International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(1): 23-33
The Society of Digital Information and Wireless Communications, 2012 (ISSN: 2220-9085)

Multi-people tracking across multiple cameras

Faouzi Alaya Cheikh[1], Sajib Kumar Saha[2], Victoria Rudakova[3], Peng Wang[4]

[1,2]Gjøvik University College, Norway
[3]École Nationale Des Ponts et Chaussées, France
[4]Univeristy Jean Monnet, France
faouzi.cheikh@hig.no, to_sajib_cse@yahoo.com,
rudakovv@imagine.enpc.fr, superwpeng@gmail.com

Abstract. Multi-target tracking (MTT) is an active and challenging research topic. Many different approaches to MTT problem exist, yet there are still few satisfactory methods of solving multi-target occlusion problem, which often appears in multi-target tracking task. The application of multi cameras in most existing researches for multi-target occlusion requires camera calibration parameters in advance, which is not practical in the case of outdoor video surveillance. Most of the proposed solutions for this problem require camera calibration parameters that make them impractical for outdoor video surveillance applications. To address this problem we propose in this paper a probabilistic approach, the foremost consideration of which is to reduce the dependency on camera calibration for multiple camera collaboration. More robustness on target representation and object tracking has been ensured by combining multiple cues such as border information of the object with color histogram, while Gale-Shapley algorithm (GSA) has been used for finding the stable matching between objects of two or more camera views. Efficient tracking of object ensures proficient recognition of target depicting parameters (i.e. apparent color, height and width information of the object) as a consequence provides better camera collaboration. Initial simulation results prove the validity of the proposed approach.

Keywords: Surveillance, multi-camera tracking, multi-target tracking, particle filter.

1 INTRODUCTION

Object tracking has received tremendous attention in the video processing community due to its numerous potential applications in video surveillance, human activity analysis, traffic monitoring, insect / animal tracking and so on. Over the last couple of years, many algorithms and results have been presented with regard to the problem of object tracking and recently the focus of the community has been concentrated on multi-target tracking (MTT) that requires determining the number as well as the dynamics of the targets. Even though single object tracking is considered as a mature field of research, due to a combination of several factors, reliable multi-target tracking still remains a challenging domain of research. The underlying difficulties behind multi-target tracking are founded mostly upon the apparent similarity of targets and then multi-target occlusion. MTT for targets with distinctive appearance is comparatively easier, since it can be solved reasonably well by using multiple independent single-target trackers. But MTT for targets with similar appearance, such as pedestrians in crowded scenes, is a much more difficult task. In addition, MTT must deal with multi-target occlusion, namely, the tracker must separate the targets and assign them correct labels. This is a particular problem in a multi-object scenario, where object-object interactions and occlusions are likely to occur. It is not possible to reliably maintain a target's identity using single clue such as continuity of appearance or motion; besides, computational complexity also plays an important role, as many applications require the tracking to be done in real time. All these issues make target tracking or multi-object tracking a sturdy task even now.

The proposed methodology in this paper is an enhancement of the scheme proposed in [33], where the main concentration was to use color and size information of the object for multi camera collaboration in order to reduce the dependency on camera calibration. Here more robustness on object tracking has been ensured by combining the centroid information calculated from the detected border of the object.

2 LITERATURE SURVEY

Most of the early works for MTT were based on monocular video [1]. A widely accepted approach that addresses many problems of MTT is based on a joint state-space representation and infers the joint data association [2, 3]. A binary variable has been used by MacCormick and Blake [4] to identify foreground objects and then a probabilistic exclusion principle has been used to penalize the hypothesis where two objects occlude. In [5], the likelihood is calculated by enumerating all possible association hypotheses. Zhao and Nevatia [6, 7] used a different 3D shape model and joint likelihood for multiple human segmentation and tracking. Tao et al. [8] proposed a sampling-based multiple-target tracking method using background subtraction. Khan et al. in [9] proposed an MCMC-based particle filter which uses a Markov random field to model motion interaction. Smith et al. presented a different MCMC-based particle filter to estimate the multi-object configuration [10]. McKenna et al. [11] presented a color-based system for tracking groups of people. Adaptive color models are used to provide qualitative estimates of depth ordering during occlusion. Although the above solutions, which are based on a centralized process, can handle the problem of multi-target occlusion in principle, they require a tremendous computational cost due to the complexity introduced by the high dimensionality of the joint-state representation which grows exponentially in terms of the number of objects tracked.

Several researchers proposed decentralized solutions for multi-target tracking. In [12] the multi-object occlusion problem has been solved by using multiple cameras where the cameras are separated widely in order to obtain visual information from wide viewing angles and offers a possible 3D solution. The system needs to pass the subjects identities across cameras when the identities are lost in a certain view by matching subjects across camera views. Therefore, the system needs to match subjects in consecutive frames of a single camera and also match subjects across cameras in order to maintain subject identities in as many cameras as possible. Although this cross-view correspondence is related to wide baseline stereo matching, traditional correlation based methods fail due to the large difference in viewpoint [13].

Yu and Wu [14] and Wu et al. [15] used multiple collaborative trackers for MTT modeled by a Markov random network. This approach demonstrates the efficiency of the decentralized method. The decentralized approach was carried further by Qu et al. [16] who proposed an interactively distributed multi-object tracking framework using a magnetic-inertia potential model.

However, using multiple cameras raises many additional challenges. The most critical difficulties presented by multi-camera tracking are to establish a consistent label correspondence of the same target among the different views and to integrate the information from different camera views for tracking, which is robust to significant and persistent occlusion.

Many existing approaches address the label correspondence problem by using different techniques such as feature matching [17, 18], camera calibration and/or 3D environment model [18, 19], and motion-trajectory alignment [20]. A kalman filter based approached has been proposed in [13] for tracking multiple object in indoor environment. In addition with apparent color, apparent height, landmark modality and homography, epipolar geometry has been used for multi-camera cooperation. Recently, Qu et al. have presented a distributed Bayesian framework for multiple-target tracking using multiple collaborative cameras. The distributed Bayesian framework avoids the computational complexity inherent in centralized methods that rely on joint-state representation and joint data association. Epipolar geometry has been used for multi-camera collaboration. However, a lot of dependency on epipolar geometry makes that approach impractical for outdoor video surveillance application, since we need to know the angle of view for each camera with respect to others accurately; which can change very frequently for outdoor surveillance cameras due to environmental maladies.

3 PROPOSED FRAMEWORK

3.1 BAYESIAN SEQUENTIAL ESTIMATION

The Bayesian sequential estimation helps to estimate a posterior distribution noted as $p(x_{0:t}|z_{1:t})$ or its marginal $p(x_t|z_{1:t})$, where $x_{0:t}$ includes a set of all states up to time t and $z_{1:t}$ - set of all the observations up to time t accordingly. The evolution of the state sequences $\{x_t, t \in N\}$ of a target is given by equation 1; and the observation model is given by equation 2:

$$x_t = f_t(x_{t-1}, v_{t-1}) \quad \textbf{(1)}$$

$$z_t = h_t(x_t, u_t) \quad \textbf{(2)}$$

Where $f_t\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_x}$ and $h_t\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_z}$ can be both linear and nonlinear, and v_t and u_t are sequences of i.d.d. process noise and measurement noise respectively; at the same time n_v and n_u are their dimensions.

In Bayesian context, the tracking problem can be considered as recursive calculation of some belief degree in the state x_t at time step t, given the data observation $z_{1:t}$. That is, we need to construct a pdf $p(x_t|z_{1:t})$. It is assumed that the initial state of the system (also called prior) is given.

$$p(x_0|z_0) = p(x_0) \quad \textbf{(3)}$$

Then, the posterior distribution $p(x_t|z_{1:t})$ can be calculated by the following next two steps given in equations (4) and (5).

Prediction:

$$p(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1}) p(x_{t-1}|z_{1:t-1})\, dx_{t-1} \quad \textbf{(4)}$$

Update:

$$p(x_t|z_{1:t}) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{\int p(z_t|x_t)p(x_t|z_{1:t-1})\ dx_{t-1}} \quad \textbf{(5)}$$

In equation (5) the denominator is a normalization constant and it depends on the likelihood function $p(z_t|x_t)$ - as described by equation (2). But presented recursive propagation is just a conceptual solution and cannot be applied in practice.

Monte-Carlo simulation [26] with sequential importance sampling (SIS) technique, allows us to approximate equations 4 and 5 in the discreet form:

Prediction:

$$p(x_t|z_{0:t-1}) \approx \sum_{n=1}^{N_s} p(x_t|x_{t-1}^{(n)})\omega_{t-1}^{(n)} \quad \textbf{(6)}$$

Update:

$$p(x_t|z_{0:t}) \approx \sum_{n=1}^{N_s} \omega_t^{(n)}\, \delta(x_t - x_{t-1}) \quad \textbf{(7)}$$

Where $\omega_t^{(n)} \propto \frac{p(z_t|x_t^{(n)})p(x_t^{(n)}|x_{t-1}^{(n)})}{q(x_t^{(n)}|x_{t-1}^{(n)},z_{0:t})}\omega_{t-1}^{(n)}$ as well as $\sum_{n=1}^{N_s} \omega_t^{(n)} = 1$

Nonetheless in order to avoid degeneracy (one of the common problems with SIS) re-sampling of the particles need to be done. Therefore, the main idea is to update the particles which almost do not make any contribution to the approximation to $p(x_t|z_{1:t})$ and pay an attention on the more promising particles. It generates a new set $\{\tilde{x}_t^{(n)}, N_s^{-1}\}_{n=1}^{N_s}$ by re-sampling with replacement N_s times from an approximate discreet representation of $p(x_t|z_{0:t})$, so that $Pr\left(\tilde{x}_t^{(n)} = x_t^{(m)}\right) = \omega_t^{(m)}$ and weights must be reset to $\omega_t^{(n)} = \frac{1}{N_s}$ (we denote it N_s^{-1}).
For this project we used the re-sampling scheme proposed in [27].

State estimate:

The mean state ('mean' particle, 'mean'shape or weighted sum of particles) has been used.

$$\varepsilon[X_t] = \sum_{n=1}^{N_s} \omega_t^{(n)} x_t^{(n)} \quad \textbf{(8)}$$

3.2 MODELING OF DENSITIES

Modeling the densities for SIS is an important task, at the same time plays a critical role in the system performance. The models that have been used in our proposed method are different from [1], and the following subsections will cover the construction of each of those densities.

- **Target representation**

Since ellipses can provide more precise shape information than a bounding box [28], simple 5D parametric ellipse was used in order to decrease computational costs, and at the same time, it is sufficient to represent the tracking results. Therefore, one state x_t^i is defined by $x_t^i = (cx_t^i, cy_t^i, a_t^i, b_t^i, \rho_t^i)$, where i is target index, t - current time, (cx, cy) - coordinates of the center of ellipse, (a, b) major and minor axis, and ρ - orientation angle, as shown in Figure 1.

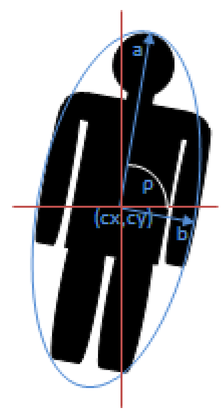

Fig. 1. Ellipse representation for describing a human body.

- **Local observation model**

As a local likelihood $p\left(z_t^i \middle| x_t^{i,(n)}\right)$, single cue - color histogram like [29] has been used. Color models are obtained in the Hue-Saturation-Value (HSV) color space in order to decouple chromatic information from shading effects. As for the bin numbers, we have used $N_h = N_s = 8$ and N_v=4 (Hue, Saturation and Value bin numbers respectively).

- **State dynamics model**

The better prediction results were achieved by using motion information. It was decided to use Lucas-Kanade (LK) optical flow algorithm [30] when we dealt with single independent tracker. However, this method works well for the only case when targets do not interact with each other.

- **Interaction model**

When targets start to interact with each other (i.e., occlusion), we cannot rely on the motion based proposal anymore. Hence magnetic-repulsion inertia re-weighting scheme [29] established on random based proposal was used instead of the re-weighting scheme [31], since [29] gives better result than [31] in our experiment.

3.3 IMPROVED DETECTION OF MOVING OBJECT

Another important point to consider is how to include prediction of a, b and ρ the ellipse parameters for stating dynamics model. Until now it was possible to predict only the center coordinates (cx, cy) of the ellipse based on the LK optical flow [30] calculation. At the same time parameters a, b and ρ propagate according to the random-based prediction which is not very effective sometimes since it can lead to over-expanding or over-squeezing of the ellipse bounds. Hence to overcome this problem, objects' contour information has been considered for calculating a, b and ρ.

Detecting the contour information consists of three phases like in [32]; in the first phase gradient information has been extracted since it is much less affected by the quantization noise and abrupt change of illumination. Sobel operators have been used instead of Roberts operators or Prewitt operators, as they are generally less computationally expensive and more suitable for hardware realizations [32].

$$f_E(x,y,t) = |f_E(x,y,t) * H_X| + |f_E(x,y,t) * H_Y| \quad \textbf{(9)}$$

where $H_X = \frac{1}{4}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ and

$$H_Y = \frac{1}{4}\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Where, f(x, y, t) denotes a pixel of a gray-scale image, f_E (x, y, t) denotes the gradient of the pixel, H_X and H_Y denote the horizontal and the vertical transform matrix, respectively.

Color image has been converted to gray image using the following equation

$$f(x,y,t) = 0.587 \cdot f(x,y,t)_G + 0.114 \cdot f(x,y,t)_B + 0.29 \cdot f(x,y,t)_R \quad \textbf{(10)}$$

In the second phase three frame differencing scheme [32] has been used instead of commonly used two frame differencing method for better detection of the contours of moving objects. The operation is detailed in the following equation

$$D(x,y,\Delta t) = |f_E(x,y,t) - f_E(x,y,t-1)| \cdot |f_E(x,y,t) - f_E(x,y,t+1)| \quad \textbf{(11)}$$

The background contour in an image resulted from such a gradient-based three-frame differencing can be largely eliminated, whereas the contours of moving objects can be remarkably enhanced.

In the third phase a step of eight-connectivity testing is appended with three-frame differencing for the sake of noise reduction.

Once the contours of moving objects have been detected using the above mentioned procedure, then (*cx*,*cy*) detected by particle filter has been used to calculate a', b', ρ'. A point $(x_{a'}, y_{a'})$, that is the farthest from (*cx*,*cy*) within a certain perimeter defined by (*cx*, *cy*, *a*, *b*+ $\frac{b}{3}$, *ρ*) has been calculated. Then $\rho' = \tan^{-1}(\frac{y_{a'} - cy}{x_{a'} - cx})$ has been calculated. Now another farthest point $(x_{b'}, y_{b'})$, which makes $90^o \pm 5^o$ angle with respect to line joining (*cx*,*cy*) and $(x_{a'}, y_{a'})$ has been calculated.

Once $(x_{a'}, y_{a'})$ has been detected then $(x_{a'}^{180}, y_{a'}^{180})$ has been calculated in the direction that makes $180^o \pm 5^o$ angle with respect to line joining (*cx*,*cy*) and $(x_{a'}, y_{a'})$. Likewise $(x_{b'}^{180}, y_{b'}^{180})$ has also been calculated in the direction that makes $180^o \pm 5^o$ angle with respect to line joining (*cx*,*cy*) and $(x_{b'}, y_{b'})$. Now (a', b', cx', cy') has been calculated as following

$$a' = \frac{L(x_{a'}, y_{a'}) + L(x_{a'}^{180}, y_{a'}^{180})}{2} \quad \textbf{(12)}$$

$$b' = \frac{L(x_{b'}, y_{b'}) + L(x_{b'}^{180}, y_{b'}^{180})}{2} \quad \textbf{(13)}$$

where $L(a, b)$ is the length of the line joining (*cx*,*cy*) and (a, b)

Likewise

$$cx' = \frac{x_{b'} + x_{b'}^{180}}{2}, cy' = \frac{x_{a'} + x_{a'}^{180}}{2} \quad \textbf{(14)}$$

Now the ellipse that tracks the object(s) has been defined by the parameters $(cx_{new}, cy_{new}, a_{new}, b_{new}, \rho_{new})$ where $cx_{new} = 0.4 * cx + 0.6 * cx'$, $cy_{new} = 0.4 * cy + 0.6 * cy'$, $a_{new} = 0.4 * a + 0.6 * a'$, $b_{new} = 0.4 * b + 0.6 * b'$ and $\rho_{new} = 0.4 * \rho + 0.6 * \rho'$

If for any reason (shadow, occlusion etc; Fig. 7), the calculated a' becomes unreliable while a_{new} is being calculated, the weighting factors (0.5, 0.5) have been used instead of (0.4, 0.6); and the last used (from known previous frame) value of a' will be used; if the value of a' is taken from a frame which is not more than 3 frames before the current frame. These factors are (0.6, 0.4) if the value of a' has been taken from a frame which is more than 3 frames but less than 9 frames before from the current frame; otherwise (1, 0). Same is true for b', cx', cy', ρ'.

3.4 MULTI-CAMERA DATA FUSION

In order to correctly associate corresponding targets (assign the same identity of objects irrespective of camera views) Gale-Shapley algorithm (GSA) [24] has been applied, which uses the color, height and width information of the detected object in 2 camera views. Each time a new object appears in the camera view(s), its normalized color histogram (normalized by apparent height and width; we will refer it as initialized histogram) has been stored in the camera view(s), at the same time the corresponding labeling of objects with respect to one reference camera has been done by using GSA that uses the histogram distance between objects of two different camera views to generate the preference lists for objects. For the system using more than 2 cameras that labeling should be done for all the cameras with respect to one reference camera. After labeling has been done each camera can work independently and track individually. Once an occlusion occurs in the reference camera, the histogram distance of objects has been calculated by considering the normalized color histograms of objects (of that frame) with respect to its initialized histogram on that camera view; and then occluded object has been identified by comparing the histogram distances of objects that are interacting. If an (or more) occlusion has been detected, the idea is to figure out which camera can be suggested for that object. For that perspective the reference camera will look for that object in the camera right to it. If that object is also occluded there then the reference camera will look for that object in the camera just left to it to check whether that object is occluded on that camera view or no and so on.

4 EXPERIMENTAL RESULT

4.1 EXPERIMENTAL SETUP

Two USB Logitec web cameras have been used. All the video sequences with people were recorded from these cameras. For the purpose to make synchronized recordings, software developed by Nils Fjeldsø at HIG was used for two cameras. For the initialization of the targets the code implemented in [31] has been used.

Several experimental camera set-ups were organized, where people number, their activities (hand-shaking, walking and occluding each other) and also illumination (daylight, artificial room light) varied. Original videos were recorded with the frame size of 640 × 480. For our tests it was decided to decrease the frame size to 320 × 240 to lower computational costs needed for processing one frame. For all the sequences, we have used 50 particles for each target.

4.2 EXPERIMENTAL ANALYSIS

For tracking individual object the model proposed here that combines border information of the object with particle filter information gives better result than using only particle filter information (see figure 2, 3; in figure 2 the tracker wraps considerable amount of background data). Even though the extra processing for combining border takes some extra time (15 ms per frame [32]), that time is very negligible compared to the time taken by particle filter but this operation ensures good tracking and overall good performance of the proposed framework.

The proposed methodology guarantees quality tracking and gives constant labeling of objects irrespective of camera views. For all the tested video sequences (15 video sequences, 2 camera views with overlapping field of view) the tracking and labeling of objects were appropriate, unless the objects were too far (more than 8 meters) from the camera.

Fig. 2. Tracking of objects using partilcle filter only.

Fig. 3. Tracking of objects using augmented particle filter augmented with border information.

Fig. 4. Video sequences obtained from two camera views and tracking of objects (where the color of ellipses ensures that the objects are labeled correctly irrespective of camera views).

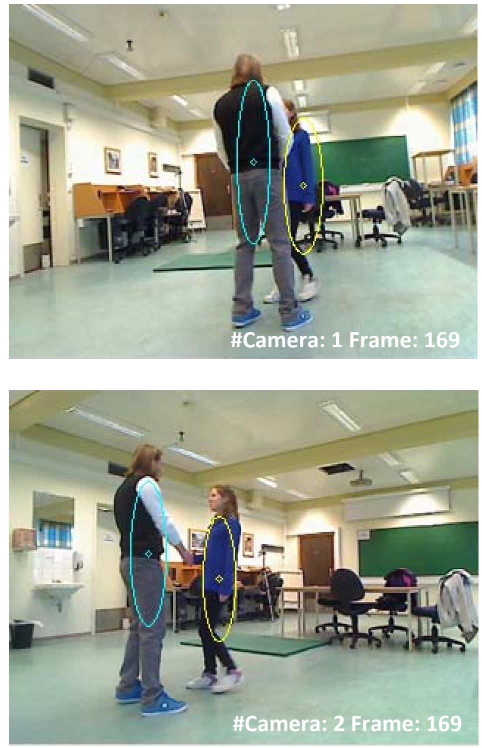

Fig. 5. Video sequences obtained from two camera views and tracking of objects.

Fig. 6. Video sequences where the yellow tracker loses its track.

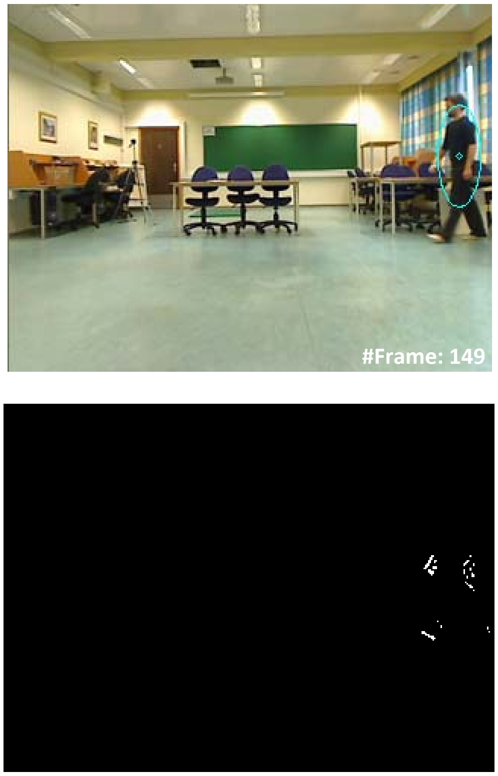

Fig. 7. Video sequence where detected border of the object is not efficient.

5 CONCLUSION AND DISCUSSION

Provided that epipolar geometry is not as apposite solution for multi-camera data fusion, a methodology for multi-camera data fusion has been proposed based on Gale-Shapley Algorithm. The idea to use GSA here is not to check the correctness of the coordinates of tracked objects in the image coordinate system (like most of the systems do), but to maintain correct identities of targets among the different camera views and gives more reliable results. While GSA based on color histogram (normalized with respect to height and width) is applied, it is very important that the tracked ellipse should contain mostly the object. Experiment results indicate that the proposed methodology ensures better tracking of objects. As a consequence, the overall accuracy of GSA increases as well. At the same time the efficient tracking of objects by the proposed methodology guarantees better identification of occlusion in camera view(s). The framework is general and can be easily adapted for tracking any class of objects.

Due to the design of the approach itself, there is a limitation of the performance. Like any color based detection methods, for objects of apparent similarity (objects wearing similar cloths and having same height and width) the system may not work properly. Besides, it has been observed from the experiment that when the object becomes too far from the camera (more than 8 meters) the tracker loses its track (see figure 6). Even though HSV color space has been used to ignore the lightness effect on objects with respect to the camera distance, it does not give very robust result in reality. Hence more investigation on advanced color spaces can be done in future.

There is room for improvement in the modeling of the appearance of the targets. All ellipse paramenters are generated based on the centroid of the ellipse. To get more precise description of targets, geometric formula can be applied for more robust detection of centroids in future.

REFERENCES

[1] Qu, W., Schonfeld, D., & Mohamed, M.: Distributed Bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. EURASIP Journal on Applied Signal Processing, Issue 1, 21–21 (2007)
[2] Y. Bar-Shalom and A. G. Jammer: Tracking and Data Association. Academic Press, San Diego, Calif, USA (1998)
[3] C. Hue, J.-P. L. Cadre, and P. P´erez: Sequential Monte Carlo methods for multiple target tracking and data fusion. IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 309–325 (2002)
[4] J. MacCormick and A. Blake: A probabilistic exclusion principle for tracking multiple objects. International Journal of Computer Vision, vol. 39, no. 1, pp. 57–71 (2000)
[5] N. Gordon: A hybrid bootstrap filter for target tracking in clutter. IEEE Transactions on Aerospace and Electronic Systems, vol. 33, no. 1, pp. 353–358 (1997)

[6] T. Zhao and R. Nevatia: Tracking multiple humans in crowded environment. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04), vol. 2, pp. 406–413, Washington, DC, USA, June-July (2004)

[7] T. Zhao and R. Nevatia: Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1208–1221 (2004)

[8] H. Tao, H. Sawhney, and R. Kumar: A sampling algorithm for detection and tracking multiple objects. In Proceedings of IEEE International Conference on Computer Vision (ICCV'99) Workshop on Vision Algorithm, Corfu, Greece, September (1999)

[9] Z. Khan, T. Balch, and F. Dellaert: An MCMC-based particle filter for tracking multiple interacting targets", In Proceedings of 8th European Conference on Computer Vision (ECCV '04),vol. 4, pp. 279–290, Prague, Czech Republic, May (2004)

[10] K. Smith, D. Gatica-Perez, and J.-M. Odobez: Using particles to track varying numbers of interacting people. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), vol. 1, pp. 962–969, San Diego, Calif, USA, June (2005)

[11] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler: Tracking groups of people. Computer Vision and Image Understanding, vol. 80, no. 1, pp. 42–56 (2000)

[12] L. Lee, R. Romano and G. Stein.: Monitoring activities from multiple video streams: Establishing a common coordinate frame. IEEE Transactions on Pattern Analysis and Machine Intelligence, Special Issue on Video Surveillance and Monitoring: 758–767 (2000)

[13] Ting-hsun Chang , Shaogang Gong. Tracking Multiple People with a Multi-Camera System. IEEE Workshop on Multi-Object Tracking (2001)

[14] T. Yu and Y. Wu: Collaborative tracking of multiple targets. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04), vol. 1, pp. 834–841, Washington, DC, USA, June-July (2004)

[15] Y. Wu, G. Hua, and T. Yu. Tracking articulated body by dynamic Markov network. In Proceedings of 9th IEEE International Conference on Computer Vision (ICCV '03), vol. 2, pp.1094–1101, Nice, France, October (2003)

[16] W. Qu, D. Schonfeld, and M. Mohamed: Real-time interactively distributed multi-object tracking using a magnetic-inertia potential model. In Proceedings of 10th IEEE International Conference on Computer Vision (ICCV '05),vol. 1, pp. 535–540, Beijing, China, October (2005)

[17] Q. Cai and J. K. Aggarwal.: Tracking human motion in structured environments using a distributed-camera system. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 11, pp. 1241–1247 (1999)

[18] P. H. Kelly, A. Katkere, D. Y. Kuramura, S. Moezzi, S. Chatterjee, and R. Jain: An architecture for multiple perspective interactive video. In Proceedings of 3rd ACM International Conference on Multimedia (ACM Multimedia '95), pp. 201–212,San Francisco, Calif, USA, November (1995)

[19] J. Black and T. Ellis.: Multiple camera image tracking. In Proceedings of 2nd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS '01), Kauai,Hawaii, USA, December (2001)

[20] L. Lee, R. Romano, and G. Stein: Monitoring activities from multiple video streams: establishing a common coordinate frame. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 758–767 (2000)

[21] Carin Hue, Jean-Pierre Le Cadre, and Patrick P´erez.: Sequential monte carlo methods for multiple target tracking and data fusion. IEEE Transactions on Signal Processing, 50(2):309–325, February (2002)

[22] P. P´erez, C. Hue, J. Vermaak, and M. Gangnet.: Color-based probabilistic tracking. In European Conference on Computer Vision, number 2350 in Lecture Notes in Computer Science, pages 661–675 (2002)

[23] Gale, D. & Shapley, L. S.: College admissions and the stability of marriage. American Mathematical Monthly, 69, 9–14 (1962)

[24] http://en.wikipedia.org/wiki/Stable_marriage_problem

[25] Guraya, F. F. E., Bayle, P.-Y., & Cheikh, F. A.: People tracking via a modified CAMSHIFT algorithm. (2009)

[26] Maskell, S. & Gordon, N.: A tutorial on particle Filters for on-line nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 2001, 50, 174–188 (2001)

[27] Kitagawa, G. Monte Carlo: Filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics, 5, No. 1, 1–25, 1996

[28] Tarng Chen, Yu-Ching Lin, Wen-Hsien Fang: A VIDEO-BASED HUMAN FALL DETECTION SYSTEM FOR SMART HOMES. YieJournal of

the Chinese Institute of Engineers, Vol. 33, No. 5, pp. 681-690 (2010)

[29] Nummiaro, K., Koller-Meier, E., Gool, L. V., & Gaal, L. V.: Object tracking with an adaptive color-based particle Filter. http://www.koller-meier.ch/esther/dagm2002.pdf. (2002)

[30] Bouguet, J. Y.: Pyramidal implementation of the Lucas Kanade feature tracker: Description of the algorithm. Intel Corporation Microprocessor Research Labs, http://robots.stanford.edu/cs223b04/algo_tracking.pdf. (2002)

[31] Blake, A. & Isard, M. : The Condensation algorithm - conditional density propagation and applications to visual tracking. In Advances in Neural Information Processing Systems (NIPS)1996, 36–1, Denver, CO, USA. The MIT Press. December 2-5 (1996)

[32] Zhao, S., Zhao, J., Wang, Y., & Fu, X.: Moving object detecting using gradient information, three-frame-differencing and connectivity testing. AI 2006: Advances in Artificial Intelligence, 4304, 510–518 (2006)

[33] Victoria Rudakova, Sajib Kumar Saha, Faouzi Alaya Cheikh,: Multiple collaborative cameras for multi-people tracking using color-based particle filter and border information of the objects. DICTAP 2011: International Conference on Digital Information and Communication Technology and its Application, 315-326(2011)