran University of Science and Technology, Tehran, Iran in 1993, the M.S. degree in computer engineering from Amirkabir University of Technology, Tehran, Iran in 1996, and the Ph.D. from the Queensland University of Technology, Brisbane, Australia in 2004. His research interests include biomedical signal processing, time-frequency signal processing and analysis, new architectures in computer design, text syntax analyzing and image processing.