Multiple Soccer Players Tracking

Nima Najafzadeh
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
najafzadeh@ce.sharif.edu

Mehran Fotouhi
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
fotouhi@ce.sharif.edu

Shohreh Kasaei
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
skasaei@sharif.edu

***Abstract*— This paper, describes a solution for tracking multiple soccer players, simultaneously, in soccer ground. It adapts Kalman filter for tracking multiple players. Adapting Kalman filter is divided to four main tasks. The first task is defining the state vector for multiple object tracking. The second task is determining a motion model for estimating the position of soccer players in the next frame. The third task is defining an observation method for detecting soccer players in each frame. Finally, the fourth task is tuning the measurement noise covariance and estimating noise covariance. In the third task, a novel observation method for detecting soccer players is proposed. This method divides the player body into three parts and calculates the histogram of each part, separately. Also, an algorithm for updating the reference object patch is introduced in observation method. Each task is discussed in detail and the promising performance of the proposed method for tracking soccer players when run on the Azadi dataset is shown.**

Keywords—Multi-object tracking; Soccer player tracking; Observation method; Kalman filter.

I. INTRODUCTION

One of the most important processes that is performed on consecutive video frames is object tracking. There are several reasons for accomplishing this task. It is required for some high level tasks including automated surveillance, traffic monitoring and navigation, human-computer interaction, and motion-based recognition [1].

Object tracking can be considered as two different problems of data association and trajectory estimation. Data association is related to assigning consistent and compatible labels to objects. The farther is a serious problem especially when occlusion occurs. The latter, is important in multi-target tracking when the tracked objects are similar to each other. On the other hand, frame quality, shadow of objects, and change in lighting condition affect the trajectory estimation methods. Therefore, object tracking remains as a challenging task in real world applications.

Trajectory estimation can be carried out by well-known filters like Kalman or particle filters. Every estimation filter utilizes an object detection approach to improve its estimation or to initialize its algorithm. Other approaches act in a tracking-by-detection strategy, where the targets are detected in a pre-processing step. Clearly, object representation and modeling are important parts of both approaches.

Recently, many works are reported based on tracking by detection methods. In these approaches, the target is represented by histogram or shape features or other object modeling techniques and is detected in every frame, independently. The trajectories are then estimated. These approaches can handle re-initialization if a target has been lost or the number of objects is not predetermined in each frame. Avoiding the excessive model drift is another advantage of these approaches [2]. The drawback of detection-based object tracking is in multi-object cases in which data association is significantly more difficult.

To deal with such mentioned problems, researchers did a lot of work and gained many good achievements. In [3], researchers proposed a real-time object detection-based tracking method which selects features based on a semi-supervised online learning technique. A multi-object tracking method using Kalman filter is suggested in [4]. A background subtraction method is utilized to detect and extract moving objects. Gao *et.al* in [5], used a multi-Kalman filtering to overcome multi-target tracking challenges in cluttered environments. The suggested approach was semi-automatic and relied on a 3-D motion estimation-based framework for object tracking in cluttered environments.

Recently, researches tries to find an optimal set of trajectories by energy minimization. In [6], Milan *et al.,* have reported an energy minimization approach for multi-target tracking which handles occlusion. In [7], an optimization approach is used to overcome data association and trajectory estimation challenges. Discrete optimization with label costs has applied for data association while trajectory estimation has been determined by continues optimization.

In [9], Berclaz et.al proposed a graph based approach which ignores appearance completely. So it is robust against the appearance changes. This method uses K-shortest path for data association and has a comparatively low computation complexity. Because of ignoring appearance, this method may produce identity switches when people come close to each other. Therefore, the mentioned approach is not suitable in crowded frames especially for sports videos. In [10], Ben

Shitrit et.al have addressed this problem and solved it using extending the formalism introduced in [9] to account for individual identities.

In this paper, a multi-target object tracking method is proposed for tracking soccer players while the number of targets is predetermined. Kalman filter is used as an estimator filter. A novel object modeling is used to describe players in soccer matches. Kalman filter is adapted for multi-target tracking while just one object is tracked directly and considered as the main target. The trajectory of other targets is estimated based on the main target. Furthermore, observation and estimation noise tuning is discussed in detail based on the motion and observation model that is used. Our future work will be on tracking unknown number of targets.

The rest of this paper is organized as follows. In Section II, the basic concepts of Kalman filter object tracking, object detection method, and the tracking model for individual targets are reviewed. The proposed observation model and Kalman filter adaptation are introduced in Section III. The experimental results are given in Section IV followed by the conclusion in Section V.

II. Object tracking by Kalman filtering

Kalman filter is an estimator for a discrete linear process. This estimator predicts the next state of a process with a linear equation, given the current state of the process as an input. Output of this equation, called prior estimation, in the next step corrects the prior estimation by measuring the process state directly. The final result is called posterior estimation.

Kalman describes the state of a process by a vector, called state vector. It is an arbitrary column vector that contains some parameters. State of a process is described by these parameters.

Object tracking is a discrete linear problem. It is a discrete problem because the position of the object is needed in every frame. It is linear because the movement of the object is considered to be linear in short time periods.

When Kalman filter is used for object tracking, the prediction equations estimate the object position in the next frame. Observation method obtains the object position directly from the next frame. And, correction equations, correct the estimated position with respect to the observed position of the object.

III. Propose method

Kalman filter is a general framework for object tracking. For tracking object, it must be adapted and its parameters must be tuned. Adapting Kalman filter includes four main tasks. The first task is defining the state vector. This vector contains parameters that describe position and other properties of the object in each frame. The second task is specifying the motion model of the object. Motion model is a linear equation that specifies how the object moves. This model is used for estimating the object position in the next frame, given the position in the current frame. The third task is determining the observation method. It is used for detecting the object in each frame. The fourth task is tuning the noise covariance for estimation and observation. As observation and estimation that is obtained from equations may have some error, noise is used to model this error.

In this paper, we adapt Kalman filter for tracking multiple soccer players. Also, we propose a novel method for observation. In our approach, we consider one player as the main object that track it, directly. Other players are considered as subsidiary objects that are tracked indirectly. When an object is tracked directly, the state vector contains the position and velocity of that object. But, when an object is tracked indirectly, the state vector contains its relative position to the main object.

In the following of this section, adapting Kalman filter for tracking multiple soccer players is discussed in four parts. In part A, a state vector for multi-object tracking is introduced. In part B, a linear constant velocity motion model is proposed for tracking soccer players. In part C, we propose a novel observation method for soccer player tracking. And, in part D, a method for tuning observation noise and estimating noise is introduced with respect to the motion model and observation method that is proposed in Sections B and C.

A. State vector

When Kalman filter is used for object tracking, each frame is considered as a state and it needs a state vector for describing it. Here, we introduce a state vector for multi-object tracking. The propose state vector is

$$X_k = [w\ h\ x\ y\ v_x\ v_y\ x_1\ y_1\ \ldots x_n\ y_n].$$

In object tracking, the object that is tracked is specified by a bounding box that surrounds the object. This box is called tracking window. In above vector, w and h are the width and height of tracking window.

As discussed before, in proposed method for multi- object tracking one object is considered as the main object for tracking. In the above state vector, x and y are the position of the main object and, v_x and v_y are the velocity of the main object in direction of x and y axis. The velocity of an object is the displacement of that object divided to the time interval that displacement happens. On the other hand, video is a sequence of images that have a fixed time interval between two adjacent frames. Therefore, we can ignore time in the velocity formula and calculate the velocity of an object as

$$v_x = x_k - x_{k-1} \tag{1}$$

$$v_y = y_k - y_{k-1}\,. \tag{2}$$

Other objects are considered as subsidiary objects. These objects are not tracked directly. So, their relative positions are in the state vector. In the above state vector, x_m and y_m are the position of the m^{th} subsidiary object relative to the position of the main object.

B. Motion model

Kalman filter consist of two main groups of equations, the prediction and correction. When Kalman is used for object tracking, prediction equations estimate the object position in the next frame, given the object position in the current frame. Estimation equations are linear equations that use the state vector of current frame for estimating the state vector of the next frame. Prediction equation is shown below [8]

$$\hat{x}_k^{-} = A\hat{x}_{k-1}^{-} + Bu_{k-1} . \quad (3)$$

In above equation, matrix A is the motion model. It relates the state vector of frame k to the state vector of frame k-1. In this section, we propose a motion model for tracking soccer players. There are many motion models that can be used for tracking, like constant acceleration motion model and constant velocity motion model. Since the video frame rate in the dataset is high, the object displacement in two adjacent frames is not too large. Thus, we do not need to use a complex motion model. Our proposed motion model and estimation equation are

$$x_k = Ax_{k-1} \quad (4)$$

$$A=\begin{bmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

According to above motion model and the state vector that was introduced in previous section, the relations between state vector parameters in state *k* and *k-1* are

$$x_k = x_{k-1} + v_x \quad (5)$$

$$y_k = y_{k-1} + v_y \quad (6)$$

$$v_{x\,k} = v_{x\,k-1} \quad (7)$$

$$v_{y\,k} = v_{y\,k-1}. \quad (8)$$

As illustrated in above equations, a constant velocity motion model for tracking soccer players is used. Equations (5) and (6) show the relation between object position in frame *k* and *k-1*. As the constant velocity motion model is used, velocity of object in frame *k* and *k-1* are equal. Also, the displacement of object between every two adjacent frames is constant and is equivalent to $\sqrt{v_x^2 + v_y^2}$.

C. Observation model

The second main group of Kalman equations are correction equations. These groups of equations, correct the estimation that is obtained from prediction equations. The correction is done with respect to observation that obtains object position directly from each frame.

Observation is detecting the object that is being tracked in each frame. In tracking, each object is specified by a bounding box that surrounds the object. Detection is done by comparing the reference object patch with the candidate object patch in each frame and choosing the candidate that has the least difference with the reference patch. This candidate is called the best candidate object patch.

For describing an object, an observation method is needed. The most common method for describing an object is the histogram of object patch. This method calculates the histogram on the bounding box that surrounds the object. Although, this method can be useful in some applications but for object tracking applications in crowd scenes (like tracking soccer players in the soccer field) it is not a good method for object description. This is because the histogram of patch only considers the color without considering the spatial distribution of colors.

Here, we propose a novel observation method for tracking of soccer players. It is a type of human tracking. Human body consists of three parts: head and neck, torso, and foots. According to these three main parts of human body, our method divides each patch horizontally into three distinct and same-size sub patches; as shown in Figure 1. We calculate the histogram for each sub patch, separately, and use the *Bhattacharyya* distance for comparing two sub patch histograms. For calculating the distance between the reference patch and the candidate patch, we use a weighting average formula. It calculates the difference between corresponding Sub patches and gives a weight to each difference according to its importance. Thus, the torso sub patches difference gains more weight than others. The formula that we use for weighting average differences is

$$\begin{aligned} &Bata \\ &= \big(bata(candidate\ first\ subpatch, reference\ first\ subpatch) \\ &+ 3 \\ &* bata(candidate\ second\ subpatch, reference\ second\ subpatch) \\ &+ bata(candidate\ third\ subpatch, reference\ third\ subpatch)\big) \\ &/5. \end{aligned} \quad (9)$$

Another important issue in observation method is specifying the reference object patch for each frame. The simplest way for specifying the reference object patch is determining the object patch in the first frame as a reference object patch for all other frames. But, as the appearance of the object changes with time this is not a good solution.

Figure 1: Three part patches.

In following, we propose an algorithm for specifying the reference object patch for each frame. Our algorithm updates the reference object patch with respect to its appearance changes. The pseudo code of the propose algorithm is shown in Figure 2.

Our algorithm updates the reference object patch based on the moving average formula. This algorithm uses all the best candidate object patches from the first frame to the current frame. But, the best candidate patches near the current frame gains more weight.

```
D=Histogram distance (reference patch , best candidate
patch)

  If  D<threshold1
      reference patch= best candidate patch
   Else if  (D>threshold1&&D<threshold2)
              Increase alpha until alpha <0.5
             & use moving average formula for updating
              Reference patch
    Else if (D>threshold3&&D<threshold3)
              Decrease alpha by rate 0.05 until alpha>0.2
              & use moving average formula for updating
              Reference patch
    Else
          Decrease alpha by rate 0.1 until alpha>0.2
          & Don't update reference patch
     End
```

Figure 2: Pseudo code of patch updating algorithm.

D. Parameters tuning

Two parameters that can have a significant impact on tracking results are the measurement noise covariance R and the estimation noise covariance Q. If these two parameters are considered to be constant during the process, the error covariance P_k and the Kalman gain K converge to a constant magnitude after several steps. It means that the filter gives a constant weight ration to the observation and estimation after those steps. Whereas, the weight ratio of observation and estimation must be dynamic and change with respect to the filter belief on the observation and estimation. Therefore, considering R and Q as constant parameters is not a good idea and they must be changes in time.

Here, we propose a method for tuning these parameters for tracking soccer players.

1) Tuning error covariance in estimation phase

As mentioned before, Q is the estimation noise covariance which relates to our belief on Kalman filter estimation. When Kalman has more belief on estimation than observation, Q decreases, and vice versa.

With respect to our motion model for tracking soccer players, players move linearly with a constant velocity in a specific direction in short time periods. Thus, when the direction or velocity of players changes, Q must increase. In this situation, we have less belief on estimation and weight of estimation must decrease.

2) Tuning error covariance in observation phase

As mentioned before, R is the measurement noise covariance which relates to our belief on Kalman filter observation. When Kalman has more belief on estimation than observation, R increases, and vice versa.

According to our observation method, we compare the candidate object patches with the reference object patch. When the difference between the best candidate object patch and the reference object patch is small, it means that we find a good candidate and thus R must decrease. But, when the difference between the best candidate patch and the reference patch is large, it means that we have found an improper candidate. In this situation, R must increase to decrease the weight of observation.

E. Oclussion handeling

In this paper we use merge and split method for handling occlusion problem. As discussed in previous sections, player detection is done as a part of tracking approach in each frames and, each player determined by bounding box that surrounds it. Therefore when two bounding boxes overlap with each other, it means that an occlusion occurs. In this situation two overlapped players consist a merged object. Tracking merged object continues until objects split from each other. When split happens we use object features that is used before merging for detecting players and labeling them correctly.

IV. EXPRIMENTAL RESULTS

We have evaluated the proposed algorithm for multi-soccer player tracking on the Azadi dataset and have compared its results with the baseline algorithm. The baseline algorithm uses the same state vector as our algorithm. For motion model, the baseline algorithm considers object position in current frame as an estimated object position in the next frame. Also, it uses histogram of patch as an observation method.

The Azadi dataset is prepared by *image processing lab* (IPL) of Sharif University of Technology. This dataset includes some video sequences that are captured from one of the domestic football matches in Iran Pro League, from several views. We used one of these sequences that is captured by PTZ cameras. The video sequences are captured with 25 frames per second sampling rate and size of each frame is 720×576. The reason for using Azadi dataset, for evaluating the proposed algorithm is that this dataset is a very challenging dataset. Because, the resolution of video sequences is low and therefore detecting players for tracking is very difficult in this dataset. Therefore, in addition to detection, estimation plays an important role for tracking players in this dataset.

a

b

Figure 3: Results of baseline method.

a

b

Figure 4: Results of proposed method.

We have implemented the proposed method and baseline algorithm with MATLAB 2013a. Both algorithms are tested on a computer with 4 Gigabyte of ram and Intel Core 2 duo CPU (2.2 Ghz). It is worth mentioning that the baseline method is faster than the proposed method. But, if the proposed method be implemented with C++ on a fast computer it can perform near real-time. Also, Results showed that proposed method perform much better than baseline method in tracking multiple soccer players.

Figures 3 and 4 show the result of the baseline algorithm and the proposed method, respectively. As it is shown in Figure 3, when two players with the same color come close to each other,

The baseline algorithm misses players. The cause of this error is that the baseline algorithm uses the histogram of a patch as an observation method. That histogram only considers colors without considering the spatial distribution of colors. Thus, the algorithm gets a region as a player that contains the right part of the left player and the left part of the right player (because it only considers the amount of yellow and green colors in a patch without considering spatial distribution of these colors). Also, the baseline algorithm considers the first frame object patch as a reference object patch. Therefore, when appearance of players change with time the baseline algorithm makes mistake in detecting players.

The proposed method was run on the same video sequence. As we see in Figure 4, the proposed method didn't miss the players. Because, when it searches in a region for a player, it does not search for a patch for which its histogram is similar to the reference patch. It searches for head, torso, and foots of the player. Also, as the reference object patch updates, it does not miss the player when the appearance of the player changes.

V. CONCLUSION

In this paper, an algorithm for tracking multiple soccer players was proposed. It used Kalman filter as a general framework and adapted it for tracking soccer players. Adaptation of Kalman filter was divided into four main steps. In the first step, a state vector for tracking multiple objects was introduced. In the second step, a motion model for soccer players was designed. In the third step, a novel observation method for detecting soccer players was proposed and an algorithm for updating the reference object patch according to the changes in players' appearance was introduced. In the fourth step, a method for tuning the measurement covariance noise and observation covariance noise with respect to the motion model and observation method was proposed. Finally, the proposed method was compared with the baseline algorithm that uses the object patch as an observation method. The results showed the superiority of the proposed method.

VI. Reference

[1] A. Yilmaz, O. Javed and M. Shah, Object tracking: A survey ACM Computing Surveys, ACM Computing Surveys Vol. 38, No. 4, 1-45, 2006.

[2] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. CVPR, 2000

[3] H. Grabner, C. Leistner, and H. Bischof. Semi-supervised on-line boosting for robust tracking. ECCV, 2008.

[4] Li, Xin, et al. A multiple object tracking method using Kalman filter. Information and Automation IEEE International Conference on., 2010.

[5] Gao, Jean, Akio Kosaka, and Avinash C. Kak. A multi-Kalman filtering approach for video tracking of human-delineated objects in cluttered environments. *Computer Vision and Image Understanding* Vol 99, Issue 1,1-57, 2005.

[6] Milan, Anton, Stefan Roth, and Konrad Schindler. Continuous energy minimization for multi-target tracking. CVPR, 2013.

[7] Andriyenko, Anton, Konrad Schindler, and Stefan Roth. Discrete-continuous optimization for multi-target tracking. CVPR, 2012.

[8] G. Bishop and G. Welch. An Introduction to the Kalman Filter. in SIGGRAPH 2001, Course 8, 2001.

[9] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple Object Tracking Using K-Shortest Paths Optimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33, No. 9, 1806-1819, 2011.

[10] H. Ben Shitrit, J. Berclaz, F. Fleuret and P. Fua. Tracking multiple people under global appearance constraints. ICCV, 2011.