Journal of Engineering Sciences

Vol 11, Issue 7, July /2020

ISSN NO:0377-9254

Automatic Change Detection System over Unmanned Aerial Vehicle Video Sequences Based on Convolutional Neural Networks

[1] BORRA SUDHAKIRAN [2] JENNE HANUMANTHU

[1] Assistant Professor, Dept. of Electronics and Instrumentation Engineering, Adikavi Nannaya University, Rajamahendravaram, East godavari, AP, India.

[2] Assistant Professor, Dept. of Electronics and Communication Engineering, Adikavi Nannaya University, Rajamahendravaram, East godavari, AP, India.

ABSTRACT: In recent years, the use of unmanned aerial vehicles (UAVs) for surveillance tasks has increased considerably. This technology provides a versatile and innovative approach to the field. However, the automation of tasks such as object recognition or change detection usually requires image processing techniques. In this paper we present a system for change detection in video sequences acquired by moving cameras. It is based on the combination of image alignment techniques with a deep learning model based on convolutional neural networks (CNNs). This approach covers two important topics. Firstly, the capability of our system to be adaptable to variations in the UAV flight. In particular, the difference of height between flights, and a slight modification of the camera's position or movement of the UAV because of natural conditions such as the effect of wind. These modifications can be produced by multiple factors, such as weather conditions, security requirements or human errors. Secondly, the precision of our model to detect changes in diverse environments, which has been compared with state-of-the-art methods in change detection. This has been measured using the Change Detection 2014 dataset, which provides a selection of labelled images from different scenarios for training change detection algorithms. We have used images from dynamic background, intermittent object motion and bad weather sections. These sections have been selected to test our algorithm's robustness to changes in the background, as in real flight conditions. Our system provides a precise solution for these scenarios, as the mean F-measure score from the image analysis surpasses 97%, and a significant precision in the intermittent object motion category, where the score is above 99%.

1 INTRODUCTION

The use of change detection algorithms is crucial in high precision surveillance systems. The methods that make use of those algorithms aim to detect the differences between information acquired at the same location, e.g., an image captured in different moments. Unmanned aerial vehicles (UAVs) became a revolution in the surveillance sector due to the lower cost and reduced human workload needed compared to previous systems. In addition, UAV operations can be automatized. This need of automation increases the importance of change detection methods. These methods are based on image sequences analysis, usually acquired by mobile vehicles. Image acquisition from the mentioned vehicles entails a considerable issue for change detection algorithms: The camera movement. This is the fundamental challenge of the algorithms, as the movement produces a variable background, thus the flight's route will be moderately modified from one flight to another. Furthermore, the weather conditions and the precision of GPS positioning influence the relation between the acquired frames and the location of the UAV. Compared to moving cameras, stationary cameras significantly reduce the complexity of the change detection problem, as the background is common to every image [1]. Moving cameras introduce complexity to the problem because the reference image is continuously changing. As a result, the system needs to detect the background of each image to provide precise detection. Therefore, variable backgrounds introduce a considerable computational load. This generates a complex problem to solve for real-time video surveillance tasks. An interesting approach for change detection with moving cameras. Our method is based on reconstruction techniques. However, the reconstruction process alone would not be precise enough for our task. This is a consequence of the limited precision of CNNs to generate detailed images. In addition, it conforms an additional process that could slow down the system. Nevertheless, it provides a distinct perspective from the stationary camera algorithm mentioned above. After reviewing the different state-of-the-art approaches we have observed an increasing tendency to use CNNs on image processing systems for change detection. Furthermore, studied implementations do not consider a moving camera, as in [3,5]. Consequently, a supplementary component has been considered to overcome the problems introduced by a variable background.

2 PROPOSED WORK

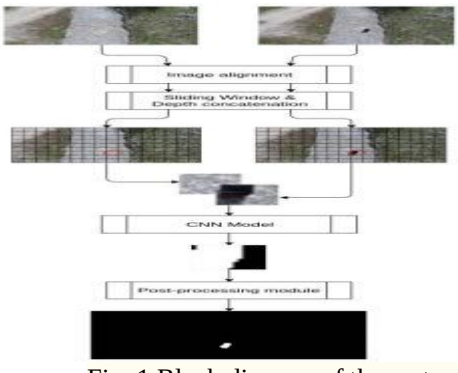

Fig: 1 Block diagram of the system

Both reference and foreground images are introduced into our image alignment system. After that, sliding window algorithm is applied. The two resultant patches are concatenated along the depth axis. Ultimately, our CNN model predicts a grayscale image, which is post processed to obtain the final binary patch depicted.

2.1 IMAGE ALIGNMENT

As stated on previous sections, our system is applied to moving cameras. Our objective is to compare two video sequences (background and foreground) to detect the changed regions between them. Because of implementation purposes, a reference video from the UAV's route is required, which will be considered as background. More recent videos from the same route are considered as our foreground scenarios. In real-world situations, the new images will not be completely aligned to the reference video. This can be caused by multiple reasons such as lack of GPS precision or weather variations. To solve the problem, an image alignment system has been developed using ORB [15].

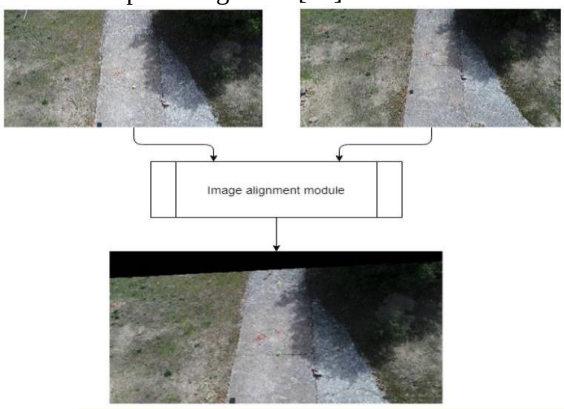

Fig: 2 Example of our image alignment module's output with a foreground image with significant difference from the reference.

The idea is similar to the feature alignment performed in [22]. Both the reference and the new route's images are compared. Feature extraction is performed with ORB algorithm to obtain the most significant zones of each picture. After that, a descriptor matcher algorithm from [23] is created. The descriptor performs an analysis of the obtained values and outputs the relation between them by distance difference. This output is filtered and sorted to obtain the most adjacent descriptive points of both images. Lastly, a geometric transformation is performed to generate a modified version of the acquired image, aligned to the reference. An example of the results obtained by this system is depicted in Figure 2. As can be observed, the resultant image contains a black zone that represents the pixels from the foreground image not included in the reference. To prevent the appearance of false positives because of these black zones, only the mid section of the image is selected automatically to perform the change detection. The alignment system entails an innovation to other implementations such as [3,12], which consider a scenario with a static camera.

2.2 SLIDING WINDOW

The image provided to our system can vary in size depending primarily on the UAV's camera device or time processing requirements. For instance, the processing requirements of a real-time detection system varies from those of a post-flight analysis. Moreover, deep learning models usually struggle to work with very high-resolution images. Our approach to resolve these problems is described in this subsection. In the first place, the maximum input size is defined as a parameter of the system. If the input images overcome the defined dimensions, they are resized to the specified size. After that, we have to divide the pictures into small regions. This is a consequence our network's requirements to process images in reasonable time. Therefore, we can obtain pixel-level precision results.

To do so, we have employed an algorithm named sliding window. This algorithm iterates over the image's dimensions, retrieving a matrix of a specific size (window) which contains a region of the initial image. The dimensions of the window are identical in width and height to prevent any further complication of the segmentation process. Another parameter of the algorithm is defined as the step. The step of a sliding window algorithm represents the distance, in each dimension, from the starting point of a window in iteration i to the starting point of the region in iteration i + 1. Adaptive to each dimension, the step remains a crucial factor in terms of computational cost, just as the dimension of the window. With these two parameters we can control the overlapping of zones among adjacent windows. If overlapping occurs, the system will benefit from an increment of precision on the analysis. In this case, four predictions are generated for the most part of the image, except the limits of each dimension. As a conclusion, the variation of both parameters provides an extremely efficient tool that can modify the algorithm's performance to adapt it for multiple purposes: real-time segmentation, maximum precision prediction or sample generation. The application of this algorithm in both the reference image and background image is depicted in Figure 3.

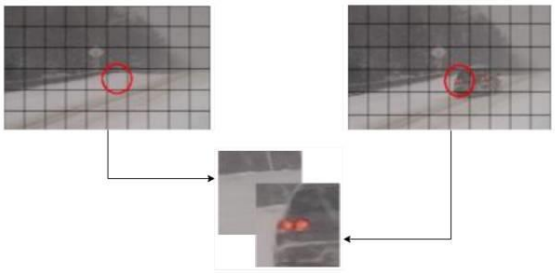

Fig: 3 Representation of the sliding window and depth concatenation process on a reference image and a foreground image.

2.3 Deep Neural Network Architecture

The model is based on the concatenation of two input images: The reference background scene and the updated scene image which may contain some changes. Both images are merged in depth dimensions, as it is performed in other state-of-the-art methods such as [3,4,12]. This input form allows the model to learn hierarchical features. As a result, CNNs conform an effective tool to obtain relevant information from images. An exhaustive description of the CNN architecture is described in detail in Figure 4 illustrates the complete architecture. The deep neural network is composed by four CNNs with increasing number of filters and a kernel size of 3 × 3 pixels. As our inputs consist of images with reduced dimensions, we employ this kernel size to extract the features as detailed as possible to obtain a precise detection. For the activation layers of the CNN, we have used Rectified Linear Unit (ReLU) activation. ReLU activation applies Equation to its inputs. This function is widely implemented in CNNs as mentioned in Section 2.2 because of its reduced computational cost and the acceleration of the optimization process. Following this layer, we have employed a max-pooling layer with 2 × 2 kernel size.

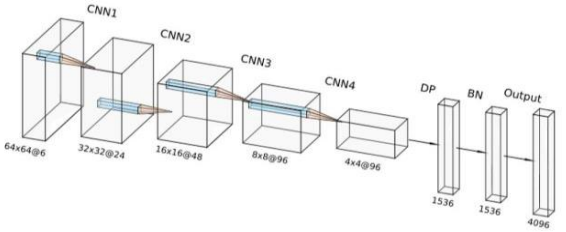

Fig: 4 Diagram of the deep neural network architecture

4 SIMULATION RESULTS

There are several key points that are highlighted through comprehensive evaluation and results. First, the results support the claim that using RGB and YCbCr color spaces produce more accurate results in comparison to when they are employed individually. This is clear from overall higher position and ARC of both MB2S-Standard and MB2Sdefault than MB2S-RGB and MB2S-YCbCr. Second important point is the performance of MB2S-RGB and MB2S-YCbCr for Night Videos(NV) and Bad Weather category. NV has low lighting conditions and BW has poor color discrimination problem. We can see that MB2S-RGB has an average ranking of 5.28, which is not only higher than that of 6.71 of MB2S-YCbCr but also higher than 6.28 and 7.71 of MB2S-Standard and MB2S-default, respectively. Likewise for BW, MB2S-RGB has higher average ranking than other three. This supports our earlier claim that RGB is more robust under low lighting conditions or when color discrimination is poor.

Third, in general, RGB performs well in simple background scenes with minimum noise, whereas YCbCr is more robust against noise. This is evident from higher average ranking of MB2S-RGB in categories such as BL and LFR and higher average ranking of MB2S-YCbCr in CJ and DB categories. Fourth, despite use of standard parameter setting and fixed number of BG models N, our proposed system performs well in 7 categories, whereas it performs poorly in 3 categories; Thermal, Bad Weather and Turbulence. This has resulted in overall position to drop down to 4th. The main reason is that in some of video sequences the scene changes over time and model update is required. This is lacking in our current system and has resulted in poor performance in aforementioned categories. For example in case of thermal, in one of video sequences when a person sitting on a chair stands up after a while and leaves, the higher temperature of chair results in misclassifications of chair as foreground. Another example is from one of test sequences in bad weather in which when snow is cleared from pathway, it becomes foreground and remains foreground since model is not updated resulting in poor performance.

Fig: 5 Effect of the image alignment component on the system's output

5 CONCLUSION

In this paper we have presented a change detection system for static and moving cameras using image alignment based on ORB algorithm and convolutional neural networks. Because of the use of UAV imagery acquired by moving cameras, the problem of dynamic backgrounds has been addressed. As we have detailed along the paper, our mayor improvement from other state-of-the-art implementations consists on the use of an image alignment process. The objective of this element is to compensate the possible variations during UAVs flights described in previous sections. In addition to that, the inclusion of the sliding window algorithm reduces the computational cost of the CNN model by reducing the dimensions of the input images. Moreover, this method adds versatility to the system. The sliding window algorithm can be adjusted to provide overlapping sections to improve accuracy with an increment in computational cost. As far as we know, a moving camera scenario has not been taken into account in any of the state-of-the-art methods for change detection compared along the paper. Only dynamic backgrounds on static cameras have been studied on mentioned implementations. Our system is capable of adapting to these conditions using image alignment techniques and the idea of a reference video or image. Datasets with dynamic backgrounds have been selected to train the network to achieve meaningful outcomes for real-world applications.

Results from experiments indicate a precise detection in scenarios with adverse weather such as a snowfall or a blizzard. The comparison with other state-of-the-art methods reflects that our system is the most accurate on the studied scenarios.

Future Work

The precision of the reference's acquisition is crucial for the system's performance. As a solution to this, we are working on GPS data processing for improving the alignment system. In addition to that, we consider the option to include deep learning in the alignment process as another of our futures lines of work. Our objective with that is to compare the performance of deep learning against our current image alignment system based on ORB.

REFERENCES

[1] Wang Y., Jodoin P., Porikli F., Konrad J., Benezeth Y., Ishwar P. CDnet 2014: An Expanded Change Detection Benchmark Dataset; Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops; Columbus, OH, USA. 23–28 June 2014; pp. 393–400.

[2] Minematsu T., Shimada A., Uchiyama H., Charvillat V., Taniguchi R.I. Reconstruction-Base Change Detection with Image Completion for a Free-Moving Camera. Sensors. 2018;18:1232. doi: 10.3390/s18041232.

[3] Babaee M., Dinh D.T., Rigoll G. A Deep Convolutional Neural Network for Video Sequence Background Subtraction. Pattern Recogn. 2018;76:635–649. doi: 10.1016/j.patcog.2017.09.040. [4] Braham M., Van Droogenbroeck M. Deep background subtraction with scene-specific convolutional neural networks; Proceedings of the 2016 International Conference on Systems, Signals and Image Processing (IWSSIP); Bratislava, Slovakia. 23–25 May 2016; pp. 1–4.

[5] St-Charles P., Bilodeau G., Bergevin R. Universal Background Subtraction Using Word Consensus Models. IEEE Trans. Image Process. 2016;25:4768–4781. doi: 10.1109/TIP.2016.2598691.

[6] Stauffer C., Grimson W.E.L. Adaptive background mixture models for real-time tracking; Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149); Fort Collins, CO, USA. 23–25 June 1999; pp. 246–252.

[7] Elgammal A.M., Harwood D., Davis L.S. Non-parametric Model for Background Subtraction; Proceedings of the 6th European Conference on Computer Vision-Part II; Dublin, Ireland. 26 June–1 July 2000; London, UK: Springer; 2000. pp. 751–767.

[8] St-Charles P., Bilodeau G., Bergevin R. SuBSENSE: A Universal Change Detection Method With Local Adaptive Sensitivity. IEEE Trans. Image Process. 2015;24:359–373. doi: 10.1109/TIP.2014.2378053.

[9] Barnich O., Van Droogenbroeck M. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. IEEE Trans. Image Process. 2011;20:1709–1724. doi: 10.1109/TIP.2010.2101613. [10] LeCun Y., Haffner P., Bottou L., Bengio Y. Shape, Contour and Grouping in Computer Vision. Springer; London, UK: 1999. Object Recognition with Gradient-Based Learning; pp. 319–345.

[11] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S.E., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions; Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Boston, MA, USA. 7–12 June 2014.

[12] Zeng D., Zhu M. Background Subtraction Using Multiscale Fully Convolutional Network. IEEE Access. 2018;6:16010–16021. doi: 10.1109/ACCESS.2018.2817129.

[13] Liu S., Deng W. Very deep convolutional neural network based image classification using small training sample size; Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR); Kuala Lumpur, Malaysia. 3–6 November 2015; pp. 730–734.

[14] Coombes M., McAree O., Chen W., Render P. Development of an autopilot system for rapid prototyping of high level control algorithms; Proceedings of the 2012 UKACC International Conference on Control; Cardiff, UK. 3–5 September 2012; pp. 292–297.

[15] Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF; Proceedings of the 2011 International Conference on Computer Vision; Barcelona, Spain. 6–13 November 2011; pp. 2564–2571.

Borra Sudhakiran is a Assistant professor Dept. of Electronics and Communication Engineering in Adikavi Nannaya University, Rajamahendravaram. He Received AMIETE in IETE. He Received his M. Tech in Embeded Systems in Lenora College of Engineering affiliated to JNTU Kakinada. He Currently Research in Image Processing, Video Signal Processing and Control Systems.

Jenne Hanumanthu is a Assistant professor Dept. of Electronics and Instrumentation Engineering in Adikavi Nannaya University, Rajamahendravaram.

He Received B. Tech in Electronics and Instrumentation Engineering from Sree Vidyanikethan College Tirupathi in 2001. He Received his M. Tech in 2011 From JNTU Kakinada. He Currently Research in Image Processing, Video Signal Processing and Control Systems.