Object Recognition, Tracking and Trajectory Generation in Real-Time Video Sequence

Hira Fatima, Syed Irtiza Ali Shah, Muqaddas Jamil, Farheen Mustafa, and Ismara Nadir

***Abstract*—The aim of tracking is to detect objects through images. This paper focuses on detection of moving objects in a scene and tracking of the objects as long as they remain in camera view. Initially noise removal and image enhancement is carried out. The objects are detected from the supervised Euclidean segmentation of the image frames. An object recognition algorithm then classifies the segmented objects by employing minimum distance classifiers approach, and by comparing various length and shape descriptors. Finally, object trajectory is generated based on centroid location of geometric object. For the case study, we have considered tracking of a tennis ball and shown its trajectory after tracking. This method assures accurate image segmentation via specifying color intensities of the object. Although median based approach gives much better results but it's rather much more expensive. Same case is with Census Transform method that is computationally expensive and is complex if multiple objects are there per frame.**

***Index Terms*—Object tracking, tennis ball, segmentation, pattern recognition, and trajectory.**

I. Introduction

Object detection and tracking in complex environment with multiple objects and their tracking is a significant objective in the area of computer vision, due to its extensive applications in many areas, such as traffic monitoring system, video surveillance, sports replays and robot navigation etc. The handiness of good quality and low price video cameras and the increasing requirement for automated video analysis has generated a good deal of attention in the areas of motion detection, object tracking and gesture analysis. Thus on a very high-level, it is possible to identify three steps in video analysis: detection of a moving object, tracking of the detected object in different frames, and analysis to recognize or characterize the object. This paper focuses mainly on the problem of lusty foreground segmentation of moving objects. The paper is ordered as follows: Section II discusses relevant research work and various techniques in the area of the object detection. Section III talks about the problem of tracking objects and our proposed solution. Section IV describes the results of the experiment, and finally conclusions and future work is presented in Section V.

Manuscript received June 14, 2013; revised September 4, 2013. This work was supported in part by National University of Sciences & Technology, Islamabad, Pakistan (NUST)/HEC.

The authors are with National University of Sciences & Technology NUST, Pakistan (e-mail: hf_262@yahoo.com, irtiza@smme.nust.edu.pk, muqaddas.jamil@seecs.edu.pk, farheen.mustafa@yahoo.com, n.ismara@hotmail.com).

II. Relevant Research Work

Despite much research work done on object recognition from a set of images, it is still an active area of research in computer vision. Content-based indexing through automatic targeted object detection and its recognition procedure has become one of most promising and intriguing problem for the years to come in, in order to face the termination of traditional informational systems. The objective of detection and its recognition is to find e.g. human faces, vehicles and buildings etc [1]. It has been applied successfully to adult content filtering of the web data, traffic controls, safety controls, geo-localization of images, videotele conferences, or sagacious man made machine intercourse. A survey on moving object detection and tracking in video surveillance is presented in [2]. Another paper [3] illustrates the detection and tracking of humans from a single video sequence using a pixel based motion detector to find region of interest (ROI) while using a model based approach. Paper [4] tracks a locomotive object by the Kalman Filter and its features. It detects all mobile objects in the frame. The method can also differentiate and track all locomotives individually and work in muss scenes satisfactorily. They detect all mobile objects, and for tracking they use the Kalman filter, color traits and position that is telling the distance of the targeted object from one frame to the next.

Object tracking is one of the most important aspect of sports video study. It has been prominently used to detect and track a player and a ball in a video. Common tracking method is based on Kalman filter usage for direct detection [5] and circular Hough transform based detection/tracking using neural classifier [6]. One more oncoming technique for tracking has foundations on trajectory report of the objects that are in motion within the frame [7]. Pose estimation and tracking a baseball in a video sequence is shown in [8], where upon utilizing the trajectory information, the ball position is judged and at the end ball is tracked efficiently along its trajectory. Another motion tracker [9] resolves typical tracking challenge, which is object's occlusions in the scene based on cluster information. A new approach has been presented in the context of pose determination of the object recognition and object class determination [10]; an algorithm has been tested on the simulated data as well as on the real data available. Overall experimental results have been extremely efficient with a recognition rate of about 97%, for 122 frames per second for about 60 objects.

Another research that has been carried out regarding object recognition and pose estimation deals with rigid body segmentation in a set of 2D images and estimates its 3D pose based on primitive knowledge of 3D model of the object [11]. The fundamental approach includes basic shape

DOI: 10.7763/IJIEE.2013.V3.395

and feature classification according to which modelling paradigms were predicted. Segmentation has been carried out by combining the edge based Geometric Active Contour (GAC) framework and region based technique that uses global image statistics. Since dynamic cluttered environments pose problems in the detection of objects of interest and their tracking, the possibility of occurrence of occlusions or missing information in image data set has been taken into account. Allili and Ziou [12] have also driven a model for tracking a video sequence entirely in the presence of non-static background environment.

III. Proposed Approach/Methodology

Our object tracking and detection algorithm is designed as a loop having three steps:

1) Detecting object based on supervised segmentation
2) Validating model and classification on the basis of shape based descriptors,
3) Tracking of the model in consequential frames.

Input to our algorithm is a sequence of camera images, where region of interest could either be chosen by manually selecting the colour of the object to be tracked, asking the user to select ROI interactively. Alternatively, the outcome of pixel based motion detector could be assigned to the ROI(background subtraction).

After a segmented object region is achieved, which is merely raw data with a boundary, we try to describe the region using shape based description and representation. Some of techniques used for this purpose are freeman chain codes [13], minimum perimeter polygons, shape numbers, boundary and texture descriptors. Pattern recognition is performed, that assigns the objects to their respective classes with as little human involvement as possible. This is usually carried out using decision theoretic methods such as minimum distance classifiers, correlation and adaptive learning. The minimum distance classifier approach has been used here to detect round shapes from the video sequences.

Based on the change of the object locations in successive frames, a trajectory is estimated and hence prediction of the ball locations for future frames is possible by modelling the ball trajectory using a cubic spline. This approach is widely used in cricket matches using cameras with high frame rates. The basic block diagram of our detection and tracking system is given below Fig. 1.

A. Object Segmentation

Human visual system gives very accurate evaluation for naturally existing colour image. Imitating the human way of analysing images, algorithms are designed to process a given image and to get better image quality than the original image. Processing is done on three colours R, G, and B of monochrome image separately and then combined to get a collective image. This procedure is sufficient to enhance the brightness of images [14], [15].

Noise introduction into digital images can be during transmission or acquisition. A decision based adaptive median filter is used to remove salt and pepper and other forms of noise. Restored images show significant improvement as compared with the originals. This procedure is much better for handling noise even at high levels [16].

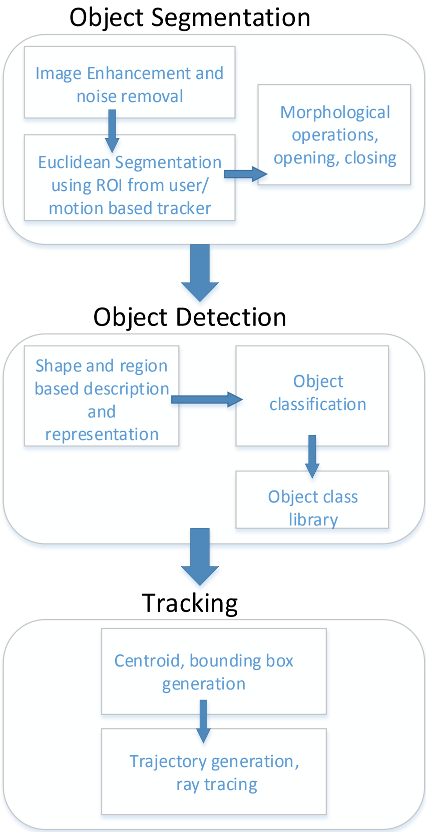

Fig. 1. Overall flowchart of the object detection and tracking method.

After image pre-processing is done, we proceed with image segmentation based on colour. A supervised human effort is required, which would be shifted to unsupervised segmentation by using pixel based motion detector. Division of image by segmentation into regions essentially divides the image into two segments; one is the concerned object or foreground pixels, and the other is background. By using series of simple operations, segmentation is done on every pixel of the image, based on Euclidean distances in colour. The Euclidean distance D_E measure is often used to figure out distance in N-dimensional vector space. It is calculated as:

$$D_E(\vec{v_1}, \vec{v_2}) = ||\vec{v_1} - \vec{v_2}||$$

where ||•|| is the L_2 vector norm. For a three-plane colour space the distance calculated is

$$D_E(\vec{v_1}, \vec{v_2}) = \sqrt{(v_{1,1} - v_{2,1})^2 + (v_{1,2} - v_{2,2})^2 + (v_{1,3} - v_{2,3})^2}$$

where v is a colour triplet, applying Euclidean based segmentation using the distance approach gives us a segmented object, as shown in Fig. 2. Additionally, morphological processing, which is a standard step in imaging, is applied on segmented images. Opening of images by using disk for ball tracking application smoothens it, removes thin protrusions and rounds the corners [17], [18].

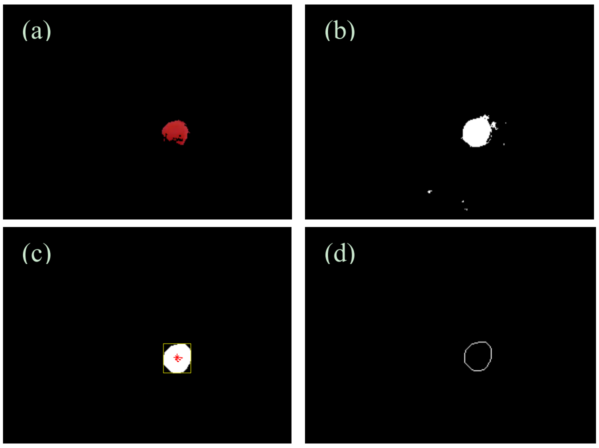

Fig. 2. Various results of the ball segmentation and morphological operations. (a) ROI (b) Euclidean segmented image (c) Opened Image showing ball shape, disk size 9, with centroid and bounding box (d) Boundary Length descriptor

B. Object Recognition

After we get the raw data from segmented images, which are merely a collection of black and white pixels, we try to represent and describe the object in terms of more understandable region and shape descriptors.

Freeman chain code is most common approach used to represent different shapes/contour i.e. planner, curves and contour of the object. Hence there is need to extract contour first that is done through segmentation. A fundamental problem occurring in geometric optimization is to compute a minimum-area or a minimum-perimeter. Let $n \in N$ is a positive integer and $h_{Q_{(n)}}$ is the support function of the $Q_{(n)}$. Where n is denoted by $P_{B,n}$ that is set of convex lattices of polygon having minimum perimeter than ‖ ·‖. $Q_{(n)}$ holding all the members of the polygon that polygon is having minimum perimeter. Its perimeter is represented as per $_{(n)}$. Since the perimeter is invariable under conditions of lattice translations, we will be supposing that $Q_{(n)}$ is positioned at the origin &

$$\lim_{n\to\infty} n^{\frac{-3}{2}} h_{Q_{(n)}}(u_1) = \frac{1}{4}\frac{\pi}{(6AB)^3}\int_{x\in B}|x.u|dx$$

where

A is area and B is its unit ball [19].

In most images, tennis ball appears in a circular or elliptical shape. Hough transform is an effective method that helps in detecting shapes such as points, lines and circles and behaves very well in partially occluded shapes [20].

Object detection in our case has been carried out based on the minimum distance classifiers approach [21]. Supposing that each pattern class is stated by an average vector $\mathbf{m}j$, which is the interpreter of that class of the vectors;

$$\mathbf{m}j = \frac{1}{N_j}\sum_{x\omega j}\mathbf{x} \quad j = 1,2,3 \ldots W$$

where N_j the number of training pattern vectors from class ω_j and recapitulation is taken over these vectors while W defining the number of pattern classes. $\boldsymbol{x}$ is the unknown pattern vector whose class membership is to determine. One method to determine the class of $\boldsymbol{x}$ (unknown pattern vector) is to allocate it to the closest prototype class. Using Euclidean distance of similarity as a measure of similarity reduces the issue to compute distance measures.

$$D_j(\boldsymbol{x}) = \left||\boldsymbol{x} - \boldsymbol{m}j|\right| \quad j = 1,2,3 \ldots W$$

Then $\boldsymbol{x}$ is allocated to class ω_i when $D_j(\boldsymbol{x})$ is the least distance.

C. Object Tracking

Object tracking has been done by highlighting the trajectory between centroids of the geometric shape (tennis ball in our case) between two consecutive frames. An arrow has been used to mark the centroid difference between two consecutive points and an overall path of the ball is achieved as shown in Fig. 3.

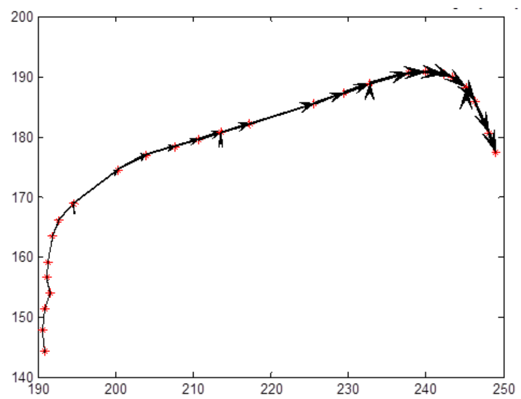

Fig. 3. Ball trajectory generated after centroids are labeled in each frame.

IV. Results

The videos used for experimentation were 30 frames per second, with a resolution of 320x240. They were taken in an indoor environment with a video camera. For the case of experimentation, an orange coloured tennis ball has been tracked. The algorithm was implemented in MATLAB and works offline as of now. Appreciable results were obtained by supervised object segmentation, and they became better after performing morphological operations i.e. opening/ closing. Object / pattern recognition correctly classifies the rounded object as ball in our case. The Fig. 3 is a trajectory of the tracked ball in the video sequence.

V. Conclusion

The proposed method guarantees satisfactory nice image segmentation by specifying colour intensities and tracks objects of interest. A minimum distance classifier approach is used for object classification. Tracking is achieved by specifying centroids of the object in each frame. Intensity specification guarantees that the object of interest is tracked in each image instead of the median based approach, which is more efficient for background estimation but is rather expensive while for census transform that computes for every pixel by comparing its grey value with the neighbouring grey value. As the size of window used here increases it ultimately increases the requirement of variable's size i.e. variables required to store census value would be 2^3 for 3×3 window, this results in to the increment of the computational complexity. Future work should concentrate on implementing segmentation which would work with occluded images with multiple objects and is more computationally efficient. Adaptive learning could also be integrated in object classification in future. 3D trajectory reconstruction is recommended for more knowledge about the trajectory of the ball/object tracked for more exhaustive analysis and visual perception.

ACKNOWLEDGMENT

This study has been funded in part by the National University of Sciences & Technology (NUST).

REFERENCES

[1] H. Yutaka, G. Christophe, and S. Rahul, *Industry and object recognition: Applications, applied research and challenges*, Berlin Heidelberg: Springer, vol. 4170, pp. 49–64, 2006.

[2] K. A. Joshi and D. G. Thakore, "A survey on moving object detection and tracking in video surveillance system," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, issue 3, July 2012.

[3] K. Filip and H. V ́aclav, "Detection and tracking of humans in single view sequences using 2D articulated model," *Series Computational imaging and vision (CIV)*, vol. 36, pp. 105-130, 2008.

[4] A. Salarpur, A. Salarpour, M. Fathi, and M. Dezf, "Vehicle tracking using kalman filter and features," *Journal of Signal & Image Processing*, vol. 2, 2011.

[5] Y. Seo, S. Choi, H. Kim, and K. Hong, "Where are the ball and players soccer game analysis with color based tracking and image mosaic", in *Proc. ICIAP*, pp. 196–203, Sep. 17–19, 1997.

[6] T. D. Orazio, G. Guarangnella, M. Leo, and A. Distante, "A new algorithm for ball recognition using circle Hough transform and neural classifier," *Pattern Recognition*, vol. 37, pp. 393-408, 2004.

[7] X. Yu, C. Xu, H. W. Leong, and Q. Tian, "Trajectory-based ball detection and tracking in broadcast soccer video," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1164-1178, 2006.

[8] B. Chakraborty and S. Meher, "2D trajectory-based position estimation and tracking of the ball in a basketball video," in *Proc. 2nd Int. Conf. on Trends in Optics and Photonics*, Calcutta, India, 2011.

[9] Tang Sze Ling *et al*, "Colour-based object tracking in surveillance application," in *Proc. Intl. Multi conf. of Engineers and Computer Scientists*, Hong Kong, 2009.

[10] L. Shang, "Real-time Object Recognition in Sparse Range Images Using Error Surface Embedding," Ph. D thesis, Electrical & Computer Engineering, Queen's University, Kingston, Canada, 2010.

[11] D. Samuel, R. Sandhu, A. Yezzi, and A. Tannenbaum, *Robust 3d pose estimation and efficient 2d region-based segmentation from a 3d shape prior*, Berlin Heidelberg: Springer, vol. 5303, pp. 169-182, 2008.

[12] M. S.Allili and D. Ziou, "Active contours for video object tracking using region, boundary and shape information," *Journal of Springer-verlag London Limited*, vol. 1, pp.101-117, 2007.

[13] Y. L. Liu and B. Zalik, "An efficient chain code with Huffman coding. Pattern recognition," *Journal of Pattern Recognition*, vol. 38, pp. 553-557, 2005.

[14] T. Lee and S. Soatto, "Video-based description for object recognition," *Journal of Image and vision computing*, vol. 29, pp. 639-652, 2011.

[15] H. Kaiqi, W. Zhenyang, and Qiao, "Image enhancement based on the statistics of visual representation," *Journal of Image and Vision Computing*, vol.23, pp. 51-57, 2005.

[16] S. Morillas, V. Gregori, G. Peris-Fajarnes, and A. Sapena, "Local self-adaptive fuzzy filter for impulsive noise removal in color images," *Journal of Signal Processing*, vol. 88, pp. 390-398, 2008.

[17] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing*, Bellingham, USA: SPIE, 2003, ch.7, pp. 163-191.

[18] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing*, Bellingham, USA: SPIE, 2003, ch. 9, pp. 227-265.

[19] S. B. M. Joseph, and P. Valentin, "Minimum-perimeter enclosures," in *journal of Information Processing Letters*, vol. 107, pp. 120-124, 2008.

[20] R. O. Duda and Hart, "Use of hough transforms to detect lines and curves in pictures," in *Proc. of Communications of ACM*, New York, 1972, vol. 15, pp. 11-15.

[21] R. Gonzalez, *Digital Image Processing Using Matlab*, 2nd ed. Natic, MA: Gatesmark, 2008, ch. 12, pp. 597-672.

Hira Fatima received the B.Sc. Chemical Engineering degrees from the University of Engineering and Technology, Lahore, in 2012 and pursuing M.Sc. Biomedical Engineering degree from National University of Science and Technology, Islamabad, Pakistan.

Her graduation work involved Dehydration of Natural Gas, in 2011 she joined Nestle Pakistan and work on focusing the Calibration and validations of the Instruments to gain the error free performance. Her theoretical and experimental research encompasses Lab-on-Chip Biotechnology for Molecular Imaging on Breast Cancer. Presently, she is serving as a researcher in molecular imaging electronics Laboratory.

Miss. Hira Fatima is a student member of American Institute of Chemical Engineers.

Syed Irtiza Ali Shah has earned his Phd in Aerospace/Mechanical Engineering from Georgia Institute of Technology, USA.

His theoretical and experimental research is based on the motion estimations for obstacles, image processing, aerial robotics and unmanned air vehicles. Presently, he is serving as Associate Professor at SMME National University of Science and Technology, Pakistan.

Muqaddas Jamil earned her bachelor's degree from College of E&ME, NUST, Pakistan in Electrical engineering with majors in DSP, Controls and Electronics in 2012. Currently she is pursuing her masters in Robotics and Intelligent Machine from the Robotics department at SMME, NUST with majors in Mobile Robotics and Advance Controls.

She is also working as hardware design engineer in Connekt lab, Seecs, NUST on a KACST funded project, titled "Reliable monitoring of oil and Gas pipelines using wireless sensor networks". Her interests include wireless sensor networks, controller design for energy harvesting systems, and development of data acquisition system for pipeline systems.

Farheen Mustafa has completed BS (Hons) Biotechnology and Bioinformatics from University of Veterinary and Animal Sciences, Lahore in 2011 and now doing MS in Biomedical Sciences from National University of Sciences and Technology, Islamabad, Pakistan.

In her graduation, she did 2 months internship from Government College University, Lahore where her work was on fermentation of different kinds of fungus. Her theoretical and experimental research work in MS is on Polymer particle for water treatment.

Ismara Nadir received the Bs.Biotechnology degree from Sarhad University of Science and Technology, Peshawar, in 2011 and pursuing M.S. Biomedical Sciences degree from National University of Science and Technology, Islamabad, Pakistan.

Her graduate work involved of Effect of heavy metal uptake on DNA, protein and proline content. She joined the Institute of Biotechnology and genetic engineering, agriculture University Peshwar, in 2010-2011 where her work was focused on the Laboratory techniques using tools of applied genetic engineering. Her theoretical and experimental investigations on Molecular techniques were based on experimental analyses using atomic absorption unit. She is presently serving as a researcher in the field of applied genetic engineering.