ICIC Express Letters
Part B: Applications
Volume **11**, Number **9**, September **2020**

ICIC International ©2020 ISSN 2185-2766
pp. **855–861**

IMPORTANT OBJECT SEGMENTATION AND TRACKING USING FEATURES AND OCR

Mie Mie Tin[1], Mie Mie Khin[2], Nyein Nyein Myo[3], Thi Thi Zin[4] and Pyke Tin[4]

[1]Faculty of Information Science
Myanmar Institute of Information Technology
73rd Street, Ngu Shwe Wah Street, Chanmyathazi Township, Mandalay, Myanmar
mie_mie_tin@miit.edu.mm

[2]Faculty of Information Science
Computer University (Meiktila)
Meiktila, Myanmar
miemie.khin9@gmail.com

[3]Faculty of Information Science
University of Computer Studies, Mandalay
Mandalay, Myanmar
vicky.mdy@gmail.com

[4]Graduate School of Engineering
University of Miyazaki
1-1, Gakuen-Kibanadai Nishi, Miyazaki 889-2192, Japan
thithi@cc.miyazaki-u.ac.jp; pyketin11@gmail.com

Received March 2020; accepted June 2020

Abstract. *Image processing and segmentation support to analyze image and big data in many different research sectors. Object tracking in a surveillance video sequence supports to the research areas. This system processes on the security surveillance camera network and extracts the important object track information. First is the extraction of the key frames from videos. These key frames are used to extract objects on different cameras and from different background areas. To extract key frames from surveillance videos, the system uses RGB color values and the block method with diagonal movement. The important object is segmented from one key frame and the system searches that object a crossing of other key frames. To segment an object, the system uses OTSU's threshold method. The comparison of an important object and the other segmented objects uses color moment features and computes the similarity value based on the histogram. The system handles all key frames to extract the similar objects. To find the tracking region of that object on different background regions, the system uses time information on key frames in the video networks. To extract time information from video key frames, the system uses character extraction and recognition with the Optical Character Recognition (OCR) method on the gray level images.*
Keywords: Image segmentation, OTSU's threshold, OCR, Feature extraction

1. **Introduction.** This research supports big data analysis and the objective is to extract moving object tracking regions from video networks. The main focus of this research is object segmentation on different background areas based on features by using key information frames.

Image segmentation supports dividing of the image into disjointed homogenous regions or classes. That is all the pixels are organized in the same classes based on their common characteristics. The segmentation is applied to bands of the RGB images and tested on

DOI: 10.24507/icicelb.11.09.855

the HSI. The result produces good segmentation by processing the intensity band of the HSI model [1]. Their technique is based on optimal, linear time, which is that computation of different weight of the data is automatically segmented. The weights are based on spatial and temporal gradients. The localization refinement step follows fast segment and accurately computes the corresponding matte functions. The adding constraints on the distance definition permits us to efficiently handle occlusions such as people or objects crossing each other in a video sequence [2]. This survey provides color image segmentation techniques. Color segmentation is based on monochrome segmentation approaches. To segment object image histogram, threshold value, feature clustering, edge detection, region-based methods, fuzzy techniques and neural networks are involved [3].

Combinations of different color spaces such as RGB, HSV, YIQ and XYZ are used for image segmentation. The combination of different color spaces gives more accurate segmentation results. K-means clustering and Effective Robust Kernelized Fuzzy C-Means (ERKFCM) are used to segment objects [4]. The authors proposed new method for object segmentation for objects in motion and color base in different backgrounds. Their algorithm presents a major extension to the state-of-the-art and the original Active Shape Model (ASM) using landmark points in a stable background [5]. This paper is presented to detect and track in thermal images. Their framework exploits raw H.264 compressed video streams and they work on Motion Vectors (MV) by using a video compression technique [6].

They have used video surveillance systems in public places with different activities and views. They have considered the scene, event and have analyzed human movements. The system recognizes objects and observes the effect of human movement on those objects. The system uses high level motion feature extraction method and embeds Markov chain models to detect object behaviors. Probability-based multiple background modeling technique is used to detect moving objects [7].

Wicaksono and Setiyono proposed the vehicle speed estimation system by using image processing from video data. Euclidean distance method is used based on many different camera and different angles views for this system. Their system extracted the foreground image by using the Gaussian Mixture Model (GMM) and filter foreground region using the median filter. Euclidean distance approach is used to find speed estimation, camera moves angle and ROI selection [8]. Huang et al. suggested the novel metric learning framework to face video recognition and to know a distance metric between Euclidean space and a Riemannian manifold. They used three basic methods for face recognition and these are Video-to-Still, Still-to-Video and Video-to-Video settings [9].

This research has used the MIIT University surveillances network and a total of 85 cameras are installed. These networks use HIKVISION network bullet cameras and HIKVISION E series network speed dome camera. All surveillance cameras are configured at different stable places and the camera view is stable in vision. This research is ongoing and five members are working as a group. This research has been tested on public and private datasets of the surveillance camera network from Myanmar Institute of Information Technology University (MIIT), Mandalay, Myanmar.

2. **Problem Statement and Methodology.** Tracking selected objects in a security camera network is very important for safety and security. It can support some accident cases in the city traffic system and security control systems. The system changes to select object tracking region from key frames. The first key frame is collected from important case point surveillance videos.

3. **Proposed Method.** To select object tracking areas in camera networks, there are many different steps, as follows:

Step 1: Key frame extraction from different surveillance videos in the network.

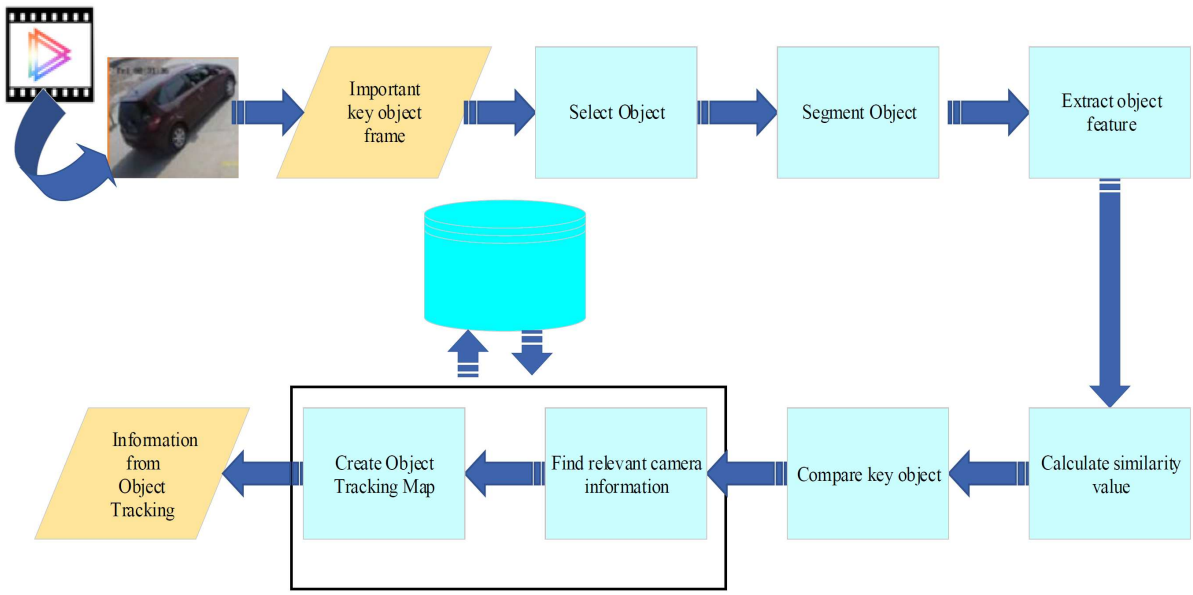


FIGURE 1. System processing steps

Step 2: Selection of the important object from one key frame and resizing that main object.
Step 3: Feature extraction from that main object.
Step 4: Foreground objects extraction from key frames in other surveillance videos and resizing that foreground object into the same dimension.
Step 5: Comparison of the main object and other objects based on feature values.
Step 6: For high similarity objects, creation of the database based on time value from frames information using character extraction and recognition.
Step 7: Analysis of object information and extraction of tracking areas based on a time schedule.

The system process includes these seven steps and steps two to six based on image processing techniques are object extraction, object segmentation, foreground detection and character image detection and reorganization processes.

3.1. **Key frame extraction on surveillance videos.** This system extracts key frames from many different surveillance cameras. To extract key frames from a surveillance camera video, the Block Diagonal Movement Technique (BDMT) is used. This technique gives excellent results by reducing the calculation time and complexity rate. The system extracts three key frames and this research uses only one excellent key frame with the best clarity [10]. These key frames are very important to extract important objects and to track object movement in video networks. The user extracts an important object from the key frame and the system extracts similar objects in a surveillance camera network.

3.2. **Selection of important key objects.** The user selects the important object from the key frame and drafts it as a rectangle. The video frame dimension is 2688×1520 pixel values. The system extracts the selected area as one rectangular image and that image is resized to dimensions of 100×100 pixels. The system segments the object using the global threshold technique.

3.3. **Segmentation and extraction of feature values.** The RGB color moment value from the selected object is compared with objects from other key frames. Object segmentation from key frames uses OTSU's method. To extract the objects from the key frame, the system segments the foreground and background by using a color segmentation algorithm. The system uses the OTSU's threshold method to segment the object from

the selected important object frame. The moment features based on RGB values and hue value are calculated for that object.

For other key frames, the foreground object and background objects are segmented using the threshold technique. The foreground object image is resized to the 100×100 image frame. That object's images are compared with the selected object based on color moment feature values. For key frames from different cameras, background information of the related cameras is already stored in the database. It supports extracting the foreground region from an image and comparing the objects with the selected object.

For the comparison of feature values on the selected object and other objects on different frames, the system classifies the similar objects including key frames. These frames have relevant camera information and time information. This information can be seen in Table 1. The object tracking region information is sorted based on time. The information in the table can help to extract the object tracking region on place.

TABLE 1. Key information frame and user selected object

Camera	Key information frames	User selected key object	Segment on RGB
GA-2030			
GF- 1020			

To compare object similarity values, the system uses this moment based on RGB color. Color Moment Mean value is

$$\mu_i = \frac{1}{N} \sum_{i=1}^{n} P_{ij} \tag{1}$$

Color Moment Standard Deviation is

$$\sigma_i = \left(\frac{1}{N} \sum_{i=1}^{n} (P_{ij} - \mu_i)^2 \right)^{1/2} \tag{2}$$

Color Moment Skewness is

$$S_i = \left(\frac{1}{N} \sum_{i=1}^{n} (P_{ij} - \mu_i)^3 \right)^{1/3} \tag{3}$$

P_{ij} = value image address $i \times j$.

3.4. **Extraction of time information on frames.** To track object movement, the time information from key frames is very important section. It can support the object movement path based on time and the system decides the tracking region based on the time schedule. To extract time information from the key information frame, the system uses character segmentation and character recognition technique to segment date and time information and classify the characters using genetic algorithm and dynamic image processing techniques.

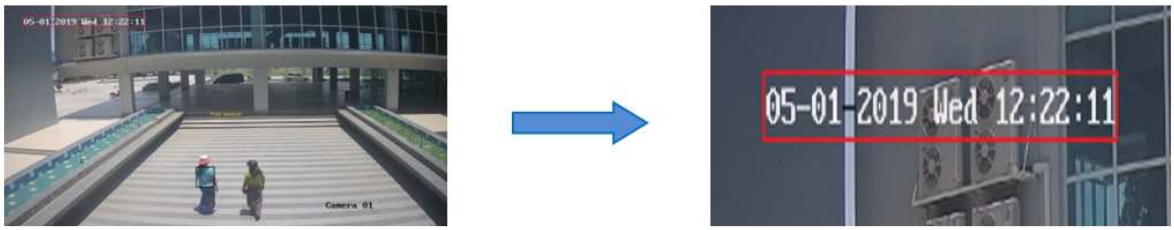
The classified information is stored in a database as per their frame information. The system finds the time difference value on object similarity frames information and calculates the object motion priority on different cameras.

Table 2. Frame information and relevant time information

Point No.	Extracted video time						Camera information	Point and place
	Year	Month	Day	Hours	Min	Sec		
1	2019	3	15	12	26	06	GA-2030	Gate point In
2	2019	3	15	01	30	34	GA-2076	Gate point Out
3	2019	3	15	12	45	12	GF-3065	Front of B
4	2019	3	15	12	56	23	GF-1020	Mid of public area
5	2019	3	15	01	20	05	GF-2451	End of Public area
...	...	...	...	...	...	...	...	...

3.5. **Finding track region based on time.** The system extracts the camera information from the camera network based on the object similarity rate between the important object and other objects from other key frames. The fining object is to accept only the smallest differences in similarity between the objects. All key frames from different video cameras are stored with the relevant information such as the camera's location and time value of a record. To store time values in the storage, the system needs to use image processing techniques.

Time information character is included in the relevant frame. To define the time information, the character image needs to process character segmentation and recognition algorithm and this information becomes real-time information for that key frame. The system selects the character region and transforms that region from a gray image to a binary image. By using Vertical Projection Profile (VPP) algorithm, characters are separated as single characters for analysis and transformed to time value [11].



Date Information: June 1, 2019 ***Time Information:*** 12 hours 22 minutes 11 seconds

Figure 2. Time information from frame

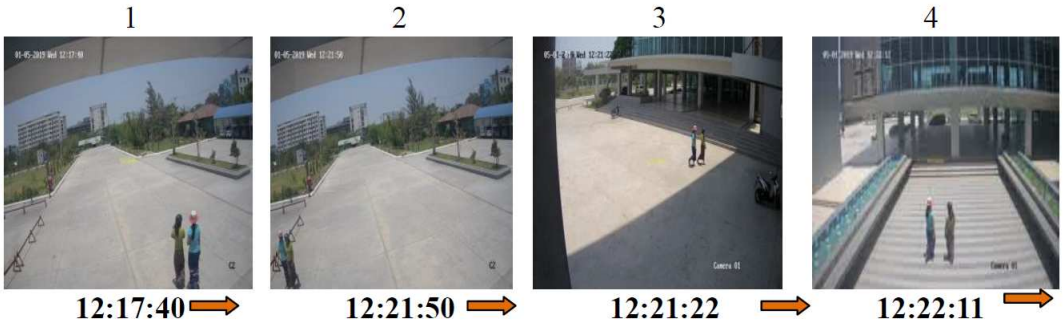
The time information extraction from key frames is based on OCR method and that information is very important to define object tracking region. This information is shown in Table 3.

Table 3. Similarity rate and camera information

Key frame time (H:M:S)	Camera information	Similarity rate	Passed parity
12:22:11	GA-2030	20.6583%	4
12:17:40	GF-3065	49.253%	1
12:21:50	GF-1020	52.34%	2
12:21:22	GF-2451	35.43%	3
01:30:34	GA-2076	2.4356%	0
...	...	...	...

4. **Expected Results.** The system extracts the results of the object passed path information based on a time schedule. The system shows the important object information and the object route on a camera network map. This basic information is shown in Table 4 based on the object movement schedule.

Table 4. Object tracking region



The important objects passed information is found based on their key frames time sequence. The system gives the object passed areas as object tracking regions and users can see if the object passed the area and the time of passing. This supports security systems, other traffic control systems and to solve criminal cases in public area.

5. **Conclusions.** This system can support many different sectors for the security and big data control systems. The color moment feature exactly supports to compare object similarity. This research has two parts, i.e., image handling and character image reorganization. When completing this research, the results supported moving object extraction from different background areas and the color object segmentation and character image analysis process.

Acknowledgment. This research is supported by the Myanmar Institute of Information Technology, Mandalay, Myanmar and they have allowed us to test this research with private datasets from the university. Professors and the entire faculty of the Engineering Department of Miyazaki University, Japan, supported the analysis and testing of this research. We express our sincere thanks to all of them for their support to this research.

REFERENCES

[1] Y. A. Alsultanny, Color image segmentation to the RGB and HSI model based on region growing algorithm, *Proc. of the 4th WSEAS International Conference on Computer Engineering and Applications*, pp.63-68, 2010.

[2] X. Bai and G. Sapiro, A geodesic framework for fast interactive image and video segmentation and matting, *IEEE the 11th International Conference on Computer Vision*, pp.1-8, 2007.

[3] H. D. Cheng, X. H. Jiang, Y. Sun and J. Wang, Color image segmentation: Advances and prospects, *Pattern Recognition*, vol.34, no.12, pp.2259-2281, 2001.

[4] C. Mythili and V. Kavitha, Color image segmentation using ERKFCM, *International Journal of Computer Applications*, vol.41, no.20, pp.21-28, 2012.

[5] B. Rouai-Abidi, S. Kang and M. Abidi, A fully automated active shape model for segmentation and tracking of unknown objects in a cluttered environment, in *Advances in Image and Video Segmentation*, Y.-J. Zhang (ed.), IGI Global, 2006.

[6] X. Wang, R. Hänsch, L. Ma and O. Hellwich, Comparison of different color spaces for image segmentation using graph-cut, *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol.1, pp.301-308, 2014.

[7] T. T. Zin, P. Tin, H. Hama and T. Toriu, An integrated framework for detecting suspicious behaviors in video surveillance, *Proc. of SPIE 9026, Video Surveillance and Transportation Imaging Applications*, 2014.

[8] D. W. Wicaksono and B. Setiyono, Speed estimation on moving vehicle based on digital image processing, *International Journal of Computing Science and Applied Mathematics*, vol.3, no.1, pp.21-26, 2017.

[9] Z. Huang, R. Wang, S. Shan, L. Van Gool and X. Chen, Cross Euclidean-to-Riemannian metric learning with application to face recognition from video, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.40, no.12, pp.2827-2840, 2018.

[10] M. M. Tin, N. N. Myo and M. M. Khin, Key information retrieval system by using diagonal block based method, *Proc. of the International Conference on Information and Knowledge Engineering (IKE)*, pp.104-107, 2017.

[11] K. Koo, J. P. Yun, S. H. Choi, J. H. Choi, D. C. Choi and S. W. Kim, Character segmentation and recognition algorithm of text region in steel images, *Proc. of the 8th WSEAS International Conference on Signal Processing, Robotics and Automation*, pp.293-298, 2009.