Targets Association across Multiple Cameras by Learning Transfer Models

Liu Suolan [1,2,3], Wang Jia [2] and Sun Changyin [1*]

[1]*School of Automation, Southeast University, Nanjing 210096, China*
[2]*School of Information Science & Engineering, Changzhou University, Changzhou 213164, China*
[3]*Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA*
liusl@cczu.edu.cn

Abstract

In this paper, we propose a novel method to solve the problem of targets association and tracking across multiple cameras of non-overlapping views. The method is divided into two parts. One is an improvement on appearance transfer model, another is an improvement on spatio-temporal transfer model. To learn inter-camera appearance transfer models, lαβ color space is used to calibrate images. By this way, the overall and local information can be used, which has advantage to color transform correction. To learn spatio-temporal transfer model, entry/exit zones of a non-overlapping topology can be effectively estimated by defining valid link and using clustering method. Then a kind of time constrain is set between two nodes to judge whether there is correlation of observations. Experiments show the effectiveness of the proposed method.

Keywords: targets association; non-overlapping; appearance; spatio-temporal

1. Introduction

Video target tracking is an important research field for its wide use in car parks, supermarkets, banks, factories, mines and other places. In order to meet the needs of monitoring, we can use multiple cameras in these regions, so that FOVS of each camera can be overlapped or adjacent. By camera calibration, we can gain continuous scene, and achieve a comprehensive monitoring of the whole region. However, in practical application, it is not realistic because of high cost and large amount of data. Actually, we only install cameras in key locations, let each camera cover only a limited area. By reasonable configuration, a little number of cameras can form a collaborative monitoring network, which covers a wide area. But there will be many blind monitoring views in different cameras, so that the targets are discrete in time and space. However, each camera is not isolated, the target will transfer from blind monitoring views of different cameras. At the same time, the captured target appearance varies greatly because of environmental illumination, angles, postures and imaging characteristics of each camera. All these cause target association very difficult. Furthermore, target matching and tracking results are affected. Therefore, in cooperative network, target tracking of multiple non-overlapping cameras has become a main research content of wide video surveillance, which plays an important role in behavior recognition [1-2]. If let pedestrian as research target, the current solution is to extract appearance and spatio-temporal feature, and construct spatial and temporal constraint between cameras based on topological relations. Then use suitable association algorithm to match different observations of different cameras, so as to obtain the corresponding relation. Color

---

[*] Corresponding Author

feature is one of the most used appearance features. While a pedestrian moves across different cameras, its color appearance will be affected by various factors. Figure 1 shows pedestrian move across a non-overlapping monitoring network. The appearance features are changed because of variance and changes of illumination, viewpoint, clothing and other related factors.
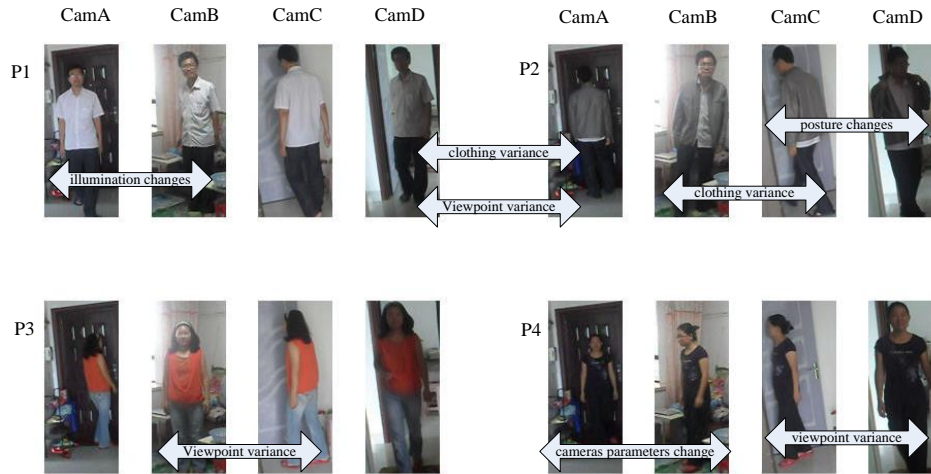

Figure 1. Appearance Features Change while Pedestrians Move Across a Non-overlapping Monitoring Network of Four Cameras from CamA to CamD.

Cheng and Madden do great research on disjoint track matching [3]. Methods based on major color spectrum histogram representation (MCSHR) matching and post-matching integration are proposed. And a two-directional similarity measurement algorithm is used to measure the similarity of the two given observations, which is based on the most similar major color searching algorithm. Furthermore, they also use the 3-5 adjacent frames to construct an incremental MCSHR (IMCSHR) matching model to cope with small pose changes occurring along the track. Capability of distinguish color is strong, its statistical characteristics can be little effected by the shooting angle, but it is easy to be influenced by light and imaging equipment. Therefore, there is much research on color consistency correction of different cameras. Porikli and Divakaran [4-5] present a novel method to solve the problem of inter-camera color calibration. A distance metric and a modeling function are used to evaluate the inter-camera radiometric properties using color histograms. It is a relative non-parametric non-linear color distortion model function of each camera combination. Gilbert and Bowden [6-8] firstly convert targets' color histogram from RGB color space to Munsell, then match them. Experiments show good matching effect of low resolution targets. Javed *et. al.,* [9-10] use conformity in the traversed paths to establish correspondence. The proposed algorithm learns this conformity and hence the inter-camera relationships in the form of multivariate probability density of space–time variables (entry and exit locations, velocities, and transition times) using kernel density estimation. And find that all brightness transfer functions from a given camera to another camera lie in a low dimensional subspace. This subspace is learned by using probabilistic principal component analysis and used for appearance matching. Jeong and Jaynes[11] use target's color degree to construct transfer function, which shows better performance. In addition to color features, many other features can also be used to calculate target association degree. Park proposed a visual search engine (ViSE) as a semi-automatic component in a surveillance system using networked cameras [12]. The ViSE aims to assist the monitoring operation of huge amounts of captured video streams, which tracks and finds people in the video based on their primitive features with the interaction of a human operator. In literatures [13] and

[14], height is used as the matching feature. Human gait can also be used as matching feature. Kang *et. al.,* [15] establish foregrounds' color distribution model and color model of feature point to match targets, which combine shape feature. Gheissari *et. al.,* [16] focuses on research individual reidentification algorithms that use the overall appearance of an individual as opposed to passive biometrics such as face and gait. Yu *et. al.,* [17] present an appearance model for establishing correspondence between tracks of people, which may be taken at different places, at different times or across different cameras. The appearance model is constructed by kernel density estimation. An additional feature of path-length is used. Experimental results demonstrate the important role of the path-length feature in the appearance model and the effectiveness of the proposed appearance model and matching method. Gandhi and Trivedi [18-19] proposed a new Panoramic Appearance Map (PAM) for multi-camera based person re-identification. Global features including clothing color, texture and height are extracted to match.

From the above investigation, we can find that color features are widely used in matching algorithms. And color transfer approaches [18-20] can improve the object recognition accuracy. However, they only do well while given a good training set, which has a suitable range of brightness and color value correspondences are known between training samples. If the given training set is not realizable or correspondences are not highly reliable, we will get poor matching performance. Especially, these methods have very high requirements to illumination condition. Once illumination is changed, some new training samples are needed. Therefore, we may draw a conclusion that color transfer approaches are not applicable to real systems of uncontrollable illumination conditions.

Target's spatio-temporal characteristic has two aspects including spatial relations and temporal relations. In literatures [21] and [22], cameras network is expressed by nodes graph. In camera's FOV, objects appear and disappear regions have been recorded as different nodes. These regions can be obtained by statistical learning of appear and disappear positions. To one camera, there may be multiple appear and disappear regions, which can be regarded as different nodes respectively. All the cameras are synchronized in time. Every observation found in each node has a time stamp. It expresses emergence or vanishing moment, which can be used to establish temporal constraints.
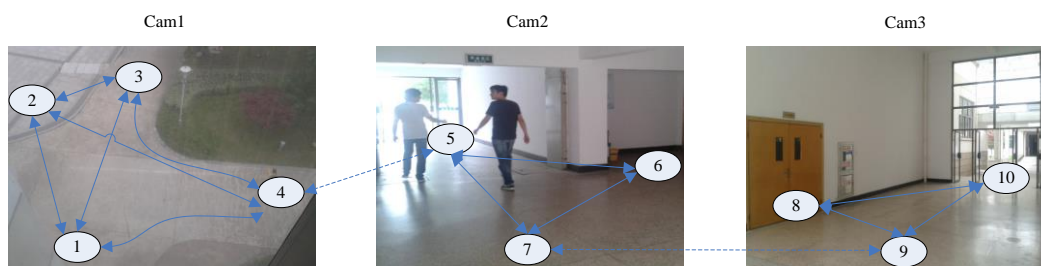

Figure 2. Topology of a non-overlapping multi-camera network. Appear and disappear positions are labeled by different numbers. The blue solid lines denote visible paths in FOV of each camera. The blue dotted lines represent valid link between two cameras.

Javed *et. al.,* [9-10] calculate spatio-temporal transition probability by using a Parzen estimator. These methods often show good estimations of the transition time distributions. However, They are complicated and difficult to solve the problem of correspondences or object tracking. Therefore, appearance feature is usually added to match observations.

In this paper, the above two issues are discussed. Our work is aimed at solving problems in multi-object tracking across non-overlapping views. Inter-camera spatio-

temporal relationships and appearance relationships are improved to provide more information in the process of objects matching.

2. Improved on Appearance Transfer Model

As to target's appearance change, literatures [21-22] modeled a transfer of colors associated with the two specific cameras. In literature [23], brightness transfer function (BTF) is proposed, which assumes that realizable foreground-background segmentation is needed for re-identification. And if appearance changes frequently, the function is not always sufficient. In literature [20], CCT method is proposed by Reinhard to solve the problem of multi-camera object tracking, which is modeled according to the color features of two hand-labeled objects or full images of FOVS. If there is more than one pair of objects, the average value is used.

Based on the work of Reinhard, we also find that the method can do well only for those images with single tone color. Otherwise, blocking process is needed for those images with composed tones color. Welsh uses energy function of L2 metric space to match images of their brightness and texture, and proposed a color transfer algorithm based on the whole [24]. While we use Welsh's method to process images, we find that if we transfer a composed and colorful target image to a single source image, the transferred result is not fit to national color. By further analysis, these two methods do not consider the pixels' correlation of local small zone. In fact, on two-dimensional space, color value of every pixel shows by a certain distribution. It only has relation with its neighborhood pixels, which is labeled as zone ε. The distribution is not affected by pixels out of the neighborhood zone. Therefore, the transferred values are also affected by ε [25]. Based on this consideration, we propose a novel algorithm to solve the problem of appearance variance across cameras, which uses statistical value of local information in the process of color transfer.

Our proposed method includes the following steps:

Step a. Convert Color Space

In surveillance video stream, the captured pedestrian is usually shown by adopting RGB format [26]. RGB spatial model is a linear representation system. It has some characteristics, such as simple and intuitive. However, it is not suitable for human vision. In addition, in RGB space, two points distance is not a linear relationship between the Euclidean distance and the actual colors distance. Thus it easily causes colors error separation. Moreover, there are strong correlations between three channels' of values, and hue, saturation, brightness information can not be directly measured. If a pixel's color information need modified, you must modify the 3 channels at the same time, which will make the operation of color mapping very complex. Ruderman et al. [27] propose a new color space (called lαβ color space) by minimizing correlation between channels for many natural scenes. lαβ is a orthogonal color space, l is brightness channel, α is blue - yellow opponent channel, β is red-green opponent channel. The three channels are perpendicular to each other, and correlation is very small, so modify can be operated on each channels, the target image does not appear visual deformatio. Therefore, to facilitate the color transform and recognition, lαβ color space is used in this paper.

Since lαβ color space is a variant of LMS color space, we firstly need to convert images from RGB color space to LMS color space, and then convert them from LMS color space to lαβ color space. While image is converted to lαβ color space, the step of decoupling is completed.

$$
\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \log(L) \\ \log(M) \\ \log(S) \end{bmatrix}, \text{ where } \begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}
$$

Step b. Calibrate Source Image

In lαβ color space, calculate the mean, overall standard deviation and local standard deviation of the source image and the target image. Let statistical characteristics of the source image's as same as possible to the target image's. That's they have the same or near mean, overall standard deviation and local standard deviation. It includes the following steps: (1) calculate the overall mean value μ, overall standard difference σ and local standard deviation σ^* of each channel in the l, α, β color space, namely the statistical parameters of three channel source image for: $\mu_{ls}, \sigma_{ls}, \sigma_{ls}^*,$ $\mu_{\alpha s}, \sigma_{\alpha s}, \sigma_{\alpha s}^*,$ $\mu_{\beta s}, \sigma_{\beta s}, \sigma_{\beta s}^*$ target image's for: $\mu_{lt}, \sigma_{lt}, \sigma_{lt}^*,$ $\mu_{\alpha t}, \sigma_{\alpha t}, \sigma_{\alpha t}^*,$ $\mu_{\beta t}, \sigma_{\beta t}, \sigma_{\beta t}^*$. Parameters of the subscript s, t represent the source image and the target image, respectively.

$$\mu_{im} = \frac{1}{M*N}\sum_{x=1}^{M}\sum_{y=1}^{N}f(x,y), \ \sigma_{im} = \sqrt{\frac{1}{M*N}\sum_{x=1}^{M}\sum_{y=1}^{N}[f(x,y)-\mu_{im}]^2} \tag{2}$$

$f(x,y)$ is gray value of point (x,y), the image size is $M*N$. A n × m template is used to calculate local standard deviation of sampling points. (2) Values of l,α and β of transferred image can be gained according to formula (1). In each channel of source image, we use every pixel's value subtract its channel's overall mean value, then scale it, and add channel's overall mean value of the target image:

$$l' = [\theta_1 * \frac{\sigma_{lt}}{\sigma_{ls}} + \theta_2 * \frac{\sigma_{lt}^*}{\sigma_{ls}^*}](l_s - \mu_{ls}) + \mu_{lt} \ , \ \alpha' = [\theta_1 * \frac{\sigma_{\beta t}}{\sigma_{\alpha s}} + \theta_2 * \frac{\sigma_{\alpha t}^*}{\sigma_{\alpha s}^*}](\alpha_s - \mu_{\alpha s}) + \mu_{\alpha t} \ ,$$

$$\beta' = [\theta_1 * \frac{\sigma_{\beta t}}{\sigma_{\beta s}} + \theta_2 * \frac{\sigma_{\beta t}^*}{\sigma_{\beta s}^*}](\beta_s - \mu_{\beta s}) + \mu_{\beta t} \tag{3}$$

θ_1 and θ_2 are two nonnegative constant, satisfying $\theta_1 + \theta_2 = 1$. l', α' and β' are result image's lαβ value.

Step c. Getting Transferred Image.

Based on step b, we can get a calibration of source image in lαβ color space. Now, we need to transfer it from lαβ color space to RGB color space, thus result image based on overall and local color transform correction is gained, which is named as transferred image.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.46791 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}, \ \text{where} \ \begin{bmatrix} \log(L) \\ \log(M) \\ \log(S) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} L \\ \alpha \\ \beta \end{bmatrix} \tag{4}$$

Since the source image and target image are equal, they can be exchanged to get different tranferred images by using our proposed appearance transfer model.
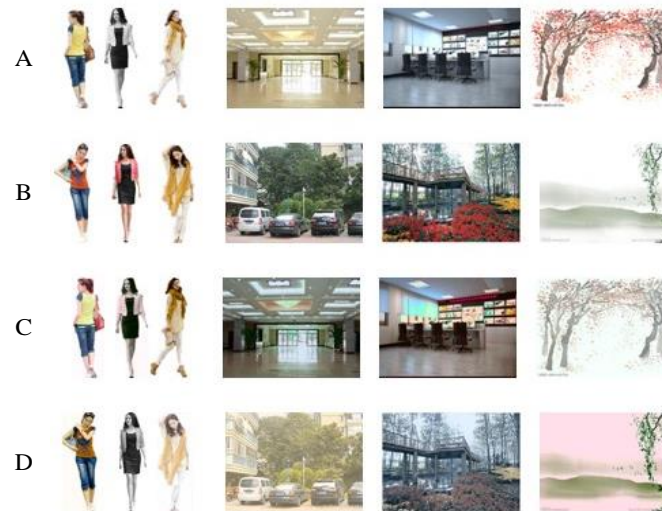
Figure 3. Some Examples of Processed Image using Different Kinds of Images

Figure 3 shows some experiments of appearance transfer. In this experiment, we select three kinds of images to test our method. The first three images are pededstrians, the later are two natural scenes, the last one is a painting. Row C shows color transfer result while row A is used as the original source image and row B is the target image. Row D shows color transfer result while row B is used as the original source image and row A is the target image. To pededstrians, we can see that there are obviously changes in their clothes and accessories. In row C, the two natural scenes show abundant colors, the level sense is stronger than before. Because of color transfer, the painting styles are changed, bring people colorful feelings.

3. Improved on Spatio-temporal Transfer Model

In literature [28], every edge of camera network topology graph represents that target may move across the adjacent two nodes. Therefore, given a pair of nodes $(Node_i^A, Node_j^B)$ from two cameras C_A and C_B , where $Node_i^A \in C_A$, $Node_j^B \in C_B$, $C_A \neq C_B$, there is a variable $LINK(Node_i^A, Node_j^B) = \{0,1\}$, 0 denotes the target can not move across the two nodes directly, unless through other nodes; 1denotes the target can move across the two nodes directly. Note that, $LINK(Node_i^A, Node_j^B) = 1$ does not mean that the target disappeared in one node will reappear in another node, only represents the probability. It may also do not reappear because of other reasons such as walking into a building or turning a corner. Figure 4 is a geometric model of the camera environment showing the relative placement of the seven cameras. Red ellipses are entry/exit zones, which can be gotten by clustering the starting and ending points in single camera monitor.
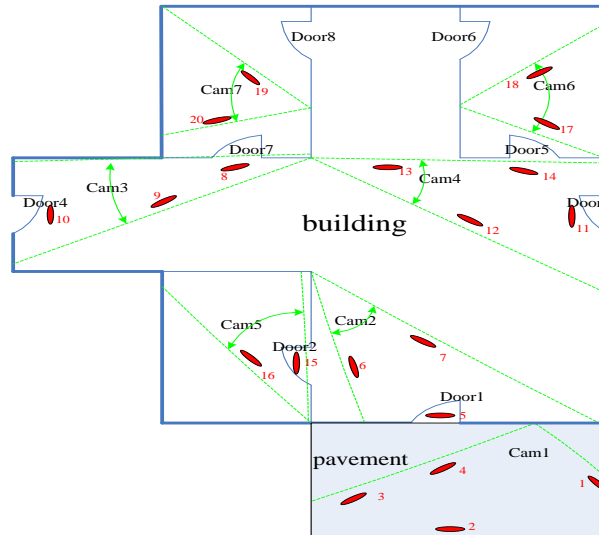
Figure 4. Geometric Model of the Camera Environment Showing the Relative Placement of the Seven Cameras of Figure 2. Entry/Exit Zones are Shown as Red Ellipses.

Figure 5 is topology structure of surveillance network. Actually, the transform relation of each camera is more complex than shown in the Figure 5. In our research, more attention is paid to the relation between cameras. Therefore, only part of the transform relationship is labeled as references.
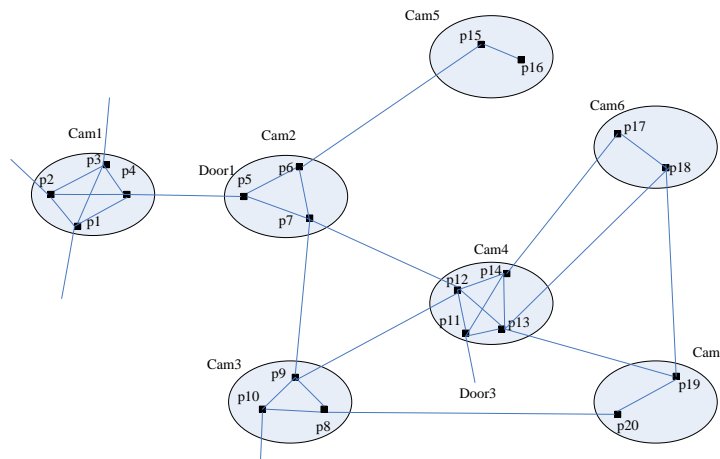

Figure 5. Topology Structure of Surveillance Network of Figure 4.

As to time estimation between cameras, we can use Gauss model. Based on literature [28], by statistics the time of different targets through two adjacent nodes, we get a time sequence $t = (t_1, t_2, \cdots, t_n)$. This time sample can be gained by manual mode, or automatic statistics by using target association algorithm. If we use automatic statistical method, for all targets, the time sequence is considered at the initial time. After a period of time, we can establish time distribution model according to target association results. If we use Gauss distribution method, distribution parameters of μ and σ^2 can be gained by maximum likelihood estimation.

$$\mu_i = \frac{1}{n}\sum_{i-1}^{n} t_i \quad , \quad \sigma_i^2 = \frac{1}{n}\sum_{i-1}^{n}(t_i - \mu)^2 \tag{5}$$

Since a new target may appear, we need update the parameters of Gauss distribution according to new target's transition time. Let t_{i+1} be transition time of target $(i+1)^{th}$, ξ be update rate, $0 \leq \xi \leq 1$, then we can construct a update formula as follows:

$$\mu_{i+1} = (1-\xi)\mu_i + \xi t_{i+1} \ , \ \sigma_{i+1}^2 = (1-\xi)\sigma_i^2 + \xi(t_{i+1} - \mu_{i+1})^2 \tag{6}$$

To nodes $Node_i^A$ and $Node_j^B$, we observe objects departing at node i and arriving at node j within a time window which is long enough. Let $DEP_i^A(t)$ and $ARR_j^B(t)$ denote the departure time sequence at node i and arrival time sequence at node j respectively. Set $T(OBJ_i^A(t), OBJ_j^B(t+\Delta t))$ as time transfer relation.

Some researchers calculate time transfer probability according to target appearing time and disappearing time. This method is not reliable, because the motion is not stable. Especially, it is difficult to estimate the state of motion in the blind zone. Therefore, the probability value is insufficient to be trusted. Based on this method, we improve it by setting time constrain, and for every target satisfying the constrain can be considered having correlation, which is constructed by extracting features. According to the characteristics of Gauss distribution, if the variable t meets Gauss distribution of parameters (μ, σ^2), the variable $y = (t-\mu)/\sigma$ meets the standard normal distribution. We need set a threshold range $(-\lambda, \lambda)$, making the probability of samples falling within the scope, which is as close as possible to 1. Based on normal distribution table, while λ is 3, $p(-\lambda < y < \lambda) \approx 0.997$, it can meet our requirement. To time distribution t, the constrain range is $(\mu - \lambda\sigma, \mu + \lambda\sigma)$. Therefore, $T(OBJ_i^A(t), OBJ_j^B(t))$ can be expressed as:

$$T(OBJ_i^A(t), OBJ_j^B(t+\Delta t)) = \begin{cases} 1, \left| ARR_j^B(t+\Delta t) - DEP_i^A(t) - \mu \right| \leq 3\sigma \\ 0, \left| ARR_j^B(t+\Delta t) - DEP_i^A(t) - \mu \right| > 3\sigma \end{cases} \tag{7}$$

Then, an improved spatio-temporal transfer model can be computed as follows:

$$ST(OBJ_i^A(t), OBJ_j^B(t+\Delta t))$$
$$= SIM(OBJ_i^A(t), OBJ_j^B(t+\Delta t)) LINK(Node_i^A, Node_j^B) T(OBJ_i^A(t), OBJ_j^B(t+\Delta t))$$
$$\tag{8}$$

$SIM(*)$ measures the similarity between each object in $OBJ_i^A(t)$ and each object in $OBJ_j^B(t)$. To measure this similarity, any effective feature can be used.

4. Experimental Results

Experiments are done to give a thorough performance test of our proposed method. They are conducted on real data collected from off-line videos to demonstrate the effectiveness. Figure 4 shows the real-life experimental network. It has seven non-overlapping cameras, containing four, three, three, four, two, two and two entry/exit zones respectively. According to dividing method of literature [28], the network can be divided into the following independent parts: $\{p4, p5\}$, $\{p6, p15\}$, $\{p7, p9, p12\}$, $\{p8, p20\}$, $\{p14, p17\}$, $\{p13, p18, p19\}$. Twenty-minute-long videos are used to test. We use Gaussian Mixture Model to detect every person arriving and leaving every node, and the arrival time and departure time are also recorded. Time window is large enough.

Distance between two correlated nodes and average transition time are shown in Table 1. Let average transition time as prior knowledge, statistic of initial time can be omitted. Set initial time constraint as $(\mu/2, 3\mu/2)$.

Table 1. Parameters between Two Nodes

nodes	distance(m)	average transition time(s)	initial time constrant
p4 and p5	2	1.7	(0.83, 2.50)
p6 and p15	3	2.5	(1.25, 3.75)
p7 and p12	7	5.1	(2.55, 7.66)
p7 and p9	24	17.4	(8.70, 26.09)
p9 and p12	26	18.8	(9.42, 28.26)
p8 and p20	4	3.3	(1.67, 5.00)
p14 and p17	3	2.5	(1.25, 3.75)
p13 and p18	22	20	(10.00, 30.00)
p13 and p19	21	19.1	(9.55, 28.64)
p18 and p19	26	23.6	(11.82, 35.45)

In this part, a 140-second video is randomly captured to do experiments. We find that there are 7 people walking in the camera network. Since the number of person remains small, specific time constraints between nodes are not needed to set. In each unit, we only need to do targets association according to the temporal relationship.
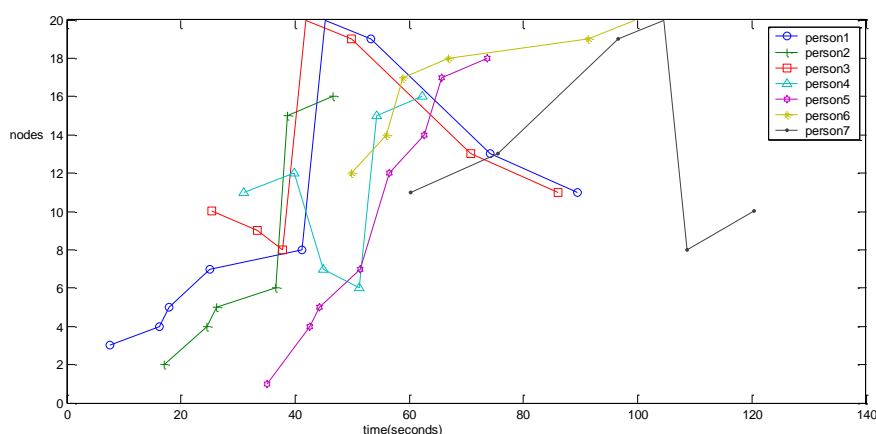

Figure 6. Association Results of Different Person

By using the association characteristics proposed in this paper, we can get satisfied matching results. Figure 6 shows the results of targets association. The horizontal axis expresses one target's emergence or vanishing moment in a node. The vertical axis expresses the number of camera nodes. 7 walking trajectories show different transfer relationship in camera networks. We use different graphic symbols to express different person.

5. Conclusion

We have presented a targets association method across multiple cameras by learning cameras' transfer model, which is designed to detect and track pedestrian trajectory from realistic videos. The inter-camera appearance transfer model measures a similarity between two pedestrians by processing images in lαβ color space. The spatio-temporal relationship is computed by defining valid link, using clustering method and setting time constrain. Through the experiments with our new dataset, we have confirmed that our

proposed method is able to track multiple targets across non-overlapping cameras based on transfer models. Our future work will focus on recognize complicated human activities in non-overlapping cameras [29].

Acknowledgements

References

[1] T. L. Song, "Multi-target tracking filters and data association: A survey [J]", Journal of institute of control, robotics and systems, vol. 20, no. 3, **(2014)**, pp. 313-322.

[2] A. Dallil, M. Oussalah and A. Ouldali, "Sensor fusion and target tracking using evidential data association[J]", IEEE Sensors Journal, vol. 13, no. 1, **(2013)**, pp. 285-293.

[3] C. Madden, E. D. Cheng and M. Piccardi, "Tracking people across disjoint camera views by an illumination-tolerant appearance representation[J]", Machine Vision and Applications, vol. 18, no. 3, **(2007)**, pp. 233-247.

[4] F. Porikli, "Inter-camera color calibration using cross-correlation model function[C]", In IEEE International Conference on Image Processing, **(2003)**, pp. 133-136.

[5] F. Porikli and A. Divakaran, "Multi-camera calibration, object tracking, and query generation[C]", In IEEE International Conference on Multimedia and Expo, **(2003)**, pp. 653-656.

[6] A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity[C]", In proceedings of Europeon Conference on Computer Vision, **(2006)**, pp. 125-136.

[7] A. Gilbert and R. Bowden, "Incremental modelling of the posterior distribution of objects for inter and intra camera tracking[C]", In British Machine Vision Conference, **(2005)**.

[8] R. Bowden and P. Kaewtrakulpong, "Towards automated wide area visual surveillance: tracking objects between spatially – separated, uncalibrated views[J]", IEEE Proceedings - Vision, Image and Signal Processing, vol. 152, no. 2, **(2005)**, pp. 213-223.

[9] O. Javed, K. Shafique and Z. Rasheed, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views[J]", Computer Vision and Image Understanding, vol. 109, no. 2, **(2007)**, pp. 146-162.

[10] O. Javed, K. Shafique and M. Shah, "Appearance modeling for tracking in multiple non-overlapping cameras[C]", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(2005)**, pp. 26-33.

[11] K. Jeong and C. O. Jaynes, "Object matching in disjoint cameras using a color transfer approach[J]", Machine Vision and Applications, vol. 19, no. 5, **(2008)**, pp. 443-455.

[12] U. Park, A. K. Jain and I. Kitahara, "ViSE: Visual search engine using multiple networked cameras [C]", In 18th International Conferences on Pattern Recognition, **(2006)**, pp. 1204-1207.

[13] C. Benabdelkader and L. Davis, "Estimation of anthropomeasures from a single calibrated camera[C]", 7th International Conference on Automatic Face and Gesture Recognition, **(2006)**, pp. 499-504.

[14] B. Song and A. K. Roy-Chowdhury, "Stochastic adaptive tracking in a camera network[C]", In IEEE International Conference on Computer Vision, **(2007)**, pp. 1-8.

[15] J. Kang, I. Cohen and G. Medioni, "Persistent objects tracking across multiple non-overlapping cameras [C]", In IEEE Workshop on Motion and Video Computing, **(2005)**, pp.112-119.

[16] N. Gheissari, T. B. Sebastian and P. H. Tu, "Person reidentification using spatiotemporal appearance [C]", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(2006)**, pp. 1528-1535.

[17] Y. Yu, D. Harwood and K.. Yoon, "Human appearance modeling for matching across video sequences[J]", Machine Vision and Applications, vol. 18, no. 3, **(2007)**, pp. 139-149.

[18] T. Gandhi and M. M. Trivedi, "Panoramic Appearance Map (PAM) for multi-camera based person re-identification[C]", In IEEE International Conference on Advanced Video and Signal Based Surveillance, **(2006)**, pp.78-83.

[19] T. Gandhi and M. M Trivedi, "Person tracking and reidentification: Introducing Panoramic Appearance Map (PAM) for feature representation[J]", Machine Vision and Applications, vol. 18, no. 3, **(2007)**, pp. 207-220.

[20] R. Erik, A. Michael, G. Bruce and S. Peter, "Color transfer between images[J]", IEEE Computer graphics and applications, **(2001)**, pp. 34-41.

[21] D. Makris, T. J. Ellis and J. Black, "Bridging the gaps between cameras[C]", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), **(2004)**, pp. 205-210.

[22] X. T. Chen, K. Q. Huang and T. N. Tan, "Object tracking across non-overlapping views by learning inter- camera transfer models [J]", Pattern Recognition, vol. 47, no. 3, **(2014)**, pp. 1126-1137.

[23] A. Datta, L. M. Brown and R. Feris, "Appearance modeling for person re-identification using weighted brightness transfer functions[C]", 21st international conference on pattern recognition, **(2012)**, pp. 2367-2370.

[24] T. Welsh, M. Ashikhmin and K. Mueller, "Transferring color to greyscale images[C]", In: Proceedings of SIGGRAPH, San Antonio, USA, **(2002)**, pp. 277-280.

[25] A. E. Alexei and K. L. Thomas, "Texture synthesis by non-parametric sampling[C]", In: Proceedings of the 7th IEEE International ConferenceonComputer Vision, Kerkyra, Greece, **(1999)**, pp. 1033-1038.

[26] N. B. Wang, X. J. Gong and J. L. Liu, "A new depth descriptor for pedestrian detection in RGB-D images [J]", **(2012)**, ICPR: 3688-3691.

[27] D. Ruderman, T. Cronin and C. Chiao, "Statistics of cone responses to natural images: implications for visual coding[J]", Journal of the optical society of America, vol. 15, no. 8, **(1998)**, pp. 2036-2045.

[28] S. H. Liu, "Moving objects detection and tracking in non-overlapping surveillance camera network [D]", National university of defense technology, **(2009)**.

[29] M. Xue, S. B. Zheng and H. Yang, "Online visual tracking based on selective sparse appearance model and spatiotemporal analysis[J]", Optical Engineering, vol. 53, no. 1, **(2014)**, pp. 1-15.

Authors

Liu Suolan, She Associate Professor, is currently Postdoctoral research fellow of school of Automation, Southeast University. She received her Ph.D. in computer application technology from Nanjing University of Science and technology in 2008. Her research interests are in pattern recognition, image analysis and computer vision.

Wang Jia, She received her Bachelor of Science degree in Computer Science in June 2012 at Changzhou Univ., China. Now, she is a postgraduate in Computer Science at Changzhou Univ. Her main research interests are Digital Image Processing and Video Surveillance.

Sun Changyin, He is currently Professor of School of Automation, Southeast University. He received his Ph.D. in Electrical Engineering from Department of Automation Control, Southeast University in 2003. His research interests are in intelligent control, neural networks, theory and design of intelligent control systems, pattern recognition.