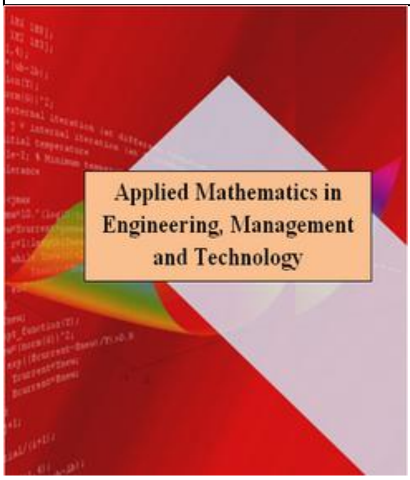
Applied mathematics in Engineering, Management and Technology 3(4) 2015:354-361
www.amiemt-journal.com

Tracking Multiple object with multi const camera in environment by Graph Cut and Filter kalman

Parisa jahanshahi
Islamic Azad University Qazvin Branch
Electronic, Computer & IT, Department of Computer Ghazvin, Iran
p.jahanshahi@qiau.ac.ir



Abstract

In this study, a powerful method of tracking people was presented based on Graph Cut and Kalman Filter in some fixed cameras. Moving targets appeared by using a background model and background separation. Homography restriction was used for camera correspondence. This method uses the matching of target points on individuals' busts as their characteristics. The resulted pictures were changed into binary pictures using graph cut segmentation, and the junction points that indicated the individuals' position were separated from other parts of the pictures. The mean position of the points was measured. Kalman filter estimated the modes of moving objects detected in the current frame with regard to the measurements done in the previous frame. The combination of the resulted junction point's position with homography transformation and bust point of the individuals in each frame through the use of graph cut was used as some measurement to improve this estimation. The results were applied to a set of videos from different scenes and were compared using Graph Cut and Kalman Filter. The experiments indicate the efficiency of the presented method.

Keywords: homography; multi const camera, multi object tracking; kalman filter; graph-cut.

INTRODUCTION

Intelligent video surveillance using multi cameras has been one of the most active research areas in computer vision. The view of a single camera is finite and limited by scene structures. Multi camera tracking aims to establish the spatio-temporal correspondence of the same object across multiple views. Multi-camera video surveillance faces many challenges with the fast growth of camera networks. As the scales of camera networks increase, it is expected that the multi-camera surveillance systems can self-adapt to a variety of scenes with less human intervention, and Tracking multiple people accurately in cluttered and crowded scenes. The goal is to efficiently extract useful information from a huge amount of videos collected by surveillance cameras by automatically detecting, tracking and recognizing objects of interest, and understanding and analyzing their activities. In this paper, we propose a tracking approach to detect and track multiple people in crowded and cluttered scenes by fixed multi camera.

Algorithms for target tracking in multi-view camera networks can be grouped based on the modalities for tracking and information fusion and can be categorized into two main classes, namely track-first, fuse-first.

A. Track-first Approaches

Track-first multi-view tracking can be performed either independently in each view or collaboratively across views. In collaborative tracking, estimated tracks in the image view and in the common view can be used to assist each other and to improve track estimates in one view. Both independent and collaborative algorithms first track objects in each camera view and then project the tracks onto the common view for fusion. The problem to be solved here is the fusion of the multiple tracks belonging to the same target.
Track-first approaches involve multiple tracking steps and hence can be computationally expensive. To reduce the complexity, fuse-first approaches can be used that defer the tracking step until when the information from each view is fused on a common view [1-2-3].

B. Fuse-first approaches

Applied mathematics in Engineering, Management and Technology 3(4) 2015:354-361
www.amiemt-journal.com

Fuse-first approaches project detection information from each view to a common view and then apply tracking. This approach, involve multiple tracking steps that can introduce sources of estimation error. These multiple steps can be eliminated by tracking on the common view only, by accumulating on the common view the information from each view.
The features extracted can be the feet location of people, the foreground pixels or the whole motion segmentation likelihood. Note that although fuse first methods involve one tracking step only, they may involve multiple detection steps in each camera view, before fusion and on the common view, after fusion [3-4-6].
Our approach is based on the fuse-first approach. Object appears in the scene be detected as foreground in each view using a background model and background difference. Next, for related between cameras used homographic constraint. Any pixel inside the foreground object in every view will be related by homographies inducted by the reference view. In this approach, the torso of a person is points of intersection in Homography plane.
Reference view Images converted to binary images by a graph-cut segmentation. This step separated the position of the intersection points from other parts inside reference images. To track, we measurement the average position of the points. The kalman filter provides an optimal estimate of its position at each time step. The filter kalman, the first one is the prediction of the next state estimate using the previous one, the second is the correction of that estimate using the measurement.

I. OVERVIEW OF THE PROPOSED METHOD

In this section the ways of detecting the background model, distinction of foreground from background, which is obtained using background modeling techniques, Correspondence between multiple cameras, segmentation of the reference image using graph cut and object tracking using Kalman filter are stated.

A. extraction forgground

Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction.
In this paper, we use of average approach for background model. Background pixels are constant over time for each incoming frame. To work, we obtain frames for a Window of time and stack them together. The sliding time window size was kept at 200 frames for each experiment. To each window of time, for each pixel of the image, the most frequent color in the time panel is considered to be the background image's color in that pixel.

B. Correspondence Between Multiple Cameras

Correspondence between multiple cameras involves at the same time instant finding correspondences between objects in the different image sequences. The existing methods for establishing correspondences can be classified (1,2), according to the types of employed features, whether the cameras are calibrated or not, and whether the correspondences are region-based or point-based.
In this paper, we used of the points located on torso to match people in multiple views, based on the homography constraint defined by the reference plane, that is point-based. This approach of detection and occlusion resolution is based on geometrical constructs and requires only the distinction of foreground from background, which is obtained inside previous part. This has the dual action of localizing people in the scene as well as resolving localization object. Warping a pixel from one image to another using the homography induced by a reference scene plane amounts to projecting a ray through the pixel onto the piercing point and then projecting it to the second camera center.
Figure 1 shows an object on a planar surface. The scene is being viewed by two cameras. H_π is the homography of the planar surface from view 1 to view 2. Warping a pixel from view 1 with H_π amounts to projecting a ray on to the plane at the piercing point and extending it to the second camera. Pixels that are image locations of scene points off the plane have plane parallax when warped. This can be observed for the red ray in the figure. Let p = (x, y, 1) denote the image location of a 3D scene point in one view and let p' = (x', y', 1) be its

coordinates in another view. Let H_π denote the homography induced by plane π between the two views. [1,5,7,8,9,10]

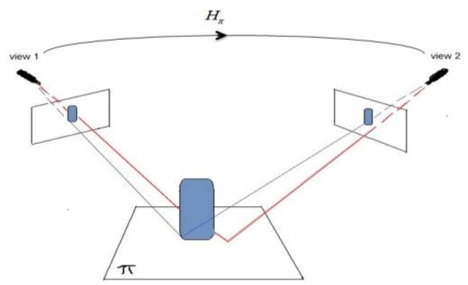


Fig. 1. the relationship between two corresponding pixels by homography matrix

Geometrically speaking warping pixel p from the first image to the second using the homography H_π amounts to projecting a ray from the camera center through pixel p and extending it till it intersects the plane π at the point often referred to as the 'piercing point' of pixel p with respect to plane π. The ray is then projected from the piercing point onto the second camera. The point in the image plane of the second camera that the ray intersects is pw. In effect pw is where the image of the piercing point is formed in the second camera.
In the present study, at any time the homography is done equal to the number of available cameras and each time the reference image is regarded to be the view of one camera. Every created reference image which is a combination of all views is a graylevel image. Some points of the image at hmography plane, where the intersection have been occurred are highly bright. . figure 2 shows, homographies result represented to the view of camera1.

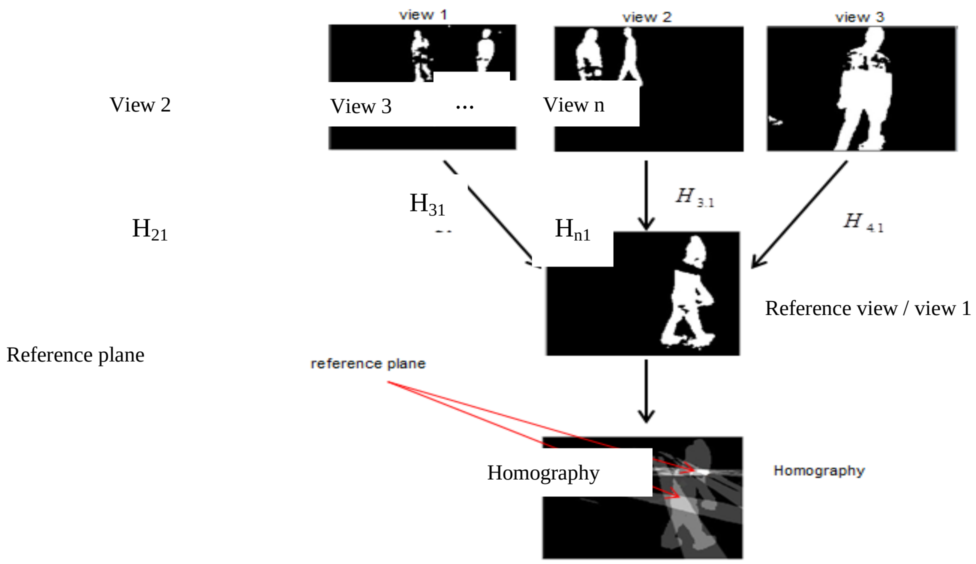


Fig. 2. the reference image with regard to the perspective of camera 1

C. separating junction points using grapg cut

In this study, graph cut segmentation algorithm is used to separate junction points at homography level in the image. This segmentation algorithm shows the image as graph segmentation problem. Node $V=\{u,x,\ldots\}$ of the image pixels and the graph N which is made of all pixels of the image are classified into N discrete sub-graphs. This is done by pruning the graph's weighted wings which have the least weights. The pruned wings between the two sub-graphs are called the cut. These sub-graphs are not related to each other at all and there is no common node among these sub-graphs. In figure 3, the red wings in graph G might be considered as a cut. eliminating these wings will result in two subgraphs A and B [3-10-11-12-13].

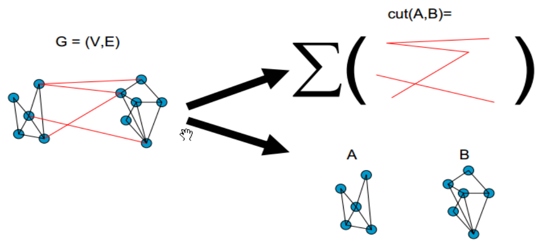
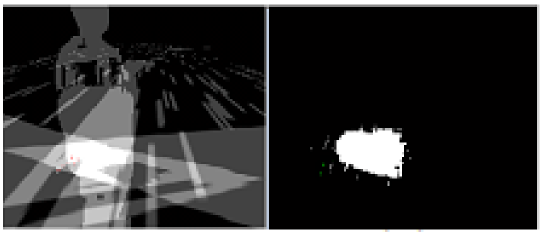


Fig. 3. an instant of graph cut performance [7]

Graph cut is done at any time on the reference images obtained from the perspective of each camera. As can be seen in figure 4, in each image the intersection points of the individuals' busts which are related to homography level and has higher brightness intensity than other parts of the image are presented.



(a) (b)

Fig. 4. a. reference image of camera 1's view. b. the image obtained from graph cut on the reference image.

D.tracking

After applying graph cut on this image, it should get prepared for tracking with Kalman filter. To do so, the output binary image is changed into some positions. For this purpose, the position of the available targets in the system are considered as P1 = {x, y} on the image. In addition to these points, the camera's access points are also considered as P2 = {x, y} .The access points are some points of the scene whose target enter the scene. P' are considered as the aggregation of points P1 and P2 which indicate the target position in the image. Each bright point on the image is related to the closest point in P'. The mean point related to each target is considered to be a measurement and is applied on the tracking system of Kalman filter.
Using these measurements, Kalman filter will better estimate the object's position compared to the available measurement. In this type of filters, tracking equations are divided into prediction and correction groups. Prediction equations first help to estimate the object's position in the next frame. After that, this position is corrected by using available measurements. This way, the intended object's moving behavior is obtained [14-15-16-17].
In prediction and correction equations shown in figure 5, each stage's output is the other one's input.

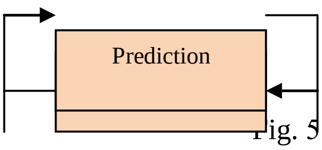


Fig. 5. Kalman filter's equations diagram [17]

Any probable position of the target which is reported through measurement is considered as measurement m, and the whole available measurements in a frame are considered as M_K. The object's real two dimentional coordinate is shown as (px, py) while the two dimentional coordinate resulted from measurment in the network is shown as (mx, my). X_K and Z_K are state and measurement vectors, respectively. Equation (1) is the equation of the target's motion which is defined as follows:

$$px_K = px_{K-1}$$
$$py_K = py_{K-1}$$

To obtain the targets' state vector at any time steps, Kalman filter is updated with new measurements and then, at the prediction stage, an estimation for x_K and y_K for the targets' state are done using the former state of x_{K-1} and y_{K-1}.

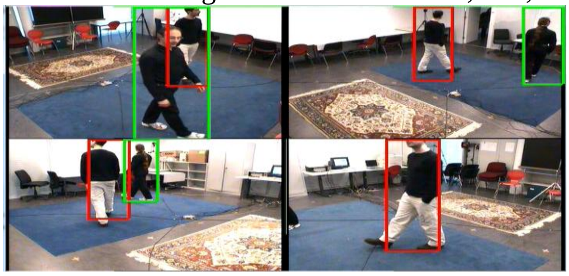
E. create and delete a target

At the measurement allocation stage, it is probable that the number of measurements be larger than the number of targets. Therefore, some measurements will be left without targets. In this case, one new target can be created per each measurement. These targets are marked as unapproved targets. If in a defined time panel adequate measurements are allocated to them, the targets will be approved as main targets.

If for any reason, the path algorithm loses an individual, that might mean that no measurement has been allocated to that path in some successive frames, and that special target will be eliminated. This is done as follows: a time panel is specified for each target. If specific number of measurements are not allocated to the intended target in this time panel, the state vector of that target will be eliminated. For example, if in the last 10 frames 6 measurements are not allocated to the intended target, the target will be eliminated.

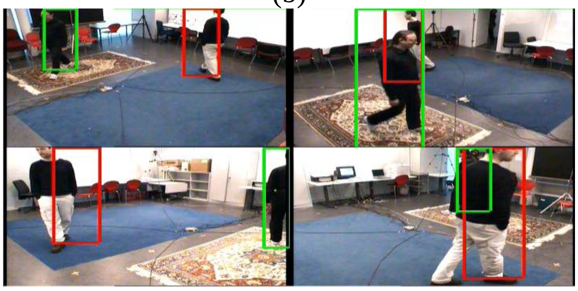
II. SIMULATION AND RESULTS

To approve this study's methodology, the obtained results are presented on EPFL website [18]. In the presented results, every individual has been shown as a colorful rectangle that shows his/her limit. This website contains several videos of different scenes. Each scene has been recorded simultaneously with four cameras at the rate of 24 frames per second. Figure 6 shows the tracking results at frames 580, 690, and 780.



(a)

(b)



(c)

Fig. 6. Results of implementation on EPFL dataset. The results in (a),(b),(c) images, in each image, from up to down and from left to right relate to views of cameras1, 2, 3 and 4, respectively.

As can be seen in table1, to evaluate the accuracy of the methodology, the obtained results were evaluated using three accuracy assessment criteria: precision, recall and f-measure. The first criterion shows the extent that the methodology recognizes and tracks the moving targets of the scene. The result is 97/55%. The second criterion shows how many people have been tracked correctly. The obtained result is 92/3%. The third criterion is a combination of the first and second ones and shows 94/8%.

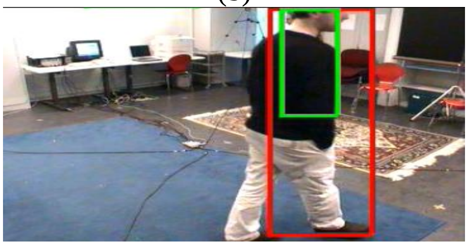
TABLE 1. Accuracy evaluation of the presented approach

	Precision	Recall	F-Measure
Proposed Method	92.3 %	97.55 %	94.8%

To evaluate the presented algorithm, a comparison was done between the algorithm results of this study on EPFL dataset and the results obtained from Kalman filter and graph cut algorithms individually (figure 7). The comparison has been done on camera1's view.

(a)

(b)



(c)

Fig. 7. Results of comparison between tracking methods of this study and Kalman filter and graph cut algorithms. (a). Kalman filter approach,(b). Graph cut approach, (c). Paper approach.

To evaluate the accuracy of tracking with Kalman filter and graph cut methods, as can be seen in tables 2 and 3, results obtained from these methods were assessed using three assessment criteria: precision, recall and f-measure.

Table 2. Evaluation of Kalman filter tracking method individually

	Precision	Recall	F-Measure
Kalman filter	91.3 %	88.3 %	89.8%

Table 3. Evaluation of graph cut tracking method individually

	Precision	Recall	F-Measure
graph cut	90.7 %	95.54 %	93.0%

Results of a comparison between tracking people through Kalman filter and graph cut tracking methods and the presented method are shown in figure 8.

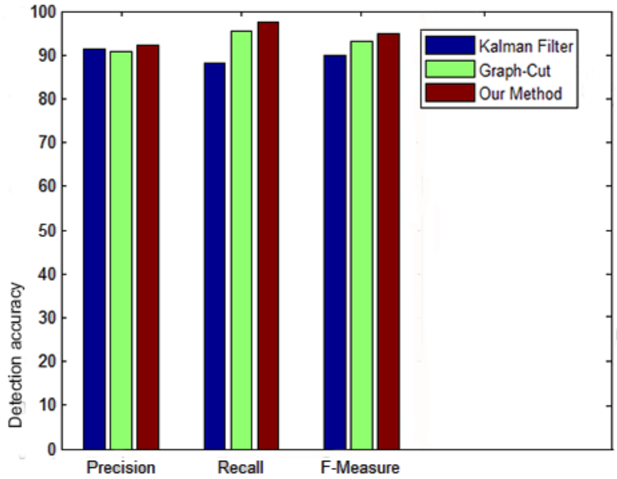


Fig 8. A comparison between accuracy of Kalman filter and graph cut tracking methods and the presented method

III. CONCLUSIONS

In this study, a method of tracking people based on graph cut and kalman filter was presented using several fixed cameras. People’s junction points at homography level in each camera appear clearly even though the people have partial or full overlaps. Using graph cut leads to more accurate junction points of people. Using Kalman filter to track people and obtain the movement path may result in the improvement of the measurements obtained from people’s positions. Therefore, people are tracked well through this method even though they have partial or full overlaps.

References

[1] X. Wang ,"Intelligent multi-camera video surveillance: A review” Vol. 34, 3–19, January 2013.
[2] roberto cipolla, janusz kacprzyk, “computer vision (detection, recognition and reconstruction),” editor- janusz kacprzyk, springer,volume 285, 2010.
[3] a. yilmaz, o. javed, and m. shah, “object tracking: a survey,” acm j. computing surveys, 2006.
[4] S. m. khan and mubarak shah," tracking multiple occluding people by localizing on multiple scene planes," ieee transactions on pattern analysis and machine intelligence, vol. 31, no. 3, march 2009.
[5] Y. ajmal sheikh and mubarak shah,” trajectory association across multiple airborne cameras”, ieee transactions on pattern analysis and machine intelligence, vol. 30, no. 2, february 2008.
[6] R. martín and j. m. martínez,” a semi-supervised system for players detection and tracking in multi-camera soccer videos”,doi10.1007/s11042-013-1659-6,springer + business media new york 2013.
[7] s.m. khan and m. shah, “a multi-view approach to tracking people in crowded scenes using a planar homography constraint,” proc. ninth european conf. computer vision, 2006.
[8] dejan arsi´c, atanas lyutskanov," multi camera person tracking applying a graph-cuts based foreground segmentation in a homography framework", 978-1-4244-5504-1/09/$25.00,2009 ieee.
[9] s. m. khan and m. shah, “a multiview approach to tracking people in crowded scenes using a planar homography constraint,” in eccv, 2006.
[10] d. freedman and m.w. turek, “illumination-invariant tracking via graph cuts,” proc. ieee conf. computer vision and pattern recognition, 2005.
[11] Wei Zhang, Xiaochun Cao, " Detecting and Extracting the Photo Composites Using Planar Homography and Graph Cut", ieee transactions on information forensics and security, vol. 5, no. 3, september 2010.
[12] Yuxian Li1, Yifu Du2,” Image Segmentation Method Based on Improved Graph Cut Algorithm”,IEEE,978-1-4673-0089-6/12/$26.00 ,2012.
[13] V. Kolmogorov, S. Vicent, “Graph cut based image segmentation with connectivity priors,” IEEE Conference On Computer Vision and Attern Recognition,vol.12,pp.767-774,2008.
[14] w. hu, x. zhou, m. hu, s. maybank,” occlusion reasoning for tracking multiple people”, ieee trans. circuits syst. video technol,114–121, 19 (2009).
[15] khan, s., shah, m., 2009. tracking multiple occluding people by localizing on multiple scene planes. ieee trans. pattern anal. machine intell. 31 (3), 505–519.

[16] wenbing dang, scharf, l.l " extensions to the theory of widely linear complex kalman filtering", signal processing, IEEE transactions on volume:60 , issue: 12 , 6669 – 6674, 20 August 2012.
[17] pieper. m. ; kummert. a," image prediction for virtual environments by means of kalman filter based 3d object tracking", page(s): 30 – 35, publication year: 2005.
[18] http://cvlab.epfl.ch/data/pom