Sports Video Annotation and Multi-Target Tracking using Extended Gaussian Mixture model

Daneshwari Mulimani, Aziz Makandar

Abstract: Video offers solutions to many of the traditional problems with coach, trainer, commenter, umpires and other security issues of modern team games. This paper presents a novel framework to perform player identification and tracking technique for the sports (Kabaddi) with extending the implementation towards the event handling process which expands the game analysis of the third umpire assessment. In the proposed methodology, video preprocessing has done with Kalman Filtering (KF) technique. Extended Gaussian Mixture Model (EGMM) implemented to detect the object occlusions and player labeling. Morphological operations have given the more genuine results on player detection on the spatial domain by applying the silhouette spot model. Team localization and player tracking has done with Robust Color Table (RCT) model generation to classify with respect to their state data and Real Time background. Markov Chain Monte Carlo (MCMC) method has applied for background annotation process. Through which each court line tracing and labeling in the half of the court with respect to their state-of-art for foremost event handling process is performed. Extensive experiments have been conducted on real time video samples to meet out the all the challenging aspects. Proposed algorithm tested on both Self Developed Video (SDV) data and Real Time Video (RTV) with dynamic background for the greater tracking accuracy and performance measures in the different state of video samples.

Keywords: Line segmentation, Player Localization, HGT, RCT.

I. INTRODUCTION:

For eternity sports video pleads to large audiences. Recently there has been a remarkable expansion of sports science. Use of video has given more preference rather than image or documentary data. Hence, video is a collection of frames. Each frame represents the unique state of object. Object discovery from the frame and keep tracking the particular object throughout video sequence became a challenging research area in computer vision(CV) system. But player tracking is a problem of predicting the route of an object in the image plane with its dynamic moves around a scene. Here are some difficulties faced by the machine learning progression

1) Changes in objects with viewpoint, illumination 2) Biased occlusions of the target objects by other objects, 3) Complexity of the static/dynamic background subtraction 4) Environmental changes while capturing the video or noisy frames. 5) Video resolution and frames differing factors to analyze the accuracy in the results.

By considering these challenges and the research gap in the sports video analysis process proposed work initiate towards the new case study on “Kabaddi” video annotation system and implementing efficient CV model for occlusion handling and team localization for player tracking. It tends more significant dedication for the needful contribution. Respective image processing modalities are briefed in following section.

II. LITERATURE REVIEW

Many researchers have worked on improving sports video content annotation with domain specific issues in the past. The proposed work has followed the team game of more than 7 players like basket ball, soccer video and football. Much work has done on these three games in literature. The most relevant work for the team localization and player tracking methodologies are studied and reviewed as follows.

Jia Liu et al [1] proposed automatic player detection, labeling and efficient tracking in broadcast soccer video.

Fig 1: Architecture for the proposed methodology

Players’ position estimated by a boosting based detector. Player tracking is achieved by Markov Chain Monte Carlo (MCMC) method by estimating the Gaussian Mixture Model(GMM) for the pose detection. This method can reach high detection and labeling precision and consistently tracking in cases of scenes such as multiple player occlusions, sensible camera motion and pretense variation and achieved the 92.38% average precision. Likewise Amir hosseinAlavi [2] as proposed the investigation of football player tracking and labeling with Kalman Filter technique for motion based tracking for blob analysis. The algorithm shows the disparity in the accuracy that based on the slow and fast moments of player in each frame. While players run with constant speed, the accuracy of detection and tracking was very high and achieved the average precision of 87.5% in detection and tracking. Algorithm performance in [2] has not achieved improved results than [1]. Branko Markoski et al[3] proposed AdaBoost algorithm for basket ball game annotation mainly concentrating on detection of player face and body parts.
Here the video footage obtained from the single moving camera to train the player’s entire body including head, legs, arms and torso achieved 70.5% accuracy. To improve this accuracy towards player detection Tsung-Yu Tsai et al [4] introduced tactic analysis and key player detection by multiple learning instances to encode the spatial and temporal variation among the players. motion intensity map (MIM) makes more efficient to detect the multiple players simultaneously and is strong enough to distinguish players of different temporal lengths and at illogical spatial locations. Wei-Lun Liu et al [5] proposed Conditional Random Field (CRF) for player detection in basket ball game. Player prediction has done with Linear Programming (LP) Relaxation algorithm to assign the play-by-play text labeling to each player. The shot segmentation has done with Hidden Markov Model (HMM). The emission probability is formed by a Gaussian Mixture Model (GMM) RGB color histograms features and the transition probability is devised to promote a smooth change of shots. Deformable Part Model (DPM) was used to automatic localization of player in the field which has achieved the aspect ratio in the precision of 73% and a recall of 78%. Many other [6, 7, 8, 9, 10] have also introduced the player action detection and event recognition in the sports video content analysis. This bird view on sports video annotation on same game repeatedly which made us to look around the research gap towards the initiation of CV model for Asian game. Furthermore, our proposed work has extended to handle the major events [11] in the kabaddi scene which has made the remarkable contribution for the third umpire judgment.

III. OVERVIEW OF PROPOSED CASE STUDY AND METHODOLOGY

The modern gaming system made the huge demand in CV models in various sports video. Less attention towards the Asian game Kabaddi which is admired by the pro-Kabaddi tournament is taken as our case study for the research. The game has ground truth measures of 13” by 6” court view. A team contains 7 players to take a game and 2 can be substitute as defender and one rider from the opposite team. The goal is proposed methodology must achieve to track 8 players on the field (scrutiny has done in half of the court). The intrinsic chronological nature of video apparent by the evolution of video features typically shows wide variations in behavior. These variations create many complexities while raising the convention for object modeling and feature estimation. These have been conquering in the preprocessing stage. In this the input video clips are validated for frame rate, resolution and required video size. To do this MATLAB tool has its own functions to test the inputted video set. Then the input data has kept for the shot boundary (SB) detection process and Key Frame extraction (KFE) as shown in the architecture.

IV. OUR APPROACH

A. Kalman Filter For KFE:

The KF algorithm involves update steps prediction and correction. Prediction process remains measuring the state of object and its dynamic amendment in its position. Then in second, correction process, it updates the object previous state and captures the current status of the object. Then it will be taken as predicted object and labeled based on the prediction location. Each player is considered as a moving object and the background (track lines) will be taken as static object [21]. These are individually set as a measurement input. The Kalman tracker uses this data to successfully correct the system and obtain a probability state vector. The equation for the measurement matrix correction is given as

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \begin{pmatrix}
X \\\nY
\end{pmatrix} + K_t \begin{pmatrix}
X' \\
Y'
\end{pmatrix}
\]

Eq-1

The KF is a best outfit to achieve discrete time and linear State-Space system in the spatial domain [18,19]. It helps to locate the filter for tracking a moving object in a cartesian coordinate system by making an allowance for the constant velocity,

```
obj=vision.KalmanFilter
...(StateTransitionModel,MeasurementModel,ControlModel, Name,Value)) function I
```

B. Extended Gaussian Mixture Model For Object Detection:

The secretion probability is modeled by an Extended Gaussian Mixture Model (EGMM). RGB color histograms, and the transition probability features are formulated to persuade a smooth change of shots. A GMM is a probabilistic model use to assume all the data points generated from a mixture of a finite number of on average distributed gaussian subpopulation among constant fields with anonymous parameters. It is a simplified k-means clustering to integrate information about the covariance structure of the data and also the centers of the hidden gaussian. It works with Expectation-Maximization (EM) process that got implemented by the GM object for mixture of gaussian models. EM is a statistical approach to get through problem by an iterative process. Initially it assumes random components (data points, extracted from k-means clusters, those are distributed around the origin on spatial domain) and it computes probability of each component of the model. Next, it squeezes the parameters to maximize the probability of the data. Same will be repeated to assure local optimum [11]. Then Bayesian Information Criterion (BIC) is computed with the help of confidence ellipsoids for multivariate models to assess the number of clusters in the data. These GM ellipsoids gets generate the covariance of different classes with different options estimated: spherical, diagonal, tied or full covariance. The model has shown in the figure 2. These distribution follows the two scaled (variance) and shifted (mean) with normal distribution parameters.

© Copyright: All rights reserved.
The process of the model making this assumption is known to be GMM. EGMM mainly used for the feature extraction and extensively in multiple object tracking in which number of mixture components and their mean prediction object location has done at each frame in a video sequence [16].

This method can adapt to changing circumstances and a wide range of irregular foreground objects being introduced in the scene, while retaining control over the number of Gaussians in the mixtures. Using these parameters the foreground (FG) object silhouette is shown as Fig 3(i) and while tracking object rectangular patch Fig 3(c) is represented [12]. Then the prediction measure of each silhouette is estimated.

### C. ROBUST COLOR TABLE (RCT) BASED TEAM LOCALIZATION:

It is a process of high-level object representation in image pixel like separation from background with required foreground object. With thresholding it is possible to segment the image based on color parameters. Example red pixels in the image can be separated from its surroundings colors [15, 16]. In this process pixel in a range fall within the range of the threshold and remaining are rejected. The thresholding process involves the test against the function as shown

\[ T = \{ x, y, p(x,y), f(x,y) \} \]

function - 1

Where \( f(x,y) \) is the gray level at the point \((x, y)\) and, \( p(x,y) \) represents some local property of the point such as average gray level of a neighborhood values centered on \((x, y)\) coordinates. Where above threshold function represents different sets of pixels, hereby those labeled 1 communicate to foreground object, while pixels labeled 0 communicate to the background state.

In the algorithm color formatting method present the color Table generation. In MATLAB, RGB color formats use to represent any standard color or brightness using a combination of RED, GREEN and BLUE components (1, 2, 3). It typically stores as a 24 bit number using 8-bit for each color components (0 to 255). flagM indicates the Variable to select the T1 table and T2 Table for color detection 1 for Team 1, 2 for Team Then we calculate the maximum and minimum color threshold for individual team i.e. T1 and T2. Each team color Table plotted with matrix size is 5X6.

\[ \text{meanColorTeam1} = \text{colorTable(flagM)} \]

function - 2

Based on the flag value respective color table will be generated. The table for each R, G, B values for min and max are listed in the 5X6 matrix bellow.

Color Table for Team 1
team = [Max R Value, Min R Value, Max G Value, Min G Value, Max B Value, Min B Value]

function-3

team = [62, 15, 37, 8, 104, 52; 68, 11, 62, 06, 125, 56; 67, 12, 53, 06, 115, 38; 108, 29, 76, 12, 167, 57; 47, 17, 30, 03, 95, 15];

Same followed for team 2 maximum mean values for RGB has calculated using meanMax() function will be designed. Minimum Mean Values for RGB calculated using meanMin() function after team table selection compute its mean values, + or - difference is considered for accurate results. After computing maximum and minimum mean values compute their difference values i.e ± maximum and minimum R, G and B

\[
RGB(x, y) = I_R(x, y) + I_G(x, y) + I_B(x, y)
\]

function-4

Based on min max value the blob is generated. Here we generate the Mean (max, min) of RGB (mm-RGB) where RGB(x, y) is sum of RGB intensity at pixel coordinate values (x, y) where as IR(x, y), IG(x, y), and IB(x, y) represents red, green and blue intensity at pixel coordinate (x, y), respectively. While using 8 bit to cryptogram color intensity, then we can get the mm-RGB value of 0-765. Second, we divide the mm-RGB value into 16 bins, and then create the histogram of mm-RGB [17].

V. RESULTS ON THREE DATA SETS:

Proposed method applied on three data sets. One: Self Developed Video (SDV) and second: Real Time Video (RTV) with Static Background (SBG) third: RTV with Dynamic Background (DBG) explained as follows . SDV: has designed by plating the indoor KABADDI (half court) as per the ground truth measures. International Kabaddi player of KSWU Woman player were taken part of this game. With complete guidance of coach and game rules. Around 38 GB data got captured using three 16MP cameras in three different angles. Since it is very high resolution to apply on the MATLAB we converted the video pixel resolution to 16:9 aspect ratios. Line detection has done with Hough Grid Transformation method (HGT)[21]. The BG segmentation and the line detection are mapped with FG evaluation process implemented by RCT method. Both methodologies exposes the highlighting court line with player tracking process with supervised dynamic state of the each object inside the play court. The results showed the clear tracking position of the each player with respect to their team color specific bounding box generation for every dynamic position throughout the scene.

Fig 5: Team wise segmentation results of SDV data set. Team 1 indicates the rider; team 2 indicates the defender segmentation. Likewise a, b, and c are three different results are shown

Fig 6: Results of court line detection, player tracking and labeling with respect to ground truth values, and object classification for T1 rider and T2 defenders.

A. RTV with SBG: has captured the real data from the “South Zone Inter University Woman’s Kabaddi Championship 2017-18”, Organized by Alva’s, Mudabidari, from Nov 16 to 19, 2017. Videos are captured with professional camera of 32 MP HD qualities with 720p video, because it uses more frames per second and has a higher bit rate. Cameras were fixed in such a way that should cover the all possible events and game points as well. In this 50GB of data set we get the complete authentic SBG videos. The algorithm has tested on 255 to 500 frame set; the segmentation and tracking results are shown bellow.

Fig 7: Team wise segmentation results of RTV data set. Team 1 indicates the rider, team 2 indicates the defender segmentation. Likewise a, b, and c are three different results are shown

B. RTV with DBG: are the downloaded videos from the YouTube containing Dynamic Background (DBG) with moving camera. Because of the major technical issues like non rigid configuration, dynamic background, moving camera, low quality, noisy frames, and compressed data while uploading and downloading from the server, makes awful results on the algorithm implementation process. The videos downloaded have different image resolution. We have taken 1280x720 pixels and were acquired at 25 fps with. More than 20 video samples of 8 to 15 seconds have taken for the testing.
Fig 8: Player detection for 1 rider and 6 defenders in the Mangalore data set, no missed detection, and FN=0 no false detection in any frames.

Fig 9: Team localization in RTV with DBG (data set from Pro-kabaddi tournament), because of fast motion and moving camera we can find missed detection in some frames, but no wrong detection so we take FN as 0

We achieved compatible efficiency with the proposed algorithm. It proves that algorithm can be applicable to any real-time Kabaddi videos as the third umpire judgment. Some of False Positive found due to boost detection in RTV with DBG data because of several consecutive frames caused by video blur. It also can be perceived that the performance of labeling is extensively improved in many frames.

VI. EVALUATION OF RESULT AND DISCUSSION:

We measure the performance of algorithm based on the quantitative analysis in terms of binary classification taken in the object detection process. In our case study we have two objects to be classified like team 1 as T1 and team 2 as T2 with respective color bonding boxes to each player while tracking. These detection ratios are measured by standard video image processing (VIP) performance metrics. These are measured in terms of true positive (TP) correctly classified object, false positive (FP) missed detection, false negative (FN) wrong detection. The f-measure will be calculated as follows.

**Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. *also called as Positive Predictive Value (PPV)*

\[
\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq-5}
\]

**Recall (Sensitivity)** - Recall is the ratio of correctly predicted positive observations to all observations in actual class, recall is called sensitivity.

\[
\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq-6}
\]

**F-measure**: F1 Score is the weighted average of Precision and Recall.

\[
\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Eq-7}
\]

The figure 9 shows the Receiver Operator Characteristic (ROC) curve for the predicted classes. The curve represents the true positive and false positive classes. True positive is the number of correctly classified foreground pixels and false positive is the number of background pixels that is incorrectly classified as FG.

**VI. EVALUATION OF RESULT AND DISCUSSION:**

We measure the performance of the algorithm based on the quantitative analysis in terms of binary classification taken in the object detection process. In our case study, we have two objects to be classified like team 1 as T1 and team 2 as T2 with respective color bonding boxes to each player while tracking. These detection ratios are measured by standard video image processing (VIP) performance metrics. These are measured in terms of true positive (TP) correctly classified object, false positive (FP) missed detection, false negative (FN) wrong detection. The f-measure will be calculated as follows.

**Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. *also called as Positive Predictive Value (PPV)*

\[
\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq-5}
\]

**Recall (Sensitivity)** - Recall is the ratio of correctly predicted positive observations to all observations in actual class, recall is called sensitivity.

\[
\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq-6}
\]

**F-measure**: F1 Score is the weighted average of Precision and Recall.

\[
\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Eq-7}
\]

The figure 9 shows the Receiver Operator Characteristic (ROC) curve for the predicted classes. The curve represents the true positive and false positive classes. True positive is the number of correctly classified foreground pixels and false positive is the number of background pixels that is incorrectly classified as FG.

### Table 1: Quantitative Analysis Based on Binary Classification with Performance Metrics

<table>
<thead>
<tr>
<th>DATA BASE</th>
<th>T1</th>
<th>T2</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>RECALL</th>
<th>PRE</th>
<th>F-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDV</td>
<td>87.5%</td>
<td>12.5%</td>
<td>0</td>
<td>1.00</td>
<td>0.87</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTV SBG</td>
<td>87.5%</td>
<td>12.5%</td>
<td>0</td>
<td>1.00</td>
<td>0.87</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTV DBG</td>
<td>79%</td>
<td>21%</td>
<td>0</td>
<td>1.00</td>
<td>0.79</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since we have initiated the new case study in the VIP field, it is very challenging task to make the compatible study with our proposed work. Hence, as per the literature study we have analyzed the team games, their contribution and research gap while implementing this methodology.

**Comparison**

**Accuracy** = True Positive / (True Positive + True Negative)*100. Eq-8

### Table 2: Comparison Between Existing Methods on Various Team Games

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdaBoost (3)</td>
<td>70.45%</td>
</tr>
<tr>
<td>CRF (Conditional Random Field) (18)</td>
<td>85%</td>
</tr>
</tbody>
</table>

---

**Fig 9: ROC curve for the TP and FP**
Kalman Filter(2)  87.5%
KF+DPM (5)  89%
Neural Network  91.08%
Proposed Method (EGMM and HGT)  95.85%

V. CONCLUSION

The complete study has gone through with video content analysis on new case study including design of new data set in all the aspects and comparing them with BCV as fit to the MATLAB tool. On these sophisticated data set preprocessing and occlusion handling has done by EGMM and KF technique. Further considering the positive result of SDV and RTV, the RCT algorithm implementation has applied for team localization and accurate tracking for FG object. The extended HGT algorithm is used for BG annotation and line segmentation, which has used for relevant event detection in the Kabaddi video. Proposed algorithms gives desired results in both static and dynamic background videos as stated in the results clearly. This can be further useful for the complete live game evaluation process for automated third umpire judgment.

ACKNOWLEDGMENT

There is no authorized data set available for the research on the Kabaddi game, since have created real-time data-set. This was done with the supervision of Dr.Rajakumar Malipatil, Director of Sports Department, AUWVK, for the right ground truth examination and game played by AWUPED International Kabaddi players. Another data set was taken from International Kabaddi Tournament Organized by Mangalore University in Alva’s sports team Moodabidri. We are indebted to both PED team for their supervision, which made potential to develop valuable data-set to process our research study further.

REFERENCES

19. A robust method for automatic player detection in sport videos, 2017

AUTHOR PROFILE

Daneshwari Mulimani, Research scholar in Karnataka state women’s University Vijayapur. Received MSc Mastered Degree under Rani Chennamma University, Belgua. Working as Assistant Professor for department of BCA and MSc(CS) in BLDEA’s S.B Arts and KCP Science college Vijayapur with seven years of teaching experience. Area of research is “Image Processing”. During my research journey published 11 papers in reputed journals like Springer, IEEE digital library, Scopus indexed journals and UGC recognized journals. Received “Best paper” award in IEEE conference organized by ICCUBEA Pune in 2017. Received Best researcher 2016 Award by IURULA affiliated with Malaysia in 2018. Presently research paper on my complete work got accepted by Springer conference organized in London.

Aziz Makandar, is serving as dean of Akkamahadevi Karnataka State Women’s University, Vijayapur, since 2018. Before joining this university, he has put 7 years of service as Associate Professor, Department of Computer Science & Engineering, K B N College of Engineering, Gulbarga. He obtained M.Tech(Computer Science & Engineering) from Vishveshwarayya Technological University, Belgua, Karnataka. His doctoral research work was on “Image Compression Techniques with Applications to Medical Imaging”. So far, he has 25 years of teaching and research experience in the field computer science. He is a member of the IEEE and the IEEE Computer Society.