Object Tracking by Non-overlapping Distributed Camera Network

Pier Luigi Mazzeo, Paolo Spagnolo, and Tiziana D'Orazio

Istituto di Studi sui Sistemi Intelligenti per l'Automazione, C.N.R.
Via G. Amendola 122/D 70126 Bari, Italy
{mazzeo,dorazio,spagnolo}@ba.issia.cnr.it
http://www.issia.cnr.it/

Abstract. People Tracking is a problem of great interest for wide areas video surveillance systems. In these large areas, it is not possible for a single camera to observe the complete area of interest. Surveillance systems architecture requires algorithms with the ability to track objects while observing them through multiple cameras. We focus our work on multi camera tracking with non overlapping fields of view (FOV). In particular we propose a multi camera architecture for wide area surveillance and a real time people tracking algorithm across non overlapping cameras. In this scenario it is necessary to track object both in intra-camera and inter-camera FOV. We consider these problems in this paper. In particular we have investigated different techniques to evaluate intra-camera and inter-camera tracking based on color histogram. For the intra-camera tracking we have proposed different methodologies to extract the color histogram information from each object patches. For inter-camera tracking we have compared different methods to evaluate the colour Brightness Transfer Function (BTF) between non overlapping cameras. These approaches are based on color histogram mapping between pairs of images of the same object in different FOVs. Therefore we have combined different methodology to calculate the color histogram in order to estimate different colour BTF performances. Preliminary results demonstrates that the proposed method combined with BTF outperform the performance in terms of matching rate between different cameras.

1 Introduction

The specific problem we address in this paper is the surveillance over wide-areas such as an airport, the downtown of a large city or any large public area. Surveillance over these wide areas consists of the search for suspicious behavior as persons loitering, unauthorized access, or persons attempting to enter a restricted zone. Currently, these surveillance tasks are accomplished by human operators who continually observe monitors to detect unauthorized activity over many cameras. Recent researches have demonstrated that the attention level drastically drops after few hours, so it becomes highly probable that suspicious activity would go unnoticed by a human operator. A computer vision system,

J. Blanc-Talon et al. (Eds.): ACIVS 2009, LNCS 5807, pp. 516–527, 2009.

however, can monitor both immediate unauthorized behavior and long-term suspicious behavior. The system would then alert a human operator for a closer look. In most cases, it is not possible for a single camera to observe the complete area of interest because sensor resolution is finite and structures in the scene limit the visible areas.

In realistic scenarios, surveillance systems which cover large areas are composed by multiple cameras with non overlapping FOVs. In this paper we investigate on the feasibility of different techniques to evaluate intra-camera and inter-camera tracking based on color histogram. As first step we have evaluated different techniques to extract the color histogram from each foreground patches. After that, we have implemented tracking algorithms that by applying inter camera appearance matching are able to track people across multiple view separated by blind regions. We have compared different methods to evaluate the color Brightness Transfer Function (BTF) between non overlapping cameras. These approaches are based on the color histogram mapping between pairs of images of the same object in different FOVs. The experimental results show how the calculated transfer function, combined by different color histogram extraction approach, essentially improves the matching rate between non overlapped regions.

The remaining of this paper is organized as follows. In section 1.1 we give a brief overview of related work. The approaches for people tracking across non overlapping cameras are summarized in section 2. The color histogram extraction and tracking approaches are described in section 2.1. The sections 2.2 and 2.3 describe the multi-cameras correspondence detection and the multi-cameras people tracking approach. In section 3 we report the results obtained on some video sequences acquired with two cameras placed in different rooms.

1.1 Related Work

Most of the approaches presented in literature suppose the use of calibrated cameras and the availability of a site model. In [1] the conformity in the traversed paths of people and car is used to establish correspondence among cameras. The algorithm learns this conformity and hence the inter-camera relationships in the form of multivariate probability density of spacetime variables (entry and exit locations, velocities, and transition times) using kernel density estimation. To handle the appearance change of an object as it moves from one camera to another, the authors demonstrate 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. In [3] particle filters and belief propagation are combined in a unified framework. In each view, a target is tracked by a dedicated particle-filter-based local tracker. The trackers in different views collaborate via belief propagation so that a local tracker operating in one view is able to take advantage of additional information from other views. In [4] a target is tracked not only in each camera but also in the ground plane by individual particle filters. These particle filters collaborate in two different ways. First, the particle filters in

each camera pass messages to those in the ground plane where the multi-camera information is integrated by intersecting the targets principal axes. This largely relaxes the dependence on precise foot positions when mapping targets from images to the ground plane using homographies. Secondly, the fusion results in the ground plane are then incorporated by each camera as boosted proposal functions. A mixture proposal function is composed for each tracker in a camera by combining an independent transition kernel and the boosted proposal function.

Kalman filters are used in [5] to robustly track each targets shape and motion in each camera view and predict the targets track in the blind region between cameras. For multi-camera correspondence matching, the Gaussian distributions of the tracking parameters across cameras for the target motion and position in the ground plane view are computed. Matching of targets across camera views uses a graph based track initialization scheme, which accumulates information from occurrences of target in several consecutive frames of the video. Geometric and intensity features are used in [7] to match objects for tracking in a multiple calibrated camera system for surveillance. These features are modelled as multivariate Gaussians, and the Mahalanobis distance measure is used for matching. A method to match object appearances over non-overlapping cameras is presented in [8]. In his approach, a brightness transfer function (BTF) is computed for every pair of cameras. Once such a mapping is known, the correspondence problem is reduced to the matching of transformed histograms or appearance models. However, this mapping, i.e., the BTF varies from frame to frame depending on a large number of parameters that include illumination, scene geometry, exposure time, focal length, and aperture size of each camera. Thus, a single pre-computed BTF cannot usually be used to match objects for moderately long sequences. An unsupervised approach to learn edge measures for appearance matching between non-overlapping views has been presented in [9]. The probability of two observations from two cameras being generated by the same or different object is computed to perform the matching. The main constraint of this approach is that the edge images of vehicles have to be registered together. Note that this requirement for registering object images could not be applicable for non-rigid objects like pedestrians. A Cumulative Brightness Transfer Function (CBTF) is proposed [10] for mapping color between cameras located at different physical sites, which makes use of the available color information from a very sparse training set. A bi-directional mapping approach is used to obtain an accurate similarity measure between pairs of candidate objects. An illumination-tolerant appearance representation, based on online k-means color clustering algorithm is introduced in [11], which is capable of coping with the typical illumination changes occurring in surveillance scenarios. A similarity measurement is also introduced to compare the appearance representation of any two arbitrary individuals. In [12] the distortion function is approximated as general affine and the object appearance is represented as a mixture of Gaussians. Appearance models are put in correspondence by searching a bijection function that maximizes a metric for model dissimilarity.

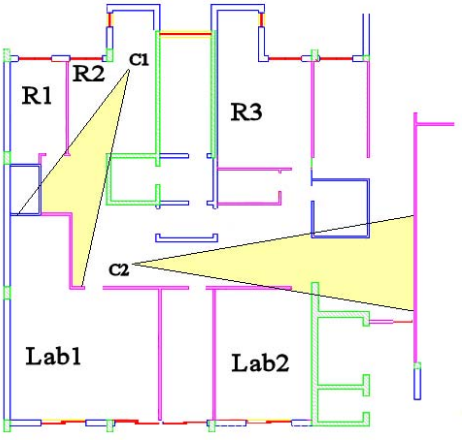


Fig. 1. The cameras configuration inside our office building

A common characteristic of the above related works is that the knowledge of model sites and particular camera positions in various scenarios allow the usage of geometrical and temporal constraints on the entry/exit areas of the image. In this way the appearance matching among different cameras is carried out on a sets of individuals that are candidate by their positions to be observed by distributed cameras.

2 Tracking in Multiple Cameras with Disjoint Views

The focus of this paper is the investigation of a multi-camera tracking system with non overlapping cameras. Two cameras C1 and C2 are arranged in two different places of a corridor as shown in figure 1. The cameras' field of views cover different non-overlapping areas. People observed in camera C2 can take a path across camera C1 turning right or also turning left in the Lab1 without entering the C1 field of view. In the same way people coming from the Lab1 are observed in camera C1 without passing trough the camera C2. The task of the multi-camera tracking algorithm is to establish correspondence across cameras finding which tracks belong to the same object. Because of the cameras' positions, it is not always possible to use space time constraints between the locations of the exits and entrances between the cameras. People can take many paths across C1 and C2, producing the same observations in the two cameras. For this reason in this paper we want to investigate the feasibility of a multi-camera tracking algorithm that relies just on the appearances of the objects in the two cameras. Anyway, the lack of entry/exit constraints renders more difficult the association task. Considering that the color distribution of an object can be fairly different when it moves in a single camera, matching appearances between different cameras is still more difficult, then it is necessary to find out

the transformation that maps the appearance of an object in one camera with its appearance in the other camera. In this paper we considered a training phase during which known objects pass trough both the cameras and their appearances is used to estimate a Brightness Transfer Function (BTF). During this phases we tested two different BTFs, ie. the mean BTF and the cumulative BTF. In the testing phase the object matches were carried out choosing those that produced the lowest values of the Bhattacharya distance between the color histograms of the considered person in one camera with all the possible persons that had travelled through the second camera.

2.1 Color Histogram Extraction and Tracking

The implemented tracker models the appearance of the target object using color histogram. The task of finding the same object from the foreground region in current frame is formulated as follows: the color histogram feature is assumed to have a density function, while a candidate region also had a color histogram feature distributed by a certain density. The problem is to find a candidate region whose associated a density is most similar to the target density. A Bhattacharraya coefficient measure is used as metric between the distribution. We have implemented different method to extract the color histogram from the foreground patches, in order to eliminate noise and possible shadow from the object patch. One or more elliptic masks are used to reach this aim. The ellipse parameter (major and minor axis) are based on the patch dimension and on the distance of the object from the camera. Basing on the person position in the FOV we have estimated, using a mapping function, his body measure in the foreground patches. In this way we try to build the elliptic masks in order to catch more useful information possible. In order to obtain discriminant color histogram we discard any possible part of the patch that could be confuse the distribution of the histogram. In this way the ellipses are drawn to cover the most part of the people body (we cut the head of the person and his eventual shadow as shown in figures 2(b), 3(b)). In particular we have compared different combination of these masks (see pictures 2(b), 2(c), 2(e), 2(f), 3(b), 3(c), 3(e), 3(f)) in order to evaluate their performance.

2.2 Establishing Correspondence Across Multiple Cameras

The correspondence problem occurs when an object enters the FOV of a camera. We need to determine if the object is already being tracked by another camera or it is a new object in the environment. Many approaches are possible to evaluate the brightness transfer function among different cameras. In this paper we compare a mean BTF approach with the cumulative BTF proposed in [10]. Suppose an object O enters the camera C1. In figure 2 we show some images of the tracks of two persons in the camera C1, while in figure 3 the same persons are observed by the camera C2.

Fig. 2. Six images of two persons in the camera C1. a) Foreground patch extracted of the first person; b) Elliptic mask of the first person; c) Double Elliptic masks of the first person; d) Foreground patch extracted of the second person; e) Elliptic mask of the second person; f) Double Elliptic masks of the second person.

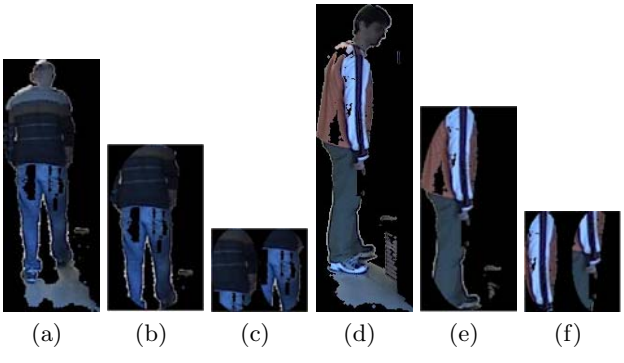


Fig. 3. Six images of two persons in the camera C2. a) Foreground patch extracted of the first person; b) Elliptic mask of the first person; c) Double Elliptic masks of the first person; d) Foreground patch extracted of the second person; e) Elliptic mask of the second person; f) Double Elliptic masks of the second person.

We evaluated the three channels RGB histograms for each image in the tracks of the camera C1. We did the same also for the tracks in the camera C2. The histograms are generated using all the 256 bins for each channel. We want to estimate a BTF $f_{1,2}$ between the cameras C1 and C2 such that, for each couple of images (i_1, j_2), given the brightness values $B^{i_1}(k)$ and $B^{j_2}(k)$ we have $B^{j_2}(k) = f_{1,2}(B^{i_1}(k))$ where $k = 0, .., 255$ represents the number of bins, $i_1 = 1, .., M$ represents the number of images in the camera C1 tracks, $j_2 = 1, .., N$ the number of images in the camera C2 tracks. In order to evaluate the BTF $f_{1,2}$ we collected a total on $N + M$ histograms obtained on the N images tracked in the camera C1 and on the M images of the same person tracked in the camera

C2. We denote the histograms obtained in the two cameras with H_{i_1} and H_{j_2} respectively. For each possible couple of histograms (i_1, j_2) we evaluated the brightness transfer function

$$f_{i_1 j_2}(B^{i_1}(k)) = B^{j_2}(k) \tag{1}$$

using the inverted cumulative histogram, that is

$$f_{i_1 j_2}(B^{i_1}(k)) = H_{j_2}^{-1}(H_{i_1}(B^{j_2}(k))) \tag{2}$$

and finally the mean BTF (referred in the following section as MBTF) $\overline{f}_{1,2}$

$$\overline{f}_{1,2} = \sum_{i_1=1}^{M} \sum_{j_2=1}^{N} f_{i_1 j_2} \tag{3}$$

We evaluated also a cumulative BTF (CBTF) as proposed in [10]. The generation of the CBTF involves an amalgamation of the training set before computing any BTFs. An accumulation of the brightness values is computed on all the training images of the camera C1 obtaining a cumulative histogram $\widehat{H_1}$. The same is done for all the corresponding training images of the camera C2 obtaining $\widehat{H_2}$. The CBTF $\widehat{f}_{1,2}$ is

$$\widehat{f}_{1,2}(B^1(k)) = \widehat{H_2}^{-1}(\widehat{H_1}(B^2(k))) \tag{4}$$

also in this case evaluated by using the inverted cumulative histogram.

2.3 Multi-camera Tracking Using BTFs

In order to solve the multi-camera people identification problem we have to choose among a set of possible correspondence hypotheses the one that produces the best match. Anyway, since our camera configuration allows people to enter into one camera field of view without passing through the other camera, we consider also the problem of finding a proper method to discard false matches. The common method to match person appearances is by estimating the similarity between color histograms. Let be $\{H^1_{k_1}, H^1_{k_2}, ...H^1_{k_{N_k}}\}$ the N_k histograms of the k-th person in the camera C1. Suppose that we have $k = P$ persons moving in the camera C1. When a new observation is taken in the camera C2 we have to decide either if it could be associated with one among the P persons moving in the camera C1 or if it is a new person entering the scene. For each person in the camera C1, i.e. $k = 1, .., P$, we evaluated the mean color histograms among the N_k observations of the k-th person obtaining the mean histograms $\overline{H}^1_k$. Anyway the mean histograms cannot be compared with those obtained by the camera C2 unless the transformation with the BTFs are applied. By using the BTFS described in the previous section we projected the K mean histograms in the new space as follows

$$\breve{H}^{12}_k = \overline{f}_{1,2}(\overline{H}^1_k) \tag{5}$$

where $\breve{H}_k$ represents the new histogram obtained using the mean BTF described in equation 3 and

$$\tilde{H}_k^{12} = \widehat{f}_{1,2}(\overline{H}_k^1) \tag{6}$$

is the transformation by using the CBTF described in 4. Let be $H_{l_1}^2$ the histogram of the first observation in the camera C2. We evaluated the similarity between couple of histograms by using the well known Bhattacharya distance (Sim_B). The association is done with the k-th person that produces the minimum distance, i.e.

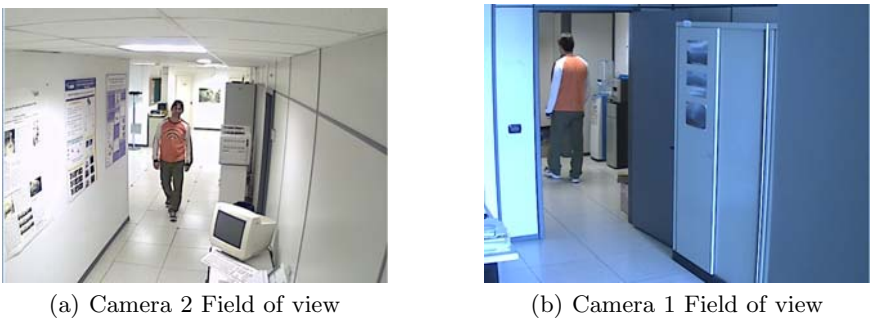
$$min_k(Sim_B(H_{l_1}^2, H_k^{12}) \tag{7}$$

3 Experimental Results

Different experiments were carried out to test the multi-camera tracking algorithm. The scenario was composed by two cameras located in two different points of our office. We used two wireless Axis IP camera with 640x480 color jpg resolution with an acquisition frame rate of 10 frames per second. The topology of this camera network is shown in figure 1, while two images acquired by the two cameras are shown in figures 4(a) and 4(b). Note that the illumination conditions and color quality vary greatly between these views.

We have divided the experiments in two different parts. In the first part we evaluate different kind of method to extract the color histogram from each foreground patche. In the second part we have evaluated different approaches to establish the correspondence across the disjoint views.

In both parts we have used the same data set. Our data-set consisted of synchronized mjpeg videos acquired simultaneously by two different cameras containing six persons. The patches data come from a single camera object detection method described in [13]. The data set were obtained by extracting people from the entire FOV of each camera. We didn't consider any geometrical constraint on the exiting and entering areas of people moving in the scenario.



(a) Camera 2 Field of view (b) Camera 1 Field of view

Fig. 4. Frames from both camera views. The same person walks from camera 1 to camera 2.

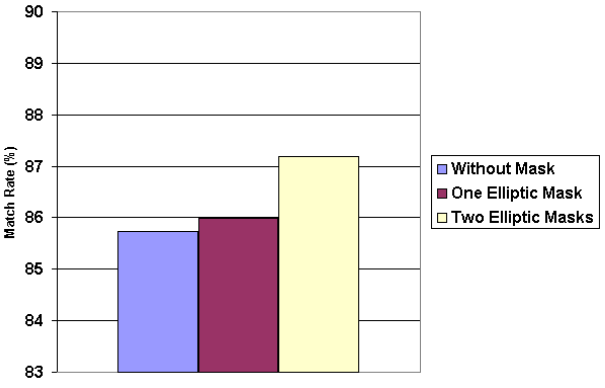


Fig. 5. A comparison of the matching success in the same FOV (Intra-camera) using different color histogram extraction method

We carried out different experiments using sets of samples as follows:

- First experiment: six individuals giving 889 appearance sample (coming from the same FOV) are used as testing data in order to evaluate the performance of different color histogram extraction methods (See section 2.1)
- In the second experiment we need a training and a testing data set: three individuals giving 561 appearance samples in both views were used in the training step, while four individuals with 493 appearance samples in both views were used in testing step (Note that in this case we added one person in the training phase and the test set remained unchanged);

In the figure 5 the results relative to the intra-camera histogram color tracking are presented. As explained in section 2.1 we have evaluated two different approaches to estimate the color histogram from extracted people patches. The similarity, between color histogram features belonging to different foreground patch, has been measured by means of Bhattacharyya distance. The lowest value of these distance among the patch and all the possible candidates (six different persons in the same FOV) determines the tracking association. It is possible to notice how the two elliptic masks approach gives better result in term of match rate. By using mask with two ellipses, in fact, it is possible to preserve the color histogram spatial information. In this way, the upper body and lower body color histograms of each patch are compared with the correspondent parts of another person (see figure 3(f), 3(c)). Results confirm that this color histogram extraction approach discriminates better among the different possible individual candidates.

In the figure 6, the results relative to the tracking across different cameras, are shown. The experiments consist of a training phase and a testing phase. During the training phase we supposed that the correspondence between the same object in the different cameras' FOV were known and this information was used

to calculate the Mean Brightness transfer function (MBTF) and the Cumulative Brightness Transfer Function (CBTF). In the testing phase the correspondences between cameras were computed using the Bhattacharya distance, and the correct match was associated with the lowest value of these distances (see equation 7) among all the possible couple of candidates.

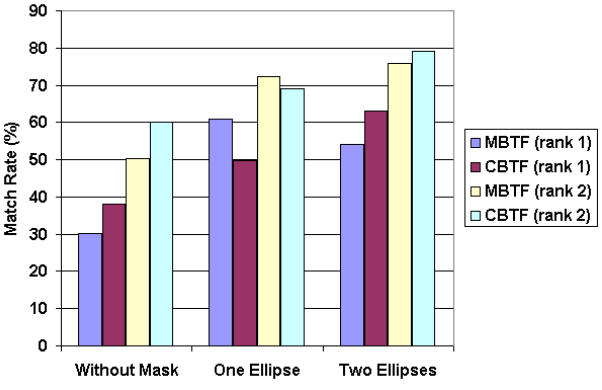


Fig. 6. A comparison of the matching success rate in establishing correspondence across cameras using varying color histogram extraction method and different Brightness Transfer Function Estimation (BTF)

As it can be noticed, in the testing phase we considered four individuals that were present in both views.

As explained in section 2.2 we tested two different approaches to calculate the BTFs between the two cameras: the MBTF and the CBTF. In order to compare these methods we used a uni-directional comparison using the Bhattacharya distance as similarity measure. For each individual we converted his RGB histograms into the target region color space (i.e. from camera 1 to camera 2). They have been compared against all individuals observed in this region. In particular we estimated the mean color histogram for the same individual in each view and we compare each converted histogram against it.

In figure 6 we report both the rank1 and rank2 results indicating the presence of the correct match as the highest and the second highest similarity score respectively. As the figure shows both methods had quite similar behaviors in term of match rates but it is possible to note that the CBTF outperform MBTF in the rank1 and rank2 results (we matched the histogram also against the mean histogram of the individual contained in the training set). Only in the case of the one elliptic mask approach MBTF gives better results than CBTF (the difference is very narrow). However the overall performances confirmed that CBTF retained more color information than MBTF and produced a more accurate mapping function. In the same figure the different color histogram extraction approaches are also compared, even in this figure the match rate shows that

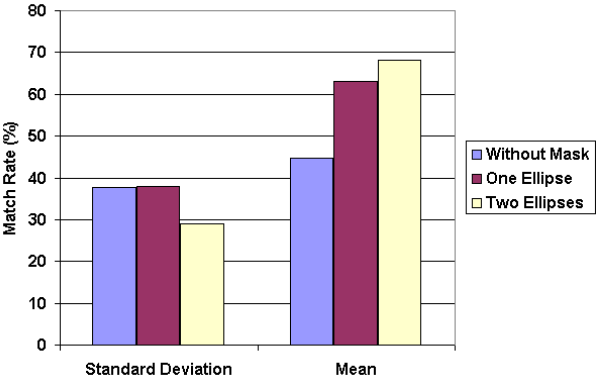


Fig. 7. Mean and standard deviation of the matching rate using different color extraction method to establish the correspondence between the two cameras

the two elliptic masks method gives the best results. This it what we expect for the reason explained in the first part of this section. Finally figure 7 shows the mean and standard deviation of the different color histogram extraction method estimated in both part of the experiment. Even these values demonstrate that double elliptic masks give the greatest mean score with the lowest standard deviation. This confirm that double elliptic masks method maps the data among the different six people classes better than the others.

4 Conclusions and Future Work

In this paper we have investigated on the feasibility of multi camera tracking algorithms based on the appearance similarity. We have considered two non overlapping cameras inside an office building. In this paper we have investigated the reliability of appearance similarity methods to track people in the same FOV and among two different FOVs. We evaluated different color histogram extraction approaches with different elliptic masks. Obtained results showed that using these masks improve overall results in terms of matching rate. We also compared two different Brightness Transfer Functions, ie. the MBTF and the CBTF. Experiments demonstrated quite similar behaviors of the two methods when the simple association problem has to be solved.

Future work will be addressed on the study of new methodologies for more reliable appearance matching. Since the people appearances can be similar in some parts of the body and differ in other parts we are thinking to apply different methodologies based on the extraction of graphs of patch histograms and use different weights in the correspondence matches in order to consider different reliability of the body parts and highlight only the significant differences among the people appearances.

References

1. Javed, O., Safique, K., Rasheed, Z., Shah, M.: Modeling inter camera space-time and apperacnce relationships for tracking across non-overlapping views. Computer Vision and Image Understanding 109, 146–162 (2008)
2. Javed, O., Shafique, K., Shah, M.: Appearance modeling for tracking in multiple non-overlapping cameras. In: CVPR, vol. 2, pp. 26–33 (2005)
3. Du, W., Piater, J.: Data Fusion by Belief Propagation for Multi-Camera Tracking. In: The 9th International Conference on Information Fusion (2006)
4. Du, W., Piater, J.: Multi-Camera People Tracking by Collaborative Particle Filters and Principal Axis-Based Integration. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part I. LNCS, vol. 4843, pp. 365–374. Springer, Heidelberg (2007)
5. Chilgunde, A., Kumar, P., Ranganath, S., WeiMin, H.: Multi-Camera Target Tracking in Blind Regions of Cameras with Non-overlapping Fields of View. In: BMVC 2004, Kingston, September 7-9 (2004)
6. Javed, O., Rasheed, Z., Alatas, O., Shah, M.: Knightm: a real time surveillance system for multiple overlapping and non-overlapping cameras.
7. Cai, Q., Aggarwal, J.K.: Tracking human motion in structured environments using a distributed camera system. IEEE Trans. Pattern Anal. Mach. Intell. 2(11), 1241–1247 (1999)
8. Porikli, F.: Inter-camera color calibration using cross-correlation model function. In: IEEE Int. Conf. on Image Processing (2003)
9. Shan, Y., Sahwney, H.S., Kumar, R.: Unsupervised learning of discriminative edge measures for vehicle matching between nonoverlapping cameras. In: IEEE Conf. on Computer Vision and Pattern Recognition (2005)
10. Prosser, B., Gong, S., Xiang, T.: Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions. In: British Machine Conference 2008 (2008)
11. Madden, C., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Machine Vision and Application 18, 233–247 (2007)
12. Jeong, K., Jaynes, C.: Object matching in disjoint cameras using a color transfer approach. Machine Vision and Application 19, 443–455 (2008)
13. Mazzeo, P.L., Spagnolo, P., Leo, M., D'Orazio, T.: Visual Players Detection and Tracking in Soccer Matches. In: IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance. AVSS 2008, pp. 326–333 (2008)