Object Detection and Tracking in Video Image

[1]Ashish Choure
[2]Amit Kumar Jha
[1]M-Tech Scholar, [2]Assistant Professor
[1]ashishchoure007@rediffmail.com, [2]akj1389@gmail.com
[12]Department of Computer Science Engineering, BUIT, Bhopal

Abstract- Object detection and tracking is a one of the challenging task in computer vision. Mainly there are three basic steps in video analysis: Detection of objects of interest from moving objects, Tracking of that interested objects in consecutive frames, and Analysis of object tracks to understand their behavior. Simple object detection compares a static background frame at the pixel level with the current frame of video. The existing method in this domain first tries to detect the interest object in video frames. One of the main difficulties in object tracking among many others is to choose suitable features and models for recognizing and tracking the interested object from a video. Some common choice to choose suitable feature to categories, visual objects are intensity, shape, color and feature points. In this thesis, we studied about mean shift tracking based on the color pdf, optical flow tracking based on the intensity and motion; SIFT tracking based on scale invariant local feature points. Preliminary results from experiments have shown that the adopted method is able to track targets with translation, rotation, partial occlusion and deformation.

Keywords: Object detection, Frame difference, Vision and scene understanding, Background subtraction, Scale invariant feature transform (SIFT)

1. INTRODUCTION

Object detection and tracking is an important challenging task within the area in Computer Vision that try to detect, recognize and track objects over a sequence of images called video. It helps to understand, describe object behavior instead of monitoring computer by human operators. It aims to locating moving objects in a video file or surveillance camera. Object tracking is the process of locating an object or multiple objects using a single camera, multiple cameras or given video file. Invention of high quality of the imaging sensor, quality of the image and resolution of the image are improved, and the exponential increment in computation power is required to be created of new good algorithm and its application using object tracking.

2. OBJECT DETECTION AND TRACKING PIPELINE

To overcome the different challenges issue as discussed in previous section there are following main component of object detection and tracking.

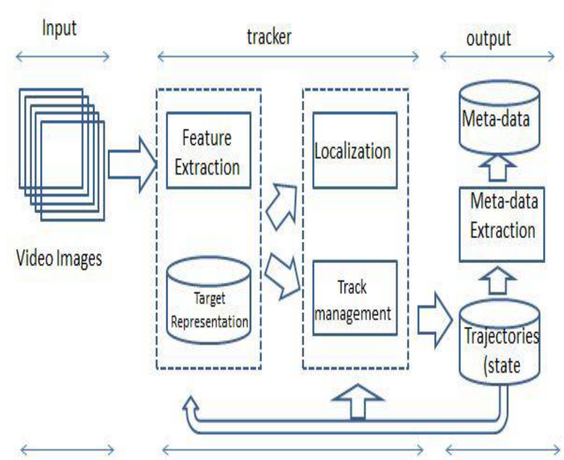

Figure 1 Basic component of object tracking algorithm

2.1 Feature extraction Method

Any object tracking algorithm can be analyzed by the quality of information that can be extracted from video frames or an image. To get more exploit information from image, we use image formation technique to extract feature which are more important, significant to identify interested object uniquely without any disambiguation.

- From the image background in the scene and
- From many another objects which are present in the scene

For any tracking algorithm extracting feature is the important step which is allowing us to highlight the information of the interested object from the video frames or target image plane. Extracted feature can be of three types.

- Low level extraction, e.g., motion, color, gradient
- Mid level extraction, e.g., edge, corner, interest point, region
- High level extraction, e.g., centroid, area, orientation, whole object

2.2 Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) is a methodology for identifying and concentrating local feature descriptors that are sensibly invariant to changes in enlightenment, scaling, pivot, image noise and little changes in perspective. This calculation is initially proposed by David Lowe in 1999, and afterward further created and moved forward.

SIFT characteristics have numerous preferences, for examples are follows:

- SIFT Features are natural feature of pictures. They are positively invariant to picture interpretation, scaling, revolution, brightening, perspective, commotion and so on.
- Great strength, rich in data, suitable for quick and precise matching in a mass of feature database.
- Richness. Heaps of SIFT feature will be investigated regardless of the possibility that there are just a couple of object.
- Moderately quick speed. The pace of SIFT even can fulfill ongoing process after the SIFT algorithm is advanced.
- Better expansibility. SIFT is extremely helpful to consolidate with other eigenvector, and create much valuable information.

2.3 Scale-space extrema detection

The primary phase of calculation inquiries over all scales and image areas. It is actualized productively by method for a difference of- Gaussian capacity to recognize potential investment point that are invariant to scale and orientation.

Interest point for SIFT characteristics relate to neighborhood extrema of difference of- Gaussian channels at diverse scales. Given a Gaussian-blurred image described

as the formula L(x, y, σ) = G(x, y, σ) * I(x, y)

Where

$$G(x,y,\sigma) = \frac{1}{2\Pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}$$

Is a variable scale Gaussian, whose result of convolving an image with a difference-of-Gaussian filter is given by

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Which is just be different from the Gaussian-blurred images at scales σ and kσ

Once this Dog are discovered, image are hunt down neighborhood extrema over scale and space. For, e.g., one pixel in a picture is contrasted and its 8 neighbors and also 9 pixels in next scale and 9 pixels in past scales. In the event that it is a nearby extrema, it is a potential keypoint. It essentially implies that keypoint is best spoken to in that scale Interest points (called keypoints in the SIFT framework) are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 8 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoint.

2.4 Kanade - Lucas - Tomasi feature tracker

Kanade-Lucas-Tomasi(KLT) feature tracker is a methodology of characteristic extraction in the area of computer vision. It is proposed basically with the end goal of managing the issue that customary image enrollment strategies are by and large immoderate. KLT makes utilization of spatial intensity information to steer the quest for the position that yields the best match. It is speedier than customary strategies for analyzing far fewer potential matches between the images.

To track the object over time, we also uses the Kanade-Lucas-Tomasi (KLT) algorithm. While it is possible to use the cascade object detector on every frame, it is computationally expensive. It may also fail to detect the object, when the object turns or tilts. This limitation comes from the type of trained classification model used for detection. The example detects the object only once, and then the KLT algorithm tracks the object across the video frames.

2.5 Mean Shift

Accurate visual item following under the state of low computational multifaceted nature displays a test. Ongoing undertaking, for example, reconnaissance and observing [15], perceptual client interfaces [16], keen rooms [17, 18], and feature layering [19] all require the capability to track moving object. As a rule, following of visual items is possible either by forward - tracking or by backward - tracking. The forward - tracking methodology appraises the positions of the areas in the current casing utilizing the division result got for the past picture. The backtracking based methodology sections frontal area areas in the current images and after that makes the correspondence of regions between the past images. For securing correspondence, a few object layouts are used. A conceivable forward-following method is mean-shift dissection. Mean movement method was initially presented in 1975, yet just following 20 years after the fact in 1995; this system has been re-presented by D. Fuiorea [20]. In his article, a portion capacity is characterized to figure the separation between specimen focuses and its mean movement, likewise a weight coefficient is reverse with the separation. The closer the separation is, the bigger the weight coefficient is. The mean movement calculation is anon-parametric method [21]. It gives faultless restriction and proficient matching without exorbitant exhaustive hunt. It is an iterative process, that is to say, first register the mean movement esteem for the current point position, then move the point to its mean movement esteem as the new position, then figure the mean movement until it satisfy certain condition. The guideline of mean movement strategy might be picked up from Figure.

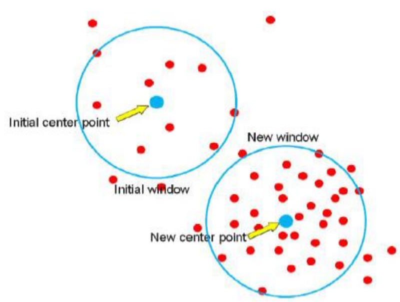

Figure 2 The principle of mean shift procedure

3. EXPERIMENTAL RESULT

Several experiments had been done to evaluate the tracking algorithms. These sequences used in experiments consist of indoors and outdoors testing environments so that the proposed scheme can be fully evaluated. First, target object of interest is defined from the first some frames. Then SIFT features are obtained from the target object. The features are stored by their key-points descriptors. Then, SIFT features are obtained from the consecutive frames to match the feature from interested object. The features of frames are also stored by other keypoints descriptors. The target object is tracked in the next frame by individually comparing each feature point found from the next frame to those on the target object. The Euclidean distance is worked out. The candidate can be preserved when the two features Euclidean distance is larger than a threshold. So the good matches are picked out by the consistency of their location, orientation and scale.

They represent the outcomes of the SIFT and tracker.

Figure 3 Detected target object in frame

A trajectory is the path that a moving object follows through space as a function of time. A trajectory can be described mathematical either by the geometry of the path or as the position of the object over time. It will store the actual path of object of interest i.e. information of target in consecutive frames. We will get the all information about target object that in which direction it moves and what is the speed of target. Trajectory of our given object of interest is give below:

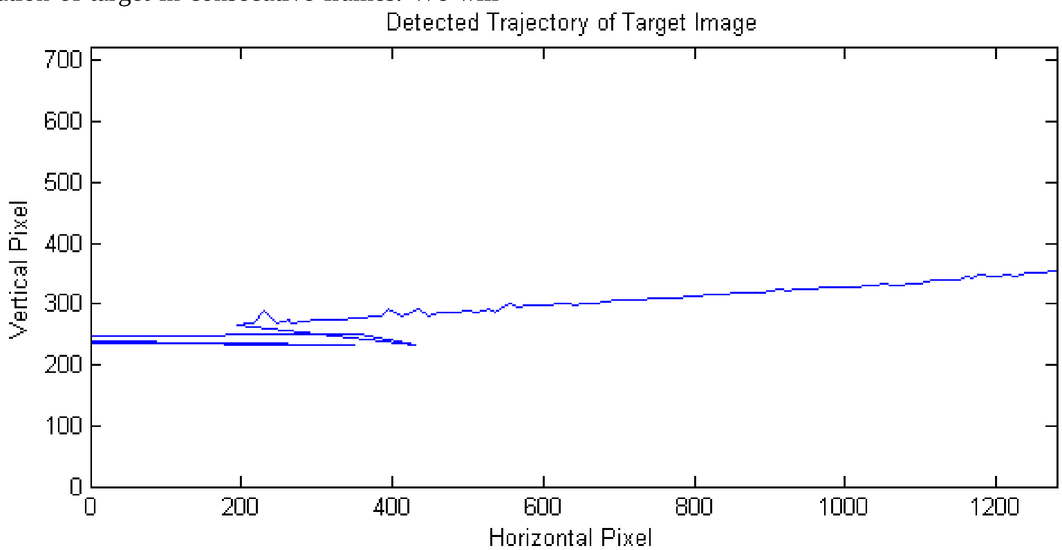

Figure 4 Trajectory of target object in video

4. CONCLUSION

Object detection and tracking is an important task in computer vision field. In object detection and tracking it consist of two major processes, object detection and object tracking. Object detection in video image obtained from single camera with static background that means fixing camera is achieved by background subtraction approach. In this thesis, we tried different videos with fixed camera with a single object and multiple objects to see it is able to detect objects. Motion based systems for detecting and tracking given moving object of interest can be created. Using SIFT feature extraction first feature of the object and the frame has detected to match the interested object. Since for feature extraction, SIFT algorithm has been used so tracker is invariant to representation of interested object.

5. FUTURE WORK

In the future, we can extend the work to detect the moving object with non-static background, having multiple cameras which can be used in real time surveillance applications.

REFERENCES

[1] Yiwei Wang, John F. Doherty and Robert E. Van Dyck, "Moving Object Tracking in Video", in proceedings of 29th applied imagery pattern recognition workshop, ISBN 0-7695-0978-9, page 95,2000.

[2] Bhavana C. Bendale, Prof. Anil R. Karwankar, "Moving Object Tracking in Video Using MATLAB", International Journal of Electronics, Communication and Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 1.

[3] Marcus A. Brubaker, Leonid Sigal and David J. Fleet, "Video-Based People Tracking", hand book of ambient intelligence under smart environments 2010, pp 57-87.

[4] Emilio Maggio and Andrea Cavallaro, "Video Tracking: Theory and Practice", first edition 2011, John Wiley and Sons, Ltd.

[5] Y.Alper, J.Omar, and S.Mubarak. "Object Tracking: A Survey" ACM Computing Surveys,vol. 38, no. 4, Article 13, December 2006.

[6] B. Triggs, P.F. McLauchlan, R.I. Hartley and A.W. "Fitzgibbon. Bundle adjustment – a modern synthesis". In Proceedings of the International Conference on Computer Vision, London, UK, 1999, 298?372.

[7] G.C. Holst and T.S. Lomheim. "CMOS/CCD Sensors and Camera Systems". Bellingham, WA, SPIE Society of Photo-Optical Instrumentation Engineering, 2007.
[8] E. Maggio, M. Taj and A. Cavallaro. "E_cient multi-target visual tracking using random finite sets". IEEE Transactions on Circuits Systems and Video Technology, 18(8), 1016?1027, 2008.
[9] G. David Lowe. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision. 2. pp. 1150?1157,1997.
[10] G. David Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), pp, 91-110, 2004.
[11] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. International Joint Conference on Artificial Intelligence, pages 674-679, 1981.
[12] Carlo Tomasi and Takeo Kanade. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.
[13] Jianbo Shi and Carlo Tomasi. Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, pages 593-600, 1994.
[14] Stan Birchfield. Derivation of Kanade-Lucas-Tomasi Tracking Equation. Unpublished, January 1997.
[15] Y. Cui, S. Samarasekera, Q. Huang. Indoor Monitoring Via the Collaboration Betweena Peripheral Senson and a Foveal Sensor, IEEE Work-shop on Visual Surveillance, Bomba y, India, 2-9, 1998.
[16] G. R. Bradski, Computer Vision Face T racking as a Component of a Perceptual User Interface, IEEE Work. on Applic. Comp. Vis., Princeton, 214-219, 1998.
[17] S.S. Intille, J.W. Davis, A.F. Bobick, Real-Time Closed-World T racking. IEEE Conf. on Comp. Vis. and Pat. Rec., Puerto Rico, 697-703, 1997.
[18] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, P_nder: Real-Time Tracking of the Human Body, IEEE Trans. Pattern Analysis Machine Intell, 19:780-785, 1997.
[19] A. Eleftheriadis, A. Jacquin. Automatic Face Location Detection and Tracking for Model-Assisted Coding of Video Teleconference Sequences at Low Bit Rates, Signal Processing- Image Communication, 7(3): 231-248, 1995.
[20] D. Fuiorea, V. Gui, D. Pescaru, and C. Toma. Comparative study on RANSAC and Mean shift algorithm, International Symposium on Electronics and Telecommunications Edition 8. vol. 53(67) Sept. 2008, pp. 80-85.
[21] Y.Cheng. Mean Shift, Mode Seeking, and Clustering, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, No 8, 790-799,1995.
[22] Stern H, Efros B (2005) Adaptive color space switching for tracking under varying illumination. Image Vis Comput 23(3):353364. doi : 10.1016 /j. imavis 2004.09.005
[23] Li S-X, Chang H-X, Zhu C-F (2010) Adaptive pyramid mean shift for global real-time visual tracking. Image Vis Comput 28(3):424437. doi : 10.1016 / j. imavis 2009.06.012
[24] Yuan G-W, Gao Y, Xu D (2011) A moving objects tracking method based on a combination of local binary pattern texture and Hue. Procedia Eng 15:39643968. doi:10.1016/j. proeng 2011.08.742
[25] Mazinan AH, Amir-Lati_ A (2012) Applying mean shift, motion information and Kalman filtering approaches to object tracking. ISA Trans 51(3):485497. doi: 10.1016/j. isatra 2012.02.002
[26] Lai S-H (2004) Computation of optical ow under non-uniform brightness variations. Pattern Recognit Lett 25(8):885892. doi : 10.1016 / j. patrec 2004.02.001
[27] Alan J Lipton, Hironobu Fujiyoshi, and Raju S Patil. Moving target classification and tracking from real-time video. In Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on, pages 814. IEEE, 1998.
[28] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., volume 2. IEEE, 1999.
[29] Ya Liu, Haizhou Ai, and Guang-you Xu. Moving object detection and tracking based on background subtraction. In Multispectral Image Processing and Pattern Recognition, pages 6266. International Society for Optics and Photonics, 2001.