

Vehicle Detection on Aerial Image

Vanniphogalu Shruti Parasayya
Electronics Engineering
SAKEC
Mumbai, India

Abstract— Detecting vehicles in aerial images provides important information for traffic management and urban planning. Detecting cars in aerial images is challenging due to the relatively small size of the target objects and the complex background in man-made areas. It is particularly challenging if the goal is near real-time detection within few seconds on large images without any additional information, e.g. road database, accurate target size. We proposed a system which can detect the vehicles on aerial video. Beside the bounding box of the vehicles we also extract an orientation and type (car/truck) information. First we classify vehicles as car or truck depending on the size of the bounding box obtained in first step. In the next step we also estimate the orientation (Right or Left) of the vehicle by comparing shift in the centroid of the vehicle. We evaluate our method on a dataset of original aerial video from a UAV (Unmanned Aerial Vehicle).

Keywords—Component; Formatting; Style; Styling; Insert

I. INTRODUCTION

The detection of vehicles in aerial images is important for various applications e.g. traffic management, parking lot utilization, urban planning, etc. Collecting traffic and parking data from an airborne platform gives fast coverage over a larger area. Getting the same coverage by terrestrial sensors would need the deployment of more sensors, more manual work, thus higher costs.

We proposed a system in which we capture aerial images over road and vehicles are detected across multiple consecutive frames. This gives fast and comprehensive information of the traffic situation by providing the number of vehicles and their positions and speed. The detection is a challenging problem due to the small size of the vehicles (car might be only 30x12 pixels) and the complex background of manmade objects which appear visually similar to cars. Providing both the type and the orientation of the detected objects supports tracking by giving constraints on the motion of the vehicles. This is particularly important in dense traffic scenes where, object assignment is more challenging.

In our system, we detect the bounding box of the vehicles by a Gaussian Mixture model and classifier in cascade structure. The bounding boxes are further classified to different orientations and vehicle type. This system is referred from paper published by Kang Liu and Gellert in the year 2015 titled as “Fast multiclass vehicle detection on aerial images” [1].

II. LITERATURE SURVEY

Detecting the cars in the images is challenging due to the relatively small size of the target objects and the complex background in man-made areas. It is particularly

challenging if the goal is near real-time detection within few seconds on large images. In this chapter will discuss previous works which are done on aerial images for traffic management and urban planning.

In the paper published by authors Vatau.A, Danescu.R titled as “Stereovision based approach for tracking multiple objects in crowded environments” where, typically the road lane markings are not visible and the surrounding infrastructure is not known [2]. This system relies on measurement data provided by an intermediate occupancy grid derived from processing stereovision based elevation map and on free-form object delimiters extracted from this grid. This system needs some practical sensors for filtering, which adds cost to the system.

The system introduced by authors Jiann-Yeou Rau; Jyun-Ping Jhan; Ya-Ching Hsu titled as “Analysis of Oblique Aerial Images for Land Cover and Point Cloud Classification in an Urban Environment” Automatically classify 3D (three dimensional) point clouds generated using oblique aerial images and vertical aerial images into various urban object classes, such as roof, facade, road, tree and grass. The generation of 3D point clouds is difficult as it requires two types of image oblique aerial images and vertical aerial images.

The system introduced by Moranduzzo and Melagni titled as “Automatic car counting method for unmanned aerial vehicle images” [4], processes very high resolution images for car detection. It uses a feature point detector and SVM (support vector machine) classification of SIFT (Scale-invariant feature transform) descriptors is applied.

The authors D. Rosenbaum, R.E. Schapire introduces a method titled as “ Detecting cars in UAV images with a catalog-based approach” uses a catalog of HOG (histogram of gradient) descriptors and later an orientation estimation[5].

In the paper published by X. Chen, S. Xiang, C. Liu, and C. Pan, titled as “Vehicle detection in satellite images by hybrid deep convolutional neural networks” cars are detected by a deep neural network in a sliding window approach on a known constant scale [6].

In the system proposed by Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, titled as “A 3D teacher for car detection in aerial images”. Vehicles are detected with online boosting on Haar-like features, local binary patterns and orientation histograms. They train the detector for cars in one direction and during testing they rotate the image in 15 degrees step. This detector is trained for a known object size 35 × 70 pixels and tested on images with the same scale[7].

In the system proposed by J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, titled as “An operational system for estimating road traffic information from aerial images” utilize the road map and stereo matching to limit the search area to roads and exclude buildings [9]. HOG features with an AdaBoost classifier are applied to detect the cars on the selected region. This method is limited to georeferenced image pairs and areas covered by the road database.

In the paper published by Jaesik Choi Titled as “Real-time On-Road Vehicle Detection with Optical Flows and Haar-like feature detector” they have used optical flow to detect coming traffics [10]. The ‘optical flow’ assumes that important features are detected in both the frames. The optical flow algorithm is not so appropriate for these traffics because they do not have any salient movement in most cases.

The authors P. Viola, M. Jones introduced a Haar-like feature detector method titled as “Detecting pedestrians using patterns of motion and appearance”, to detect traffic in same direction [11].

In the paper published by G.D.Hager and P.N.Belhumeur titled as “Efficient region tracking with parametric models of geometry and illumination” used horizontal and vertical edges(Knowledge-based methods) in HG step[12]. The selected regions at HG step are matched with predefined templates in HV step. Whereas, authors J. Wang, X. Chen, and W .Gao. Published a paper titled as “Online selecting discriminative tracking features using particle filter” used horizontal and vertical edge in HG step [13]. However; they use Haar Wavelet Transform and SVMs (Appearance based methods) in HV step. In the paper published by authors P.Viola and M. Jones titled as “Robust real-time object detection”, detected long distance stationary obstacles including vehicles [14].

In the paper published by Y. Freund and R.E. Schapire, titled as, “A decision theoretic generalization of online learning and an application to boosting”[15]. They used an efficient optical flow algorithm (Motion based methods) in HG step [15]. Although they did not use any traditional HV method, they used Sum of squared differences (SSD) with a threshold value to verify their hypothesis.

For implementing the system we proposed, we are using the technique published in paper by authors D. Hari Hara Santosh, and P. Venkatesh, P. Poornesh titled as “Tracking Multiple Moving Objects Using Gaussian Mixture Model” [16].

III. PROPOSED METHODOLOGY

A. Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as colour based tracking of an object in video. In many computer related vision technology, it is critical to identify moving objects from a sequence of videos frames. In order to achieve this, background subtraction is applied which mainly identifies moving objects from each portion of video frames. Background subtraction or segmentation

is a widely used technique in video surveillance, target recognitions and banks. By using the Gaussian Mixture Model background model, frame pixels are deleted from the required video to achieve the desired results. The application of background subtraction involves various factors which involve developing an algorithm which is able to detect the required object robustly, it should also be able to react to various changes like illumination, starting and stopping of moving objects. Surveillance is the monitoring of the behaviour, activities or other changing information usually of people and often in a surreptitious manner. Video surveillance is commonly used for event detection and human identification. But it is not easy as think to detect the event or tracking the object. There are many techniques and papers introduced by many scientists for the backend process in the video surveillance. Different automated software’s are used for the analysis of the video footage. It tracks large body movements and objects.

B. Vehicle Classification based on Type

Our next step is to classify the vehicle as car or truck, also to estimate the orientation of the vehicle. To classify it as Car or truck we use the concept of size. We first set the threshold and compare it with bounding box of the vehicle. If the Bounding box is less than threshold it will undergo category A which is car, whereas the Size of Bounding box greater than threshold will undergo Category B which is Truck. The flowchart of this algorithm is shown in Fig 3.1

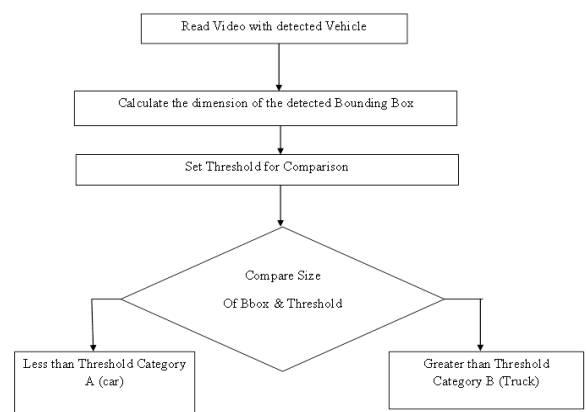


Fig 3.1 Flowchart of Vehicle Classification based on type

C. Estimation of Orientation of Vehicle

w our next target is to achieve the orientation of the vehicle. Here we first obtain the centroid of one car then after few seconds we again find the centroid of same car if there is increase in coordinates of the centroid in positive direction of x-axis we define that vehicle is traveling in right direction. If the coordinates of the centroid is decreasing in negative direction of the x-axis vehicle is defined to travel in left direction. The flow chart of this algorithm is shown in Fig 3.2.

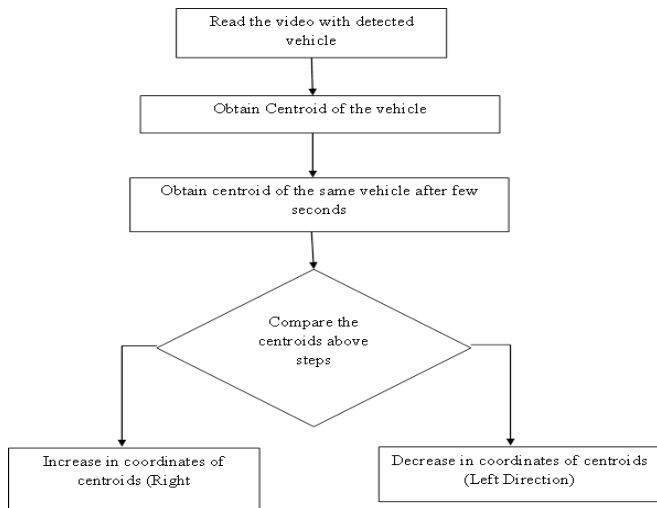


Fig 3.2 Flowchart of Estimation of Orientation of Vehicle

IV. RESULTS

A. Vehicle Detection

We are first defining three Gaussian parameters to detect foreground from one of the video frame. The Fig 4.1 shows one of the video frames extracted from the traffic video of aerial image. Fig 4.2 describes about Gaussian Parameter.

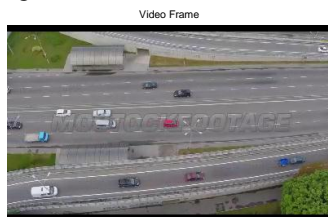


Fig 4.1 Traffic scene extracted from aerial video

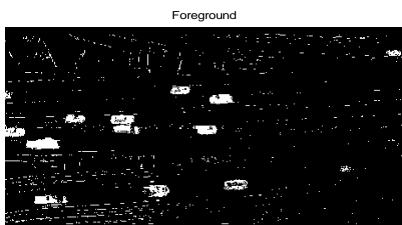


Fig 4.1 Moving Vehicle detected using GMM

The foreground contains some noise which can be removed using morphological area opening operation. We first define square as a structuring element and set some threshold which will remove noise as well as lane markings. The result of area opening operation is as shown in Fig 4.3



Fig 4.3 Clean foreground after Morphological Operation

For image processing, a blob is defined as a region of connected pixels. Blob analysis is the identification and study of these regions in an image. The algorithms discern pixels by their value and place them in one of two categories: the foreground (typically pixels with a non-zero value) or the background (pixels with a zero value). The Fig 4.4 shows output of Blob Analysis.



Fig 4.4 Vehicle detection after Blob Analysis

The above steps are then applied to all the video frames. Fig 4.5 & 4.6 shows one of the frames of two different videos. In this frame green and red colour boxes are bounding boxes of detected cars which depend on the area of blob defined in step3. By comparing the result we could conclude that efficiency of the system is above 90% if background is simple. In Fig 6.6 the background is complex and there are chances of false detection of blobs, in such case efficiency reduces to 70%

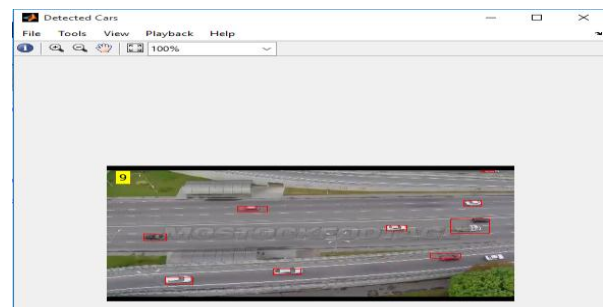


Fig 4.5 Vehicle detection in simple Infrastructure



Fig 4.6 Vehicle Detection in Complex Infrastructure

The above frame also gives some false detection of roof tops and some of the connected blobs. This disadvantage can be overcome in the next method. Table .I shows the Tracking efficiency of the above algorithm for simple infrastructure.

Frame Number	Cars passed	Cars Detected	Efficiency in %
250	10	9	90
400	8	8	100
550	10	10	100
620	10	8	80
Total Efficiency			93

Table. II shows the Tracking efficiency of the above algorithm for complex infrastructure.

Frame Number	Cars passed	Cars Detected	Efficiency in %
250	10	6	60
850	8	6	75
1335	13	10	76
1436	15	11	73
Total Efficiency			71

From Table number I and II we conclude that Gaussian mixture model is successful in case of simple infrastructure whereas in case of complex structures efficiency of the system has reduced.

B. Vehicle Classification Based On type

Each detected Bounding Box will undergo classification based on type (Car or Truck). We first set the threshold. The bounding box whose size is less than the threshold value will undergo category A i.e Car. The bounding box whose size is greater than the threshold value will undergo category B i.e. truck. Following are the results obtained after classification. Fig 4.7 shows the result obtained.

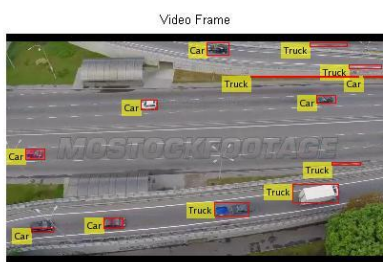


Fig 4.7 Vehicle Classification based on type

After comparing results of around 60 frames we conclude that this method is good to classify vehicle based on size. With the comparison of all the frames we conclude that this method will give accurate result when the background is simple. In case of complex infrastructure the Bounding box size will vary leading to fluctuations in result. In some frames we see that in 10 vehicles we get accurate result only for 8 or 7 vehicles whereas there is false output in case of change in size of bounding box. So this method is giving good result in case of simple background with 80% accuracy. From the Table III we

conclude that the algorithm will work efficiently for simple background with the accuracy of 89%

Frame No.	Number of cars	Correct Classification	False Detection	Accuracy in %
200	10	9	2	90
320	12	10	2	83
380	9	9	0	100

C. Estimation Of Vehicle Orientation

Here we take two frames after certain interval. In these frames we calculate the centroid of the vehicles. If the centroid of the vehicle is shifted in positive direction or if there is increment in coordinates of the centroid on x-axis that indicates vehicle moving in positive direction. Similarly if there is decrement in the coordinates of the centroid on the negative direction of x-axis it indicates vehicle moving in left direction. Results obtained for the same is shown in Fig 4.8. In the figure yellow line is indicating left direction whereas green line indicates right direction.

Result with the accuracy is well understood from the Table IV

Frame No.	Number of cars	Correct Classification	False Detection	Accuracy in %
200	10	8	2	80
320	12	10	2	83
380	9	9	0	100
400	10	8	2	80

From the above table it is clear that the estimation of vehicle orientation is giving good result with 86% accuracy. But in case of heavily traffic loaded area the accuracy will reduce to some extent which can be minimized by using some filtering methods. We can also use deep neural network classifier to classify on four way roads. The images with result of orientation estimation for two different frames are shown in following figures.

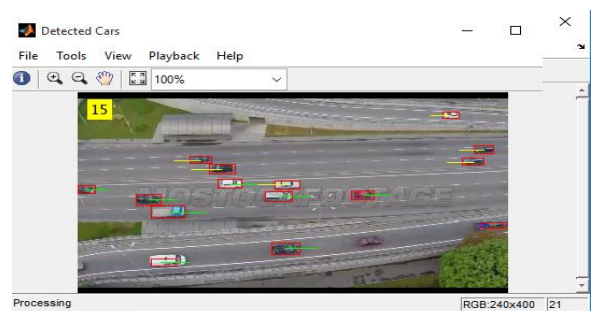


Fig 4.8 Output of Estimation of orientation of vehicle

problem is number of boxes keep changing, so this logic will fail in cases where the boxes number will change, so false data will be used, box 1 value might be compared with box 2 value Best way is to average the continuous centroid difference values rather than estimating every time in successive frame.

CONCLUSION

We have presented a method which can detect vehicles with orientation and type information on aerial videos. The application of Gaussian mixture model with cascade combination of classifier is good for applications where we would like to know in detail about the vehicle. Also the vehicle detection is useful in urban planning and traffic management. The application of Gaussian mixture model gives better efficiency when there is a simple background, whereas in case of complex background efficiency reduces to some extent. This disadvantage can be overcome by using classifier such as artificial neural network. The type classification of vehicle gives better result when the bounding boxes don't fluctuate. Whereas in case of heavily traffic loaded area there are chances of false detection. For the estimation of orientation of vehicle we are getting accuracy of 80%. It can be improved with application of kalman filtering. The accuracy of the overall system is in the range of 80 to 90%.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards my guide, Prof. Dr. Uma R. Rao, for the help, guidance and encouragement, provided during the ME Dissertation. This work would have not been possible without her valuable time, patience and motivation. I thank her for making my stint thoroughly pleasant and enriching. It was great learning and an honour being her student.

I would also like to express gratitude to all my colleagues and fellow students for their valuable guidance and co-operation as and when needed. Last but not the least I would like to thank all the helping hands which directly or indirectly helped me during my project.

REFERENCES

- [1] Kang Liu, Mattyus, Gellert. "Fast Multiclass Vehicle Detection on Aerial Images". *Geo science and Remote Sensing Letters, IEEE* Year: 2015, Volume: 12, no.9.
- [2] Vatau.A, Danescu.R, "stereovision based multiple object tracking in traffic scenarios using free form obstacle delimiters and practice filters" *Intelligent transportation system*, vol.16.
- [3] Jiann-Yeou Rau; Jyun-Ping Jhan; Ya-Ching Hsu "Analysis of Oblique Aerial Images for Land Cover and Point Cloud Classification in an Urban Environment" *Geoscience and Remote Sensing, IEEE Transactions on* Year: 2015, Volume: 53.
- [4] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 3, pp. 1635–1647, March 2014.
- [5] D. Rosenbaum, R.E. Schapire "Detecting cars in uav images with a catalog-based approach," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 10, pp. 6356–6367, Oct 2014.
- [6] X. Chen, S. Xiang, C. Liu, and C. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 10, pp. 1797–1801, Oct 2014. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [7] Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3D teacher for car detection in aerial images". in *Computer Vision*, 2007.
- [8] S. Tuermer, F. Kurz, P. Reinartz, and U. Stilla, "Airborne vehicle detection in dense urban areas using hog features and disparity maps," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 6, pp. 2327–2337, Dec 2013.
- [9] J.J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, "An operational system for estimating road traffic information from aerial images" *Remote Sensing*, vol. 6, no. 11, pp. 11 315–11 341, 2014.
- [10] Jaesik Choi Titled as "Real-time On-Road Vehicle Detection with Optical Flows and Haar-like feature detector" *ISSN: 2231-2306, Volume-2, Issue-4, May 2012*
- [11] P.Viola, M.Jones, and D.Snow. "Detecting pedestrians using patterns of motion and appearance", 2003.1,2,3
- [12] G.D.Hager and P.N.Belhumeur .Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039,1998.2
- [13] J. Wang, X. Chen, and W .Gao. Online selecting discriminative tracking features using particle filter. In *CVPR'05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)- Volume2*, pages 1037–1042, Washington , DC ,USA, 2005. IEEE Computer Society.2
- [14] P.Viola and M. Jones. "Robust real-time object detection." *International Journal of Computer Vision* to appear, 2002.
- [15] Y. Freund and R.E. Schapire. "A decision theoretic generalization of online learning and an application to boosting". In *Proceedings of the Second European Conference on Compu-1*
- [16] D. Hari Hara Santosh, and P. Venkatesh, P. Poornesh "Tracking Multiple Moving Objects Using Gaussian Mixture Model" *ISSN: 2231-2307, Volume-3, Issue-2, May 2013..*