An Improved Object Detection Algorithm that Utilizes the Motion Vector

Pyeong Kang Kim, Hyung Heon Kim, Tae Woo Kim and Young Kyun Cha

Abstract—— **With regard to the methods for the existing object detection, the CCTV images are decoded, and the object is then detected by analyzing the whole image information; the process incurred high computational burden, resulting in a speed reduction problem. Various techniques have been introduced to overcome these drawbacks. In order to maintain the real-time characteristics of multi-channel images, however, expensive hardware devices must be used. Thus, to solve these disadvantages, we suggested in this paper the object detection method wherein the motion vector based on the compressed stream is analyzed and set in motion. As a result of the performance evaluation, the detection rate was not significantly different from the existing method, but the detection speed was at least 21.5 times higher.**

Keywords— **object detection, convolutional neural network, Inception V4, Motion Vector**

I. Introduction

Today, most of the closed circuit television (CCTV) integrated control systems are expanding their domains, from human-eye monitoring and video recording to Intelligent Surveillance System. Intelligent surveillance systems analyze the CCTV image, recognize the object, identify the danger situation, and notify the administrator in real time, so it is possible to deal with the situation quickly and prevent crime.

What is important in the intelligent surveillance system is the technology for detecting valid objects in CCTV images.

In general, a background generation method or a background difference method is used as a technique for detecting an object in an image. The background generation is a method of dividing the background and the object and reconstructing the image so that only the background components exist. By the method of differentiating between input images, objects can be easily detected from the reconstructed background image. The most widely used technique for analyzing and detecting background components is the MoG (Mixture of Gaussian) [1]. MoG technology has the advantage of being able to handle slow changes of light, periodic motion in the background, slow moving objects, long-term scene changes, and camera noise effectively. In addition, techniques such as scale-invariant feature transform (SIFT) [2], sped-up robust feature (SURF) [3], and Histogram of Oriented Gradient (HOG) [5] were used. In recent years, object detection techniques using the convolutional neural network [6] have come to the fore.

Pyeong Kang Kim, Junior Engineer, INNODEP INC., KOREA
Hyung Heon Kim, Senior Engineer, INNODEP INC., KOREA
Tae Woo Kim and Young Kyun Cha, INNODEP INC., KOREA

Note, however, that such methods have a disadvantage, i.e., it is difficult to process multiple channels due to a relatively large amount of computation incurred in decoding and analyzing the input image. Such disadvantage increases the hardware cost and consequently raises the control center’s operating expenses.

Therefore, in this paper, we propose a compression stream-based object detection algorithm that detects moving objects without decoding the images to solve the disadvantage above. This paper consists of 5 chapters. Chapter 1 introduces the background and necessity of this paper; Chapter 2 describes the existing object detection technology. In Chapter 3, we map out the object detection method based on the proposed compression stream; in Chapter 4, we implement and test the designed algorithm and verify the proposed technique. Finally, Chapter 5 presents the conclusion and future direction of this research.

II. Object detection technology

The first step in detecting objects is background modeling. Typical techniques of background modeling include background generation techniques and background subtraction techniques. After background modeling is performed on the input image, objects are detected using techniques such as point, area segmentation, and statistical approach. Figure 1 shows the object detection flow diagram.

Fig. 1. Object Detection Flow Diagram

A. Background Modeling

- Background Generation Technology

Background generation method is a method of dividing the background and the object; it is a method of reconstructing the image so that only the background components exist. The most widely used technique for analyzing and detecting background components is MoG (Mixture of Gaussian). The MoG technique can effectively deal with slow changes of light, periodic motion in the background, slow-moving objects, long-term scene changes, and camera noise.

- Background Subtraction Technology

https://doi.org/10.17758/URUAE8.UL12191012

The background subtraction technique can be performed by calculating the deviation between the image and the background image. The part where there are significant changes in the difference values between the corresponding image and the background image is assumed to indicate the object in motion. In order to detect the movement area, the area where the changes appear is indicated by another symbol; the object area is then detected through connection check

B. Object Detection Technique

- Object Detection Using Points

The object detection technique using points is a method of analyzing the texture information expressed in an area to find a point having noise and continuity and determine the point as an object. Most point methods have been used for a long time in motion, stereo, and tracking problems, but the all-important technique is finding a specific location that does not change with the orientation of light and camera. The Kanade-Lucas-Tomasi Feature Tracker (KLT) [9] and the Scale-Invariant Feature Transform (SIFT) are notable examples of methods that utilize points.

- Object Detection Using Area Segmentation

Area segmentation is a method of dividing an image into similar areas by repeated calculation, using information such as color, brightness, and texture. The performance of this segmentation method varies depending on the conditions of how to define the initial values (color, brightness, texture). The mean-shift technique for finding clusters in the color space is most commonly used.

- Object Detection Using Statistical Approach

Measurements obtained from video images always contain noise due to environmental factors. In addition, motion occurs together with noise, resulting in object motion and confusion. The statistical method is a way of resolving such tracking problems by obtaining uncertainties that occur during the measurement and modification of the object state. The statistical method involves estimating the next state by modeling the object state values such as location and velocity changes.

The Kalman Filter [10] and the Particle Filter [11] are typically used

III. MOTION OBJECT DETECTION USING MOTION VECTOR

A. Problems with the Existing Technology

There are various methods for detecting objects as listed above. In recent years, an object detection method (R-CNN, etc.) combining a sliding window technique or a Region Proposal technique for detecting an existing object and a composite neural network has appeared, showing high detection performance.

In the case of the techniques above, however, the entire image information is analyzed after the CCTV image is decoded to detect the object; hence the disadvantage of high computation amount and the resulting speed reduction. In order to compensate for these drawbacks, methods such as resizing or reducing the number of image analysis frames have been applied. In order to maintain the real-time property of multi-channel images, however, expensive hardware devices have to be used, and such reduces the efficiency of resource use of the institutions adopting them.

Therefore, in order to solve these problems, we propose a method of detecting moving objects that analyzes the undecoded compressed streams.

B. Motion Vector Detection

Motion vector means a distance represented by a vector (displacement) when motion estimation is performed.

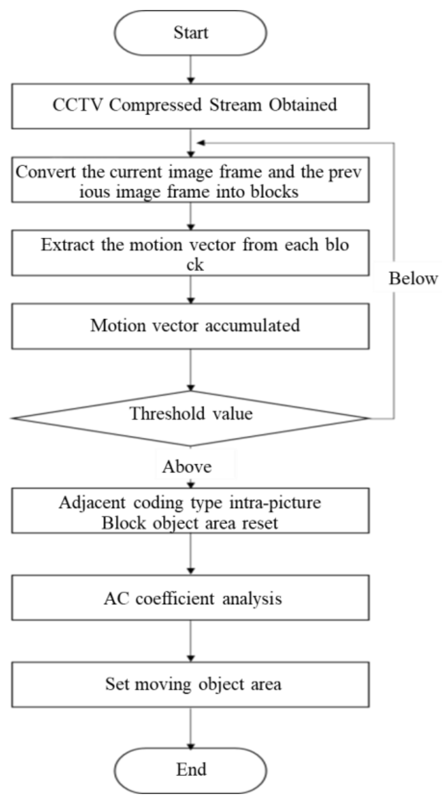

Fig. 2. Primary Object Candidate Area Detection

In this paper, the compressed stream is received from CCTV, and the previous image frame and the current image frame are compared to convert the image into blocks and extract the motion vector from each block. Fig. 2 represents the entire algorithm.

The motion vector generates a three-dimensional histogram wherein the X, Y coordinates indicate the direction of the motion vector and the Z coordinate indicates the size of the motion vector. If the direction and size of the motion vector are similar to the preset values, the motion vector is accumulated to detect the area of the moving object. The cumulative time is set to 500ms; if the cumulative value does not exceed the predetermined threshold value, the image is determined as noise and ignored in the detection phase.

The area of the motion object detected in the phase above may not be equal to the size of the object in the real image, or

https://doi.org/10.17758/URUAE8.UL12191012

multiple detection areas may be set in a single object. In addition, there is a problem of the object not being detected in the case of a still object. Fig. 2 shows the primary object candidate area detection

Fig. 3. Primary Object Candidate Area Detection

In order to solve the problem above, multiple adjacent image blocks are re-identified with focus on the primary candidate area block. The motion vector value for the various adjacent video blocks was compared with the 2nd threshold value, and the block with motion vector value exceeding the 2nd threshold value is set as the region proposal block again. In addition, among multiple adjacent blocks, those with intra-picture as coding type are reset to the object area. Intra-picture does not have a motion vector, so it is impossible to determine whether an object is in the adjacent block based on the motion vector. In this case, the adjacent intra-picture should be regarded as region proposal to include the object area. Finally, AC coefficients were analyzed to detect objects that do not show moving displacement, with a certain pattern range detected as the object area. Fig. 3 represents the results of detection of the motion object.

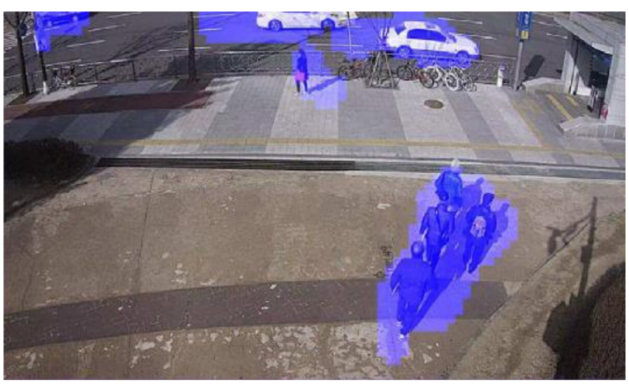

Fig. 4. Motion Object Detection

IV. Implementation and Testing

For implementation and performance evaluation, we performed with the Intel (R) Core (TM) i7-6700 CPU @ 3.44GHz and RAM 16.0GB environment in Windows 10 OS. The Video Dataset for implementation and performance evaluation was tested based on ten arbitrary images in possession. Fig. 5 shows the test image.

Fig. 5. Test Image

A. *Algorithm Implementaion*

The object detection method based on the compressed stream proposed in this paper is implemented as shown in Fig. 6. Fig. 6 (a) shows the result of detecting a correctly moving object. Figure 6 (b) shows the faulty detection result.

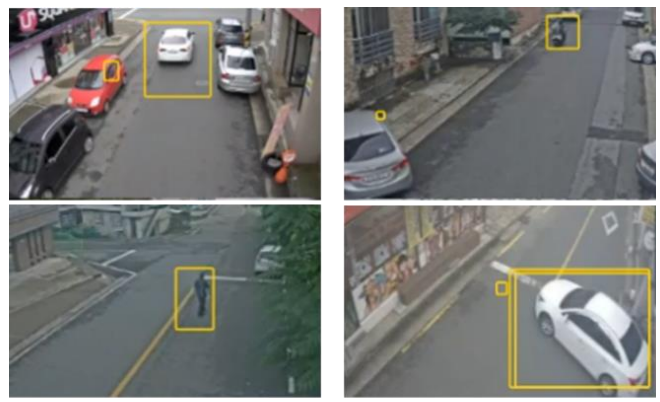

Fig. 6. Implementation Results

B. *Testing*

For the performance evaluation, the bit stream is received from the existing CCTV and decoded; using the background modeling and subtraction images, the detection rate and detection speed of the (1) conventional object detection method and (2) object detection method using the proposed compression stream-based motion vector were comparatively evaluated. Table 1 shows the evaluation results.

TABLE I : EVALUATION RESULTS

Classification	Suggested Technique		Existing Technique	
	Recognition Rate	Speed	Recognition Rate	Speed
Video 1	88.5%	930fps	89.5%	45fps
Video 2	84%	905fps	85.7%	46fps
Video 3	87%	942fps	86.4%	40fps
Video 4	83.2%	924fps	82.9%	42fps
Video 5	80.9%	940fps	83.4%	43fps
Video 6	87.4%	927fps	89.4%	45fps
Video 7	89.8%	930fps	87.6%	47fps
Video 8	81.6%	910fps	83.4%	40fps
Video 9	86.5%	915fps	84.1%	41fps
Video 10	88.7%	934fps	87.9%	43fps
Average	85.76%	926fps	86.03%	43fps

As a result of the performance evaluation, the detection rate of the proposed technique is 85.76%, which is not significantly different from the conventional technique's detection rate of 86.03%.

In terms of detection speed, the proposed technique showed 926 fps, which is about 21.5 times faster than the conventional technique's 43 fps.

V. CONCLUSION

In the existing object detection methods, CCTV images are decoded, with the entire image information analyzed to detect the object; hence the high computation amount and the resulting speed reduction. In order to compensate for these drawbacks, methods such as resizing or reducing the number of image analysis frames have been applied. In order to maintain the real-time property of multi-channel images, however, expensive hardware devices have to be used; this reduces the efficiency of resource use of institutions that adopt them. Therefore, in this paper, we have proposed a method that enhances the detection speed wherein the object is detected by analyzing the motion vector that is based on the compressed stream, in order to solve the problems above.

As a result of the comparison with the existing technique, it was confirmed that there were no significant differences in the detection rate, whereas the detection speed was at least approximately 21.5 times higher than that of the existing technique. Through the proposed method, real-time implementation is considered possible; hardware dependency can be eased so as to enhance the efficiency of resource use of the adopting institution. In the future, we aim to conduct studies on improving object class classification and algorithm

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. 2014-3-00123, Development of High Performance Visual Discovery Platform for Realtime and Large-Scale Data Analysis and Prediction)

REFERENCES

[1] R. Jain, and H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," IEEE Trans. Pattern Analysis, Machine Intelligence, Vol.1, No.2, pp.206-14, 1979. https://doi.org/10.1109/TPAMI.1979.4766907

[2] D.G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," Inte. J. Comput. Vision, vol. 60, no. 2, 2004, pp.91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[3] H. Bay et al., "Speeded-Up Robust Features(SURF)," Comput. Vision Image Understanding, vol. 110, no. 3, 2008, pp.346-359. https://doi.org/10.1016/j.cviu.2007.09.014

[4] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc.IEEE Conput. Soc. Conf. Comput. Vision Pattern Recogn., · · Kauai, HI,USA, Dec. 8-14, 2001, pp.I:511-I:518

[5] N.Dalal and B.Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn., San Diego, CA, USA, June 20-25, 2015, pp. 886-893.

[6] A. Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks," Conf. Neural Inform. Process. Syst., Lake Tahoe, NV, USA, Dec. 3-6, 2012,pp.1097-1105.

[7] R. Girshick et al., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, " IEEE Conf. Comput. Vision Pattern Recogn., Columbus, OH, USA, June 23-28, 2014, pp.580-587. https://doi.org/10.1109/CVPR.2014.81

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," In ECCV, 2014. https://doi.org/10.1007/978-3-319-10578-9_23

[9] J. Shi and C. Tomasi, "Good features to track," IEEE Conf. Computer Vision, Pattern Recognition, .593-600, 1994.

[10] G. Welch, G. Bishop, "An Introduction to the kalman filter," Technical Report, Department of Comp. Sc. and Engg., Univ. of North Carolina at Chapel Hill, 2002.

[11] Z. Khan, T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets," Euro. Conf. Computer Vision, Vol.3024, pp.279-290, 2004. https://doi.org/10.1007/978-3-540-24673-2_23

PyeongKangKIm
Graduated from Korea National University of Transportation,
Dept. of Public Administration, 2012
Graduated from Gachon University , Mobile Software, MS in 2015
2015~ Researcher of Innodep Technology Lab.
Depth field is pattern recognition, machine learning, video security.