

Skeletonization of Players in Dynamic Backgrounds Using Discrete Curve Evolution

Narra Dhanalakshmi, Y. Madhavee Latha and A. Damodaram

Abstract Skeletal part extraction of the human or player becomes important in applications like developing the gaming consoles, event prediction in sports, gait based human recognition and classification of human activity etc. The efficiency of the said applications depend on how efficiently the skeletal part is extracted and the extraction of skeletal part is influenced by dynamic background of the video, occlusion and resolution of the video. So in this work, we proposed a method to extract the skeletal part of the sports man with a varying background thereby facilitating the subsequent analysis. Histogram of Oriented Gradients (HOG) is used to detect the human region by making use of Support Vector Machines (SVMs) and then Graph Cut technique is applied to remove the background to extract only the foreground in the form of silhouette. Finally a skeletal pruning method is applied which is based on contour partitioning method such as Discrete Curve Evaluation (DCE) technique. The proposed method is tested on sports video like cricket and shows effectiveness of the method in extracting the skeletal shape from the video with dynamic backgrounds.

Keywords Histogram of Oriented Gradients (HOG) • Graph-cut segmentation • Silhouette • Skeletonization • Discrete Curve Evaluation (DCE)

N. Dhanalakshmi (✉)
ECE Department, V.N.R. Vignana Jyothi Institute
of Engineering & Technology, Hyderabad, India
e-mail: dhanalakshmi_n@vnrvjiet.in

Y. Madhavee Latha
ECE Department, Malla Reddy Engineering College for Women, Hyderabad, India
e-mail: madhuvsk2003@yahoo.co.in

A. Damodaram
Sri Venkateswara University, Tirupati, Andhra Pradesh, India
e-mail: damodarama@rediffmail.com

1 Introduction

Improvements in video processing technology and computation technology motivated researchers to use computer vision and pattern recognition methods for analysis of videos. Now a days, this kind of research tremendously increased in sports videos in indexing, retrieval and event detection and even prediction on real time. Prediction of an event will be happened by analyzing the player in successive frames from a video. This analysis will be divided into different sections (1) Detection of a player, (2) Subtracting the background from the player (3) Skeletonization, (4) Prediction of an Event. So, accuracy of the prediction depends primarily on player detection, detection of player is a challenging task due to great variability in appearances, poses and also due to large variations in the background.

The important step in skeletonization is detection of human region i.e. human silhouette. So far, many methods have been developed in this area. Different authors used different approaches to detect the players. A player can be detected using background subtraction methods [1]. Pascual [2] used a player separation algorithm without background subtraction. Some of the techniques use color models with special templates to calculate likelihood maps for color templates matching to extract an object of interest [3]. Cloud System Model (CSM) framework [4] is designed to handle 2D articulated bodies in order to segment humans in the video. This methods works when the player facing towards the camera. Matthias Grundmann [5] uses hierarchical graph-based algorithm to segment the human. The next step in Skeletonization process is removing the background in order to precisely extract the player. Removing the background can be done using different color space approaches like Hybrid Color Space (HCS), RGB color space, and $L^*U^*V^*$ color space [6]. Most of the existing methods used low level features. But these methods are not able to segment the player precisely.

After achieving human silhouette, skeleton part should be extracted. To achieve this, the methods [7, 8] use iterative algorithm to extract skeleton part, while keeping the topological structure of pixels. Accuracy and smoothness of the skeleton may be achieved by thinning.

Towards this, we presented a novel and effective way for automatic retrieval of skeletal part of the player from the sports video sequences like cricket for further processing like prediction of an event or classifying the event. We proposed a system for skeletonization of the sports man, this uses the combination of HOG and SVM for the detection of a player, Graph-Cut segmentation to subtract the background and Discrete Curve Evolution method in order to extract the skeletal part. The updated model provides more accurate skeleton for further video analysis like even prediction or classification.

This paper has been organised into following sections. Section 2 gives background information of the methods. Section 3 discusses the proposed method and Sect. 4 explains experimental results. In Sect. 5, we draw conclusions and discuss about future work.

2 Background

The proposed system employs gradient based algorithm such as Histograms of Oriented Gradients (HOG) descriptors are trained using SVM for detecting human region and Graph cut method for segmenting player and non player regions in the sports video sequence and Discrete Curve Evaluation for skeletonization. We improved our work [9] and extended the work to extract skeletal parts.

The HOG descriptor [10] was focused on the detection of pedestrian (human) by calculating gradients (G_x , G_y) in both the horizontal and vertical directions for all the pixels in the frame on overlapping basis in order to improve the performance. These HOG descriptors of both player and non player are trained using SVM [11]. However, the detected region of the player alone may not be directly suitable for action analysis of a player, since it also captures gradients of non human neighboring pixels. As a result of this, the accuracy of evaluating human action may be reduced.

Graph cut [12] method represents each frame as a graph with nodes and edges, where edges are assigned with some non negative weight or cost based on edge weighting functions. This function clusters the pixels that possess similar characteristics. Greig et al. [13] discovered the powerful optimized graph cut algorithm for solving many problems in the field of computer vision. But, the Graph Cut method alone is not enough to segment only foreground i.e., player, when the video contains signification motion in successive frames and variation in pixel intensities. Due to this background pixels tend to be segmented as foreground. So, we eliminated the problems occurred when HOG and Graph Cut individually applied by combining both HOG and Graph-cut methods in our paper.

Most of the skeletonization methods work well when proper silhouette is given. So, we have used the Discrete Curve Evaluation (DCE) method used in [8] for skeletonization. This DCE simplifies the contour of the segmented human silhouette and then pruned the skeleton by contour partitioning.

3 Proposed Method

We used the combination of above three methods to achieve our goal of skeletonization. An overview of the major steps involved in our method is shown in Fig. 1.

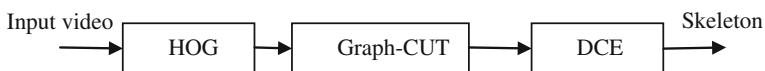


Fig. 1 Block diagram of the proposed work

HOG is used to extract only players as it is popular as human detection algorithm. HOG descriptor is derived by splitting each frame in terms of blocks and then each block into cells. Gradients (G_x, G_y) are computed in both the horizontal and vertical directions for all the pixels of all the cells in the frame. Then gradient magnitudes (G) and directions (θ) are computed by using the following expressions Eqs. 1 and 2:

$$G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$\theta = \tan^{-1}\left(\frac{G_x}{G_y}\right) \quad (2)$$

Once the HOGs are computed, we do provide the training the SVM classifier with positive class HOGs of the players and negative class HOGs of non-player or human. A player detection of the frame is done by scanning a detection window across each frame at multiple positions and scales, in each position runs SVM classifier. It results in multiple overlapping detections in 3D position and scale space, around each player object in a frame and these are combined to get the final player position using Non-Maximum suppression with mean shift seeking algorithm [14]. But this HOG alone results in a human part with some background. Example results are shown in Fig. 2.

So, next is to segment the precise human part and eliminate background. Typically, segmentation is termed as a binary labeling problem where pixels are assigned to either foreground or background by the set of labels and can be optimally solved by an execution of min-cut/max-flow such as Graph-cuts which acts like a powerful energy minimization tool. The graph cut based segmentation approach of Boykov et al. [15] is adopted in this work with an energy function of the form:

$$E(A) = \lambda \cdot R(A) + B(A) \quad (3)$$



Fig. 2 Human detection using HOG

The coefficient $\lambda \geq 0$ specifies a relative importance of the region properties term $R(A)$ (penalties for assigning a pixel p to Foreground (F)) versus the boundary properties term background $B(A)$.

$$R(A) = \sum_{p \in P} R_p(A_p) \tag{4}$$

and

$$R_p(F) = -\ln \Pr(I_p|F) \tag{5}$$

$$R_p(B) = -\ln \Pr(I_p|B) \tag{6}$$

where negative log-likelihoods is motivated by the MAP-MRF formulations in [13]. In this work, the boundary penalties are set based on the following function Eq. 7:

$$B(A) = \sum_{p \in P} \sum_{\{p,q\} \in N} B_{\{p,q\}} \cdot \delta(A_p, A_q) \tag{7}$$

where

$$\delta(A_p, A_q) = \begin{cases} 1 & A_p \neq A_q \\ 0 & A_p = A_q \end{cases}$$

$$B_{\{p,q\}} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(p, q)}$$

where I_p and I_q are the intensities of pixel p and q with a penalty discontinuities between functions σ , when $|I_p - I_q| < \sigma$ the penalty is large, when $|I_p - I_q| > \sigma$ the penalty is small. Finally, the best cut that would give an “optimal” segmentation with minimum cost among all possible cuts as the sum of costs of all edges that go from S (source) to T (sink). This assigns each pixel either foreground i.e., player or background i.e., black.

$$C(S, T) = \sum_{(p,q) \in P, p \in S, q \in T} w(p, q) \tag{8}$$

When we used Graph-cut alone without HOG, this results in a foreground that comprises of player and non-player objects as well. This is due to variations in pixel intensities and camera movement (motion). We got the impressive results when HOG output fed to graph cut method. Figure 3 shows the results with Graph-Cut alone and Fig. 4 shows Graph-CUT with HOG.

On the extracted human body in the form of silhouette, we applied DCE method for skeletonization as a final step. The extracted skeleton has more accurate for further applications.



Fig. 3 Graph-cut alone



Fig. 4 Graph-cut with HOG

4 Implementation and Result

The proposed system uses a combination of HOG and Graph-cut for getting region based silhouette for players in the sports video sequence and then DCE is used for skeletonization for further processing in event detection and classification. We selected a cell size (8×8) and block size (16×16) for HOG computation. In this paper, the SVM is trained with sports players as positive class and non human as negative class on the HOG and S-T graph cut algorithm is used for segmenting the player from the background. Results of HOG and Graph-Cut for players detection is shown in Fig. 5a. DCE algorithm has been applied on the results of HOG and

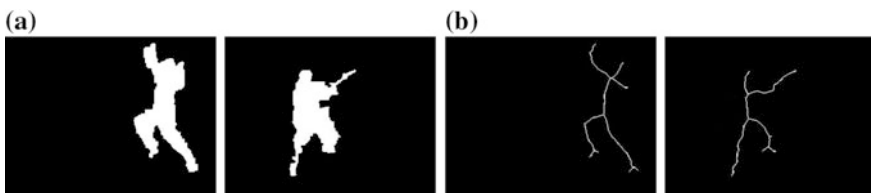


Fig. 5 From left to right: extracted silhouette using HOG-graph-CUT and extracted skeleton by DCE

Graph-cut methods and achieved the accurate skeleton as shown in Fig. 5b. This extracted skeleton further can be used for applications like event recognition, gait based human recognition, event classification and so on.

5 Conclusions and Future Work

We proposed a method to extract the skeleton of the player, which is based on a combination of Histogram of Oriented Gradients (HOG) features, Graph-cut method and DCE. This method precisely extracts human silhouette even in the varying backgrounds and then skeleton. These results also show that combination of HOG and Graph cut produces improved performance than applied individually. Silhouette features can adequately represent the movements performed by players in a video and DCE is applied on silhouette to extract skeleton for future work. Which includes, improving the accuracy of HOG-Graph-Cut further and then developing a system which extracts features from skeleton like joint features and train them to predict human action based on skeleton feature points or to classify the action or to detect the event. The actions could be bowling, batting, fielding etc.

References

1. J. Sullivan, S. Carlsson, Tracking and labelling of interacting multiple targets, in *Proceedings of the 9th European Conference on Computer Vision (ECCV 2006)* (2006)
2. P.J. Figueroa, N.J. Leite, Ricardo M.L. Barros, Tracking soccer players aiming their kinematical motion analysis. *Comput. Vis. Image Underst.* **101**(2), 122–135 (2006)
3. S. Gedikli, J. Bandouch, N. v. Hoyningen-Huene, B. Kirchlechner, M. Beetz, *An Adaptive Vision System for Tracking Soccer Players from Variable Camera Settings*
4. T.V. Spina, M. Tepper, A. Esler, V.M.N. Papanikolopoulos, A.X. Falcao, G. Sapiro, *Video Human Segmentation Using Fuzzy Object Models and Its Application to Body Pose Estimation of Toddlers for Behavior Studies*, 29 May 2013. arXiv:1305.6918v1 [cs.CV]
5. M. Grundmann, V. Kwatra, M. Han, I. Essa, Efficient hierarchical graph-based video segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, USA, June 2010
6. N. Vandenbroucke, L. Macaire, J.G. Postaire, Color pixel classification in an color hybrid color space, in *Proceedings of International Conference on Image Processing* (1998), pp. 176–180
7. W. Xie, R.P. Thompson, R. Perucchio, A topology-preserving parallel 3D thinning algorithm for extracting the curve skeleton. *Pattern Recogn.* **36**(7), 1529–1544 (2003)
8. X. Bai, L. Longin, L. Wenyu, Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3), 449–462 (2007)
9. N.D. Lakshmi, Y.M. Latha, A. Damodaram, Silhouette extraction of a human body based on fusion of HOG and graph-cut segmentation in dynamic backgrounds, in *Computational Intelligence and Information Technology, 2013. CIIT 2013. Third International Conference on*, 18–19 Oct 2013, pp. 527–531

10. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1 (2005), pp. 886–893
11. T. Joachims, Making large-scale SVM learning practical, in *Advances in Kernel Methods—Support Vector Learning*, ed. by B. Schölkopf, C. Burges, A. Smola (MIT-Press, 1999)
12. X. Lin, B. Cowan, A. Young, Model-based graph cut method for segmentation of the left ventricle, in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1–4 2005*
13. D. Greig, B. Porteous, A. Seheult, Exact maximum a posteriori estimation for binary images. *J. Roy. Stat. Soc. B* **51**(2), 271–279 (1989)
14. D. Comaniciu, Nonparametric information fusion for motion estimation, in *Proceedings of the Conference on CVPR 2003*, vol. I, Madison, Wisconsin, USA, 2003, pp. 59–66
15. Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. (PAMI)* **26**(9), 1124–1137 (2004)