Real Time Object Detection and Tracking

Akhila S[1]

[1]*DDMCA Student, Dept. of MCA, Sree Narayana Guru Institute of Science and Technology, Kerala, India*

---***---

Abstract - *Obstacle detection and tracking could also be an important research topic in computer vision with kind of practical applications. Though an ample amount of research has been exhausted this domain, implementing automatic obstacle detection and tracking in real-time remains an enormous challenge. To affect this issue, we propose a fast and robust obstacle detection and tracking approach by integrating an adaptive obstacle detection strategy within a Kalman filter framework during this paper. An appropriate salient object detection method autoinitializes the Kalman tracker for this purpose. Moreover, an adaptive obstacle detection strategy is proposed to refine things and boundary of the thing when the arrogance value of the tracker drops below a predefined threshold. Additionally, a reliable post-processing technique is implemented to accurately localize the obstacle from a saliency map recovered from the search region. The proposed approach has been extensively tested through quantitative and qualitative evaluations on sort of challenging datasets. The experiments demonstrate that the proposed approach significantly outperforms the state-of-the-art methods in terms of tracking speed and accuracy.*

Key Words: Object tracking, detection, Kalman filter, salient object.

1. INTRODUCTION

Automating visual detection and tracking of moving objects by intelligent autonomous systems, like unmanned aerial vehicles (UAVs), has been a lively research topic for the past decades in computer vision. The research has diverse applications extending from military, surveillance, security systems, aerial photography, search and rescue, visual perception, auto-navigation to human-machine interactions [1]. Recently, computer vision is being extensively utilized in roadside vehicle positioning and tracking also as in intelligent transportation systems for the vehicle also because the passengers' safety [2], [3]. Thanks to its emerging multidisciplinary usage, a handsome number of companies are developing their own UAV systems, like Google's Project Wing, Amazon Prime Air and DHL's parcelcopter. However, designing such intelligent UAVs is pragmatically challenging. It's vital to stay track of other UAVs, birds, airplanes or other possible flying objects during the flight of an autonomous UAV. Hence, identifying such potential obstacles precisely and localizing them in real-time for successful collision avoidance and autonomous navigation is important. Recognizing such possible threats during a real-time and embedding such computationally sound algorithm in flying UAVs demands an ample amount of research and engineering.

In this paper, we propose a completely unique and intelligent vision-based system which will automatically detect, localize, and track the objects in high speed. Extensive experiments demonstrate that the proposed approach stands out among all the state-of-the-art detectors and trackers in terms of speed and precision.

Mathematically, a saliency map are often understood as a probability map that expresses the probability of salient pixels in terms of intensity relative to the whole image [5], [6]. [7] claimed that a serious portion of the image is occupied by the image background and is homogeneous. Hence, the image boundary are often easily connected - connectivity prior. They also assumed that objects are generally absent on the image boundaries in order that we will presume these boundaries to be the background also i.e. background prior. Zhang et al. [8] successfully used minimum barrier distance [9], [10] along side a raster scanning algorithm utilizing both connectivity prior and background before generate saliency map in their work.

In this paper, we propose quick, reliable and accurate object localization and tracking approach for the autonomous navigation of the flying UAVs by integrating the techniques for salient object detection [8] with the kernelized correlation filter [4]. Our approach achieves better detection and tracking results compared to the state-of-the-art method in terms of speed and accuracy.

1.1 Machine Learning

Machine learning is an application of AI (AI) that gives systems the power to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the event of computer programs which will access data and use it learn for themselves. The method of learning begins with observations or data, like examples, direct experience, or instruction, so as to seem for patterns in data and make better decisions within the future supported the examples that we offer. The first aim is to permit the computers learn automatically without human intervention or assistance and adjust actions accordingly. But, using the classic algorithms of machine learning, text is taken into account as a sequence of keywords; instead, an approach supported semantic analysis mimics the human ability to know the meaning of a text.

1.2 Computer Vision

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the attitude of engineering, it seeks to understand and automate tasks that the human sensory system can do. Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the important world so on supply numerical or symbolic information. Understanding means the transformation of visual images into descriptions of the earth that add up to thought processes and should elicit appropriate action. This image understanding are often seen because the disentangling of symbolic information from image data using models constructed with the help of geometry, physics, statistics, and learning theory. The pc vision system methods are given in Fig-1.

2. RELATED WORKS

An object tracking by detection approach was proposed in [11]. However, since their detector needs training with a large number of data samples, auto-initialization is not feasible in their approach. Optical flow motion cues were leveraged in [12] to design a tracker combined with a detection scheme based on a saliency map with an auto-initialization in the first frame. However, such trackers are computationally too expensive for real-time applications. Multiple cameras in an aircraft that measure the altitude and act as a sense and avoid collision sensor were used in [13] but their method is inefficient without the use of GPS data (which may have delays), thus not feasible for real-time. Previous research on saliency map-based object detection can be broadly classified into two methods — top-down and bottom-up. In the top-down methods [5], [6], [14], detection is executed on the reduced search space since all the possible objects in an image are localized. But these methods are unrealistic for real-time object detection because they are mostly task-driven and accompanied by supervised learning. On the other hand, bottom-up methods [7], [8], [15]–[17] compare the feature contrast of the salient region with the background contrast by using the low-level features (like the color, contrast, shape, texture, gradient and spatio-temporal features) from an image. Such methods have higher possibilities to fail in the case of complex images as they do not have prior knowledge of the localization of the object or the number of objects present in an image. In contrast, the top-down methods require proper training before detection. However, our approach identifies the approximate location of the object from the previous tracking results and then performs the re-detection on a much smaller search region. Hence, our method is computationally efficient since it does not require any type of supervised training for the detection. Additionally, the reduced search region enhances qualitative efficiency during detection.

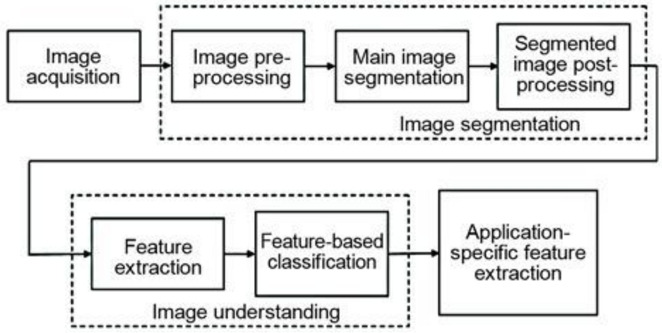

Fig -1: Computer Vision System Methods

3. METHODOLOGY

In this section, we describe the details of the proposed strategy for fast and robust object detection and tracking. First, a saliency map S of an entire image is generated to segment the salient object out from the background and autoinitialize the tracker with the current location of the salient object for tracking in the consecutive frame. In this process, we generate a saliency map, post-process the generated saliency map using the proposed post-processing technique to segment the salient object and feed the location of the salient object to initialize the tracker. Next, the filter starts training itself on the salient object on each frame while tracking of the object runs simultaneously until a low peak of filter response (confidence value) is observed. Confidence value measures the resemblance of the object in the consecutive frame compared to the previous frame where the object was being tracked. Once such low confidence value is observed for the tracker, our proposed adaptive detection approach is applied to re-detect the object. The re-detection scheme is important because it helps to increase the confidence value of the tracker to track in the later frames by re-estimating the accurate position and the size of the object being tracked. To do so, we determine an adaptive search region R based on the confidence value; the area of R is progressively increased to re-detect the object being tracked as the confidence value drops lower. This redetection scheme is much similar to the detection process as performed in the first frame. A slight variation is that we generate S only for search region R instead of an entire frame and update the tracker accordingly for a smooth training of the KCF filter throughout the tracking process.

- SALIENCY MAP

  In computer vision, a saliency map is an image that shows each pixel's unique quality.[1] The goal of a saliency map is to simplify and/or change the representation of an image into something that's more meaningful and easier to research. Saliency estimation could also be viewed as an instance of image segmentation. In computer vision, image segmentation is that the process of partitioning a digital image into multiple segments (sets of pixels, also referred to as superpixels). The goal of segmentation is to simplify and/or change the representation of a picture into something that's more meaningful and easier to research. Image segmentation is usually wont to locate objects and limits (lines, curves, etc.) in images. More precisely, image segmentation is that the process of assigning a label to each pixel in a picture such pixels with an equivalent label share certain characteristics. Saliency maps process images to differentiate visual features in images. For example, coloured images are converted to black-and-white images in order to analyse the strongest colours present in them. The Saliency Map may be a topographically arranged map that represents visual saliency of a corresponding visual scene.

- KALMAN FILTER

Kalman filtering is an algorithm that gives estimates of some unknown variables given the measurements observed over time. Kalman filters are demonstrating its usefulness in various applications. Kalman filters have relatively simple form and need small computational power. However, it is still not easy for people who are not familiar with estimation theory to understand and implement the Kalman filters.

- DATASET

Our approach is tested on 25 challenging video sequences where the object is subjected to variations of scale, partial occlusion, axial and planar rotation, illumination variation and camera instability.

4. CONCLUSION

It is vital for an intelligent autonomous UAV to have an automatic, robust and real-time object tracking system built in it. Therefore, in this paper, we have proposed a tracking method that incorporates variations in shape, size, illumination as well as degenerate conditions like partial occlusion, planar/axial rotation and camera instability in it for better performance than the existing state-of-the-art trackers. Most of the up-to-date trackers were found to fail in one or several such complex scenarios. However, our tracker is able to keep track of the object without any abrupt failures. Both qualitative and quantitative evaluation measures demonstrate that the proposed approach is more efficient than the competing trackers. Unlike other trackers, the proposed tracker is able to auto-initialize without any manual interference. Hence, our approach is found to be accurate and fast in terms of speed for realtime autonomous sense-and-avoid UAVs, drones or similar flying units. Nevertheless, some of the experiments show that our method may not perform as expected in the presence of several dubious salient objects in a scene or the object is too tiny to detect and auto-initialize the tracker.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” Acm computing surveys (CSUR), vol. 38, no. 4, p. 13, 2006.

[2] Z. Kim, “Robust lane detection and tracking in challenging scenarios,” IEEE Transactions on Intelligent Transportation Systems, vol. 9, no. 1, pp. 16–26, March 2008.

[3] M. Shan, S. Worrall, and E. Nebot, “Probabilistic long-term vehicle motion prediction and tracking in large environments,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, pp. 539–552, June 2013.

[4] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583–596, 2015.

[5] H. Cholakkal, D. Rajan, and J. Johnson, “Top-down saliency with locality-constrained contextual sparse coding.” BMVC, 2015.

[6] J. Yang and M.-H. Yang, “Top-down visual saliency via joint crf and dictionary learning,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 2296–2303.

[7] Y. Wei, F. Wen, W. Zhu, and J. Sun, “Geodesic saliency using background priors,” in European Conference on Computer Vision. Springer, 2012, pp. 29–42.

[8] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, “Minimum barrier salient object detection at 80 fps,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1404–1412.

[9] K. C. Ciesielski, R. Strand, F. Malmberg, and P. K. Saha, “Efficient algorithm for finding the exact minimum barrier distance,” Computer Vision and Image Understanding, vol. 123, pp. 53–64, 2014.

[10] R. Strand, K. C. Ciesielski, F. Malmberg, and P. K. Saha, “The minimum barrier distance,” Computer Vision and Image Understanding, vol. 117, no. 4, pp. 429–437, 2013.

[11] M. Andriluka, S. Roth, and B. Schiele, “People-tracking-by-detection and people-detection-by-tracking,” in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008, pp. 1–8.

[12] V. Mahadevan and N. Vasconcelos, “Saliency-based discriminant tracking,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 1007–1013.

[13] A. Nussberger, H. Grabner, and L. Van Gool, “Robust aerial object tracking in high dynamic flight maneuvers,” ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 2, no. 1, p. 1, 2015.

[14] A. Kocak, K. Cizmeciler, A. Erdem, and E. Erdem, “Top down saliency estimation via superpixel-based discriminative dictionaries.” in BMVC, 2014.

[15] S. He, R. W. Lau, W. Liu, Z. Huang, and Q. Yang, “Supercnn: A superpixelwise convolutional neural

network for salient object detection," International Journal of Computer Vision, vol. 115, no. 3, pp. 330–344, 2015.

[16] P. Wang, J. Wang, G. Zeng, J. Feng, H. Zha, and S. Li, "Salient object detection for searched web images via global saliency," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3194–3201.

[17] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1265–1274.