JDCS 디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 19, No. 4, pp. 757-770 Apr. 2018

선택적 히스토그램 빈 기반 열화상 영상 전경 추출

유 광 현[1] · 자 히 르[2] · 김 진 영[1*] · 신 도 성[3]
[1]전남대학교 대학원 전자컴퓨터공학부
[2]과학기술연합대학원 컴퓨터 소프트웨어학과
[3]디펜스텍 주식회사

Foreground Extraction in Thermal Videos Based on Selective Histogram Bins

Gwang-Hyun Yu[1] · Muhammd Zaigham Zaheer[2] · Jin-Young Kim[1*] · Do-Seong Sin[3]
[1]Department of Electronics and Computer Engineering, Chonnam National University, Gwangju, 61186, Korea.
[2]Department of Computer Software from the University of Science and Technology (UST), Korea.
[3]DefenseTech, Yong Bong Dong, Buk-Gu, Gwangju, Korea.

[요　　약]

열화상 영상기반 감시 시스템에서 전경추출은 매우 중요한 단계이다. 전경추출단계는 계산시간과 메모리 사용측면에서 시스템의 실시간 처리가 매우 효율적이어야 한다. 그러나 이러한 효율성은 ROI 탐지의 정확도와 매우 연관되어 있다. 본 논문에서 열화상 비디오 처리를 위하여 새로운 히스토그램 빈에 기반하여 배경과 전경을 분리하기 위한 두 가지 방법을 제시하는데, 이는 임의의 주어진 환경에서 열화상영상의 시간상에서 일관성을 갖는 다는 점과, 이러한 성질이, 간단한 시간축 메디안 필터링에 비하여 80%이상의 메모리를 절감할 수 있다.

[Abstract]

Foreground extraction is the most significant step in thermal imaging based surveillance systems. This step needs to be efficient in terms of time and memory consumption in order for the system to provide real time results but usually this efficiency reciprocates with the accurateness of the ROI detection. In this study, novel selective histogram bins based two background & foreground separation approaches for thermal videos processing have been proposed which exploit the temporal-consistency property of the thermal images in a given environment and can save over 80% memory than their simplest counterpart temporal median filtering.

색인어 : 전경추출, 열화상영상, 히스토그램빈, 감시시스템

Key word : Foreground extraction, Thermal videos, Histogram bins, Surveillance systems

http://dx.doi.org/10.9728/dcs.2018.19.4.757

Received 23 September 2018; **Revised** 22 October 2018
Accepted 05 November 2018

***Corresponding Author; Jin Young Kim**

Tel: +82-62-530-1757
E-mail: beyondi@jnu.ac.kr

Ⅰ. Introduction

Human detection and localization is a vital procedure for many camera based safety and security applications such as search and rescue, surveillance, reconnaissance or driver assistance. It can be applied for abnormal behavior or unwanted event detection, human gait characterization, living being detection, head counting in dense crowd, person identification, gender characterization etc. Various kinds of imaging systems are being used nowadays for numerous specific applications. Among those is the infrared thermography or thermal imaging, which is a significant technique used widely to obtain heat signatures of an environment. Since, infrared radiation is emitted from every object (living or non-living) having a temperature above zero, the thermal imaging can obtain situation of an environment without visible illumination. Until recent times, the use of thermal infrared cameras was limited to only military applications, helping operators to better see their environment and to detect humans and vehicles in very low light conditions. Driven by the emergence of new types of thermal cameras, more affordable, infrared imaging technology is starting to reach industrial, commercial and consumer markets.

With this increase in the variety of thermal-images capturing instruments and their possible applications, various researchers have proposed techniques to best utilize these devices with novel work on object identification or tracking. These approaches to process the thermal videos for extraction and classification of the useful information can be divided into three major categories based on:

Temperature threshold based object classification
Direct texture analysis of the whole frame
Background / Foreground separation

Temperature threshold works in the environments where a significant temperature difference between background and foreground objects exists. Hence, an optimal threshold value is determined which can separate out foreground from background. Wai Kit Wong et. el. in [1] proposed a scenario dependent threshold which is defined by the maximum and minimum of the heat range in which a human can lie. This value is then used to figure out intrusion in a trespass monitoring system. Michael Teutsch et. el. proposed to employ Maximally Stable Extremal Regions (MSER) for hotspot detection [2]. MSERs are the result of a blob detection method based on thresholding and connected component labeling [3]. In [4], K. Senthil Kumar et. el. suggested various ideas to determine optimal threshold for foreground segmentation. Histogram analysis is finally suggested for optimal threshold calculation. Threshold based background subtraction methods are although fast but mostly situation dependent. These methods lack robustness as parameters are predefined according to dataset properties and a widely speculated assumption of human body always possessing higher temperature than the surrounding environment also needs to be retained.

Direct texture analysis is one of the most commonly adopted methods to extract useful information from the thermal videos. In [5], Atousa Torabi et. el. used a mix of thermal and visual videos to construct a human blob outline. The image registration technique and finally the fusion of both videos provides a good scheme to detect and track the object. Authors in [6] suggests a thermal range based approach which takes into account the shape of the object as well. An environment is modelled as a collection of 3D humans and boxes and projected on a 2D scale to create various possibilities of human appearing behind common occlusions. Evolutionary algorithm is further used to classify a human. In [7] Wang W. et. el. suggested not to use pre-existing feature extraction methods like Haar, covariance and histogram oriented gradient (HOG) for numerous reasons like small target size, insufficient texture information and variation under different weather conditions. Instead, the idea of using shape context descriptor based adaboost cascaded classification is proposed for human detection. Authors in [8] discuss a possibility to construct a robust pedestrian representation model by using intensity distance projection space. Authors in [9] targets the area of detecting, classifying and tracking object which may be found on ocean surface. A Gaussian filter is followed by a prewit operator based edge enhancing technique. Then, a threshold is applied to subtract the ripple effect out of the image and highlight the object. Finally, classification is performed by the method proposed in [10]. Direct texture and shape analysis is a good choice for most applications. The methods formulated are generally robust and can work in any environmental condition but require clearly visible human blob of adequate size. Moreover, high computation power is required to execute complex image processing algorithms.

Because of such drawbacks, background models based methods are also employed as an alternative. In this kind of techniques, a background image is modelled and then the frames are subtracted from this background to separate out the foreground. Umesh Gupta et.el. [11], with main focus on single or multiple target tracking, suggested to use Single Frame Differencing (SRF) [12], Running Average (RA) and Temporal Median Filter (TMF) [13]. Ju Han et el built Gaussian model by using sufficient number of frames and then foreground is estimated based on this model [14]. Eun Som Jeon Et.el. in [15] adopted a two-step method to generate a background image. Based on the assumption that human body always appears greyish than the background, first a background image is estimated from a set of frames by using Median filtering. Another average image is then generated from the set of frames not belonging to the previous set. Morphological operation is applied on both images and background image is generated by finally removing the foreground candidates. Guang Han et.el. recommended not to use adjacent frame differencing method [16] for background subtraction as it may bring holes in the foreground region. Instead authors presented a pixel-wise median technique with a predefined threshold to decide if the values belongs to foreground or not [17]. An extensive study about Pedestrian Tracking by Thermal Infrared Imaging is presented in [18]. The tracking based on fused thermal and visual video produced somewhat adequate results whereas the findings shows that only thermal video on its own is not efficient enough to be used with current algorithms for the said purpose. A contour extraction based methodology is presented in [19, 20] where region of interest (ROI) is first calculated by using a pixel-wise statistical background subtraction model, whereas foreground pixels are identified by using square of Mahalanobis distance. Then each ROI is examined separately to extract person silhouette from the thermal halos. Various methodologies are further used to enhance the quality of this region. As the method is based on contour instead of color range based intensity level or texture, it is robust and more stable across various environments. Presented in [21] is a technique to detect moving objects by a thermal camera which is mounted on a vehicle. First the thermal images were converted into greyscale images by the method of contrast stretching. Warping was then used to identify relative motion of objects between stationary objects which is caused by moving vehicle. A threshold was then introduced over difference between warped and sample images to extract the foreground. Other enhancement techniques were then used to improve the quality of foreground object. Bin Qi et.el. suggested to use sparse representation to extract the foreground / background information. The method initially requires a very large amount of training data in order to successfully extract the information [22]. Most of these background modeling techniques based systems are not constrained by the environment but usually consumes higher memory or computation power hence, limit their applicability in mobile devices.

To overcome these drawbacks, we present herein a new histogram bins based representation of the thermal images which can be used to extract regions of interest from the thermal videos. The proposed research is significant in the following aspects:

The suggested methodology is robust to weather variations and the resulting intensity changes in the foreground images.

Proposed histogram bins can be used to calculate the temporal median value with extreme memory saving.

Second technique, which adopts a slight variation in the usage of histogram bins than calculating temporal median, not only overcomes the limitations of median filtering but also consumes almost 80 % less memory than basic temporal median filtering.

The robustness of the methodology is intensively examined on various open datasets as well as another more challenging self-recorded dataset.

Background estimation in such systems is a significant step as any error at this point can propagate throughout the process. Successful background estimation leads to a clearly separated foreground on which classification can then be performed. An error occurred in the foreground separation step (especially a miss of the human blob) cannot be rectified in the classification part.

The technique involves background model calculation hence it is suitable for the cases of stable cameras which is usually common in surveillance or intrusion detection systems. The main contribution of this research work is that it can separate the foreground from the background with extreme accuracy while consuming minimum memory.

The rest of this article is organized as follows: Section 2 explains the datasets used in this experiments. Section 3 briefly reviews the existing background subtraction technique and their pros and cons whereas section 4 describes the proposed methodologies in details. Section 5 presents a brief overview of thresholding and morphology techniques used to enhance the foreground images. Finally, paper is concluded with results and discussions in Section 6.

Ⅱ. Datasets

To validate the methodology various publically available thermal datasets [23]-[26] have been used in the experiments. These datasets consist of recordings in numerous environments with single and multiple pedestrians whereas various small occlusions like trees poles are also involved. In addition, to augment more challenges for the methodology, a self-recorded dataset has also been experimented.

2-1 Our Dataset

As thermal cameras take heat signatures from an environment, the videos obtained may vary a lot at different hours. The results may depend on weather conditions, day / night times, heat emitting objects inside the frame, background environment etc. With each new case, a human blob appears differently in the

(a) (b)

그림 1. (a) 녹화 세팅; (b) Argo S 모델 열화상 카메라

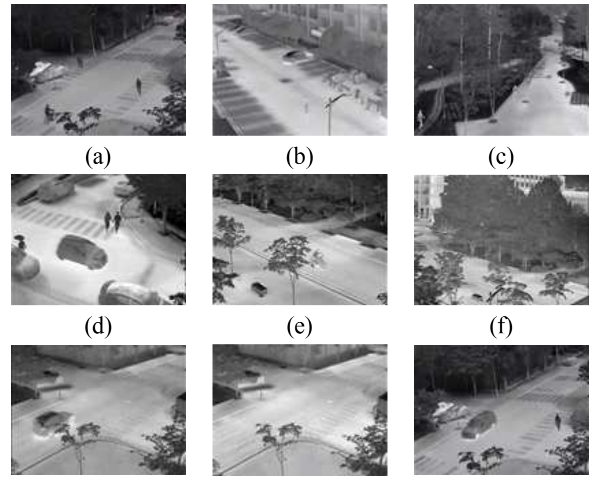
Fig. 1. (a) Recording setup; (b) Argo S Thermal Imaging Camera

thermal videos.

Our dataset was recorded within the campus area of Chonnam National University, Gwangju, South Korea, using Argo S Thermal Imaging Camera which captures 640 x 480 sized videos at the frame rate of 31 per second. The recordings were done in summer (July, August and September) when the temperature on average ranges between 18°C – 30°C. A total of 300 minutes in 12 footages were captured at five different locations at different times of the day involving numerous kinds of occlusions and intrusions. The recording setup is shown in Figure 1.

The first and the foremost challenge introduced in the dataset is related to a large temperature variation between day and night. In Gwangju, during summer, the temperature can get as low as 14°C in the night and as high as 34°C during a hot day. As a result, in our dataset, human blob also varies between black during the daytime to white during a cool night, compared with the background environment. Second, the recording is done in environments containing different background materials e.g. concrete, stones, trees, grass or buildings. These materials react differently when exposed to sun, absorb varied levels of heat during the day time and finally emit this heat differently which results in a lot of disparity in an overall thermal image. (Figure 2)

Third, varied recording locations are chosen to introduce several occlusions in the frame. These occlusions are categorized into two types: passive occlusions and active occlusions. Passive occlusions are those which may block view but do not affect the overall intensity level of the environment e.g. trees or shades etc. Although, due to strong winds, these trees show motion which can result in false movement detection in the foreground. In contrary, active occlusions are those which are heat sources themselves e.g. poles (Figure 2 (b) & (c)). One another challenge in this dataset is that all five locations involve roadside view hence, vehicles, which are very strong hotspots, enter and exit the frames frequently. These extreme hotspots not only disturb the overall intensity levels in the frame but can also overshadow the human blobs. This scenario could be very challenging for the temporal background update process as the sudden intensity



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

그림 2. 데이터셋 내의 다양한 장소에서의 이미지. 부분 밝기 변화, 다양한 흡장 및 달리고 있는 자동차를 볼 수 있음

Fig. 2. Scenes from various locations in the dataset. Blob intensities variation, occlusions variety and running vehicles can be seen

variations may induce far lasting effects.

Ⅲ. Background Extraction

Background extraction is supposedly a continuous procedure which is performed on the incoming frames to estimate the background. Then the new frames are subtracted from this background to extract the foreground. Many researchers suggested to use temporal median or averaging filtering as these are simple to implement, fast and perform efficiently [12], [13], [15]. Various other complex algorithms are also suggested by some authors to estimate the background [1]-[4], [9], [17]-[21]. This research work is mostly focused on devising an efficient and memory-economical background extraction technique which can fulfil the challenges introduced in the datasets hence, complex techniques are not analyzed in the scope of this paper.

3-1 Temporal Median Filtering (TMF)

In this technique, the background is modeled by taking pixel wise median of intensities over a small window of frames. In thermal videos, background pixels are likely to retain their intensity values and do not change drastically over time. Hence median filtering can estimate the stable intensity values which presumably belong to the background. The window size is determined by a factor of how slow persons are moving within the frame as, an adequate number of pixels must be in the median calculation sequence to determine the background. In Figure 3, a temporal window of size 63 is visualized at pixel location (257, 86) out of which, first 20 frames in the sequence possess intensity value 110, next 20 frames 111, further 14 frames 230 and last 9 frames again possess 110. The most credible value for the background pixel is 110 in this case whereas median of this sequence gives 111, which can also bear almost identical results in the foreground extraction process.

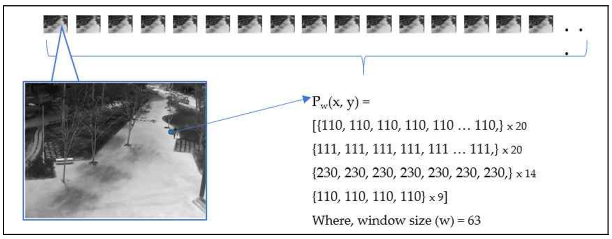


그림 3. 63프레임의 윈도우 위에 랜덤 포인트 p(x,y)를 표시하는 픽셀 스트림

Fig. 3. A pixel stream represented for a random point p(x,y) over a window of 63 frames

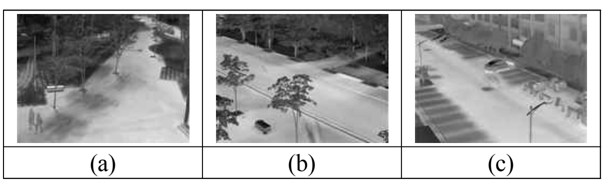


그림 4. 배경 이미지의 고스트/그림자 효과

Fig. 4. Ghost / Shadows in background images

3-2 Fast Temporal Median Filtering

The main idea behind fast median filtering is to avoid sorting on every loop. Instead, a separate array is used to keep the sorted elements whereas the main incoming stream array is used to update this sorted array. With each element coming in, the old one is deleted from the sorted sequence and the new one is put in its suitable position. This method speeds up the implementation but the results produced are identical with the normal TMF

3-3 Temporal Average Filtering (TAF)

This technique works on the same principle as of temporal median filtering for background calculation but instead of taking median values, an average of all values in the sequence is taken. But, averaging the values can only minimize the effect of an outlier and cannot fully eradicate it, which is a drawback of this technique. Taking again the case visualized in Figure 3, averaging of all the intensity values would give 137 as background pixel. This difference in value of actual background pixel and the estimated background pixel may generate shadows (Figure 4) which can lead to false region-of-interests alarms.

The background images produced are usually far less accurate than TMF as TAF leaves ghosts/shadows more often. Various researchers [15] suggested additional methodologies to improve TAF but then these techniques also require additional memory and processing-power consumption hence, are out of the scope of this research work.

Ⅳ. Histogram Bins Based Proposed Background Subtraction Method

4-1 Temporal Median Filtering (TMF)

Thermal images are different from optical images in various ways including grey scale / color representation, pixel intensity values at a given point, range of pixel values etc. In normal cases, thermal images are greyscale with 28 pixel intensity values. Moreover, overall background of the thermal videos is based on temperature signatures hence it does not change much over a small interval of time and so is the respective histogram, as shown in Figure 5. Hence, it can be inferred that for a stream of

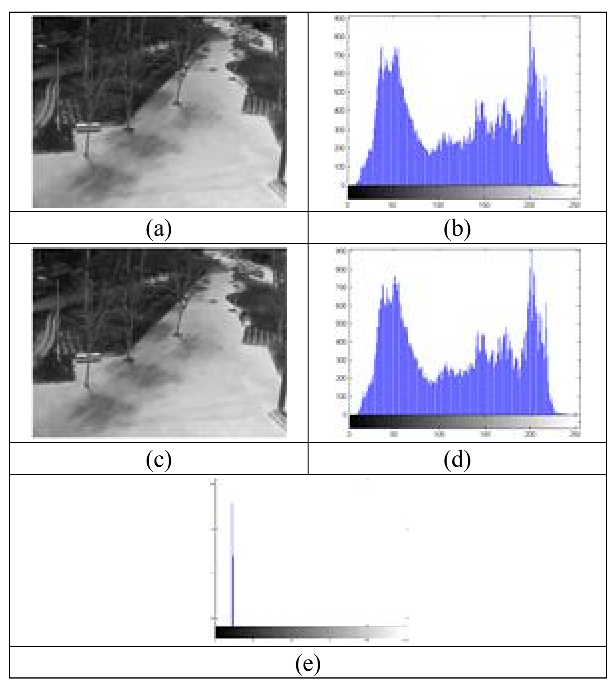


그림 5. (a)와 (c)는 데이터셋으로부터의 랜덤 프레임이다. (b)와 (d)는 그것들의 각각의 히스토그램이고; (e)는 한 픽셀 스트림의 히스토그램이다.

Fig. 5. (a) and (c) are random frames from the dataset and (b) and (d) are their respective histograms; (e) histogram of one pixel stream

intensity values at a pixel point, although possible values are from 0 to 255, the actual range of intensity value should be very small. Figure 5 (e), shows that the intensity values of pixel (20, 20) from our dataset over the time t to t+5 seconds was only 23 or 24. These intensity values, which we will term as Selective Histogram Bins, can be represented in the form of [Bin value, # of occurrence]. These bins can be manipulated in numerous ways to reduce memory consumption for pixel stream representation. In our dataset, an average of 2.64 bins were utilized per pixel to represent these intensity values over a window of 63.

1) Selective Sorted Histogram Bins Based Median Calculation

This method requires the Selective Histogram Bins to be stored in intensity-value wise sorted fashion and only these bins are then needed to identify the median value / background pixel intensity. Using values from Figure 3, selective histogram bins formation is shown in Figure 6. Ultimately, the bins representation of the form [Bin value, # of occurrence], would become { [111,20], [230,14], [110,29] } whereas, the selective sorted histogram bins would be { [110,29], [111,20], [230,14] }. In order to get the median value, number of occurrences are counted up-to w/2 value rounded off to integer. In our case, w/2 would be equal to 32 with which corresponding occurrence lies in the bin for intensity value 111. Hence, 111 is the median value of the whole sequence and consequently the intensity value of background pixel.

Once value of the background pixel is extracted, two subsequent operations are performed on the bins to temporally proceed the system: incorporate new pixel from the pixel stream and delete one oldest value. But unlike a circular array, these bins do not possess any temporal property. In order to assist the bins to track changes, supporting time-stamp methods are required which may track the pixel occurrence time. These methods are discussed in the subsequent part of this section.

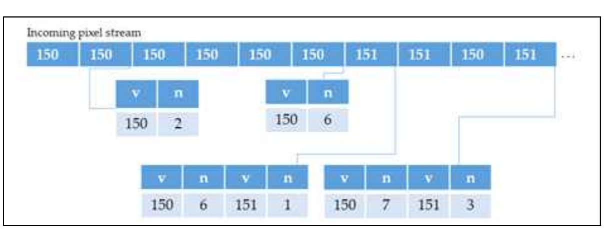


그림 6. 도달하는 픽셀스트림의 선택적 빈 형성

Fig. 6. Selective bins formation from the incoming stream of pixels

2) Previous Occurrence based Upcoming Background Pixel

Median value based estimation for background pixel is efficient in most scenarios but it is prone to error in case foreground object is moving slowly or stay still for a short time. This is because the median calculation only takes central value in the whole sequence and if pixel stream contains over fifty percent values belonging to the foreground, results would be erroneous. For example, a sequence from 63 frames from our dataset forms the selective histogram bins as:

$$Sequence1 = ([97,14],[99,15],[138,4],[155,2],[193,4],[219,14]) \quad (1)$$

The value belonging to the background here is actually 97 or 99 but taking median of this sequence would give 138 which is far away from the actual value(s). As a result, it can leave holes in the ROI blob or add noise to the background image. However, this limitation can be overcome using the occurrence information in histogram bins. Usually it is obvious that an object moving within the background frame must not appear excessively than the background itself. Hence, the most occurring pixel intensity can be selected for the background pixel instead of median value. In example sequence shown in Figure 3, this proposed technique would give 110 whereas in Sequence 1 (Equation 1), it would give 99 as the background pixel value which is more accurate than the median values of same sequences as 111 and 138 respectively. In this proposed technique, bins are not required to be sorted although the time-stamp methods are still indispensable to perform temporal updates.

4-2 Time-stamp

Time-stamp methods are required to keep the bins updated with new pixel values and delete the outdated values; complementary (but compulsory) tasks with the histogram bins based background calculation methods. Time-stamping ensures that none of the values exceeds the limit of maximum window size, goes out eventually after this specific length of the window and hence keep latest incoming values for the background update process.

1) Bit Wise Time-stamp

Occurrence of a value in a bin cannot exceed the window size. With this limit predefined, a binary sequence can be used to mark each change with a 1 or 0, depending on whether the respective bin was updated or not. Another random sequence from the dataset forms the histogram bins as:

$$Sequence1 = ([110,20],[111,18],[110,5],[230,7],[231,10],[111,4]) \quad (2)$$

The window size is set as 63 hence, each bin can have a maximum of 64 elements where the 64th element is outgoing. A

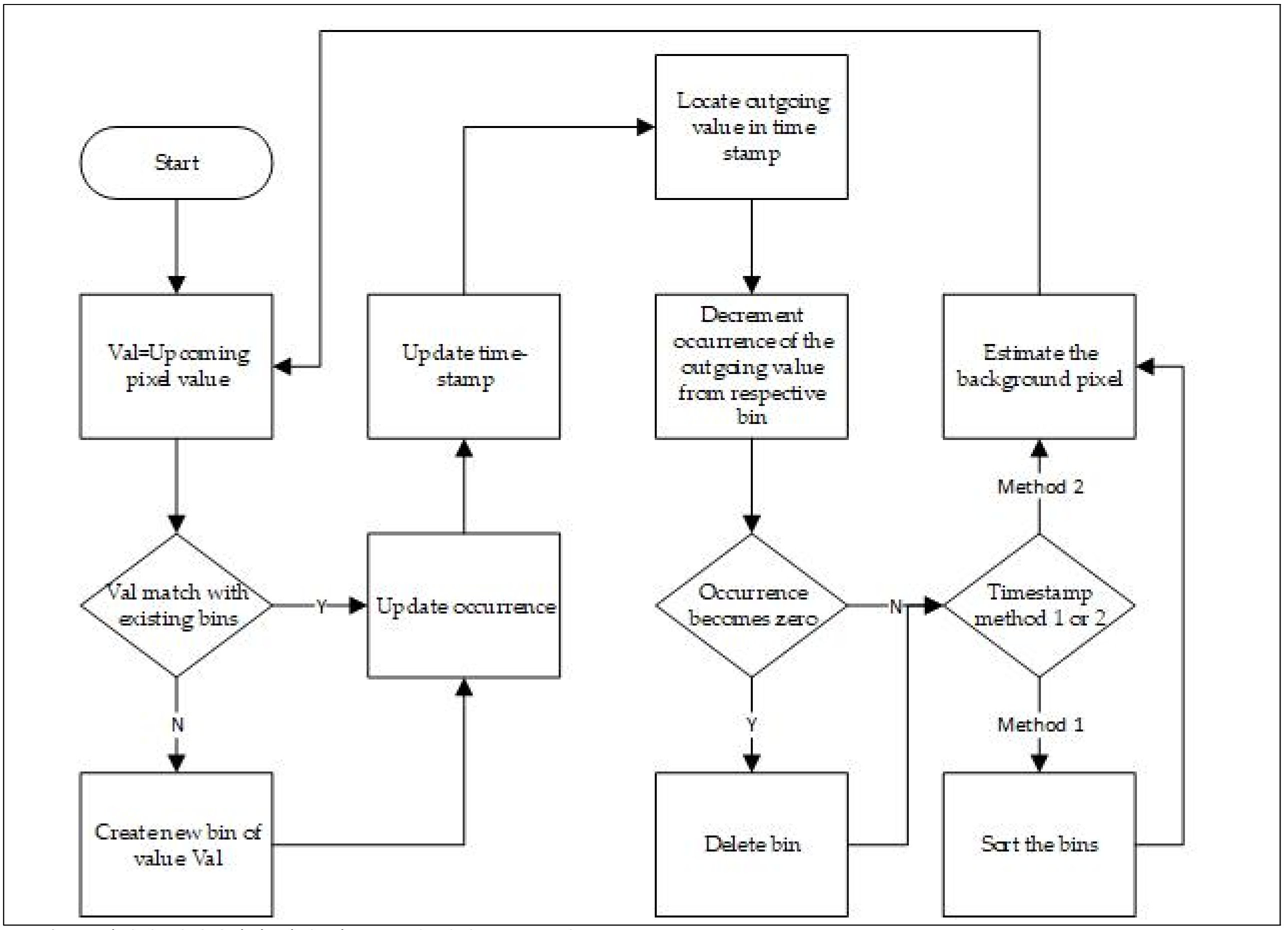


그림 7. 제안된 배경차감법 기반 히스토그램 빈의 플로우 차트

Fig. 7. Flow chart of the histogram bins based proposed background subtraction method

separate 8 bytes (64 bits) binary number is assigned to every bin, when the bin is formed and updated with every new incoming pixel value. Initial filling of the bins and their respective time stamps is shown in Figure 8 (a) whereas, once the whole stream window is incorporated, deletion of the initial value is shown in Figure 8 (b). Whenever a value is updated in a bin, its corresponding timestamp bit is marked as one, zero otherwise. After each incoming value, all the time-stamps are bitwise shifted to the left and once a 1 in any time-stamp reaches at the left most position, one occurrence value is decremented from that bin. If the occurrence becomes zero at any point, whole bin is deleted to conserve memory.

2) Bin Change based Time Tracking

This time-stamp method also relies on the fact that the change in pixel values of thermal images over a short time is not significant. As the target of a time-stamp method is to implement the queue concept i.e. pop the first element when a new element comes in. This can be done by keeping the record of changes in the pixel values among incoming stream as in thermal videos a pixel value which belongs to background usually remains same or change very little in an incoming stream of frames (Figure 5). In our dataset, on average, 5.37 changes were observed per pixel for a window of 63 frames. The proposed technique works by assigning a new lifetime value to the respective bin at every change between two consecutive values in the input stream. No change in the intensity values means no new lifetime number. This lifetime value, which needs to be equal to the window length, determines which value is going out from which bin and at what time. Figure 9 (a) and (b) shows initial fill-up of bins, lifetime numbers and then deletion of the expired values.

Ⅴ. Thresholding and Morphology

5-1 Tresholding

Once the background is subtracted (Figure 10 (b)), in order to enhance the foreground blob, a binary threshold is applied to the pixels, the value of which, in our experiments, is determined

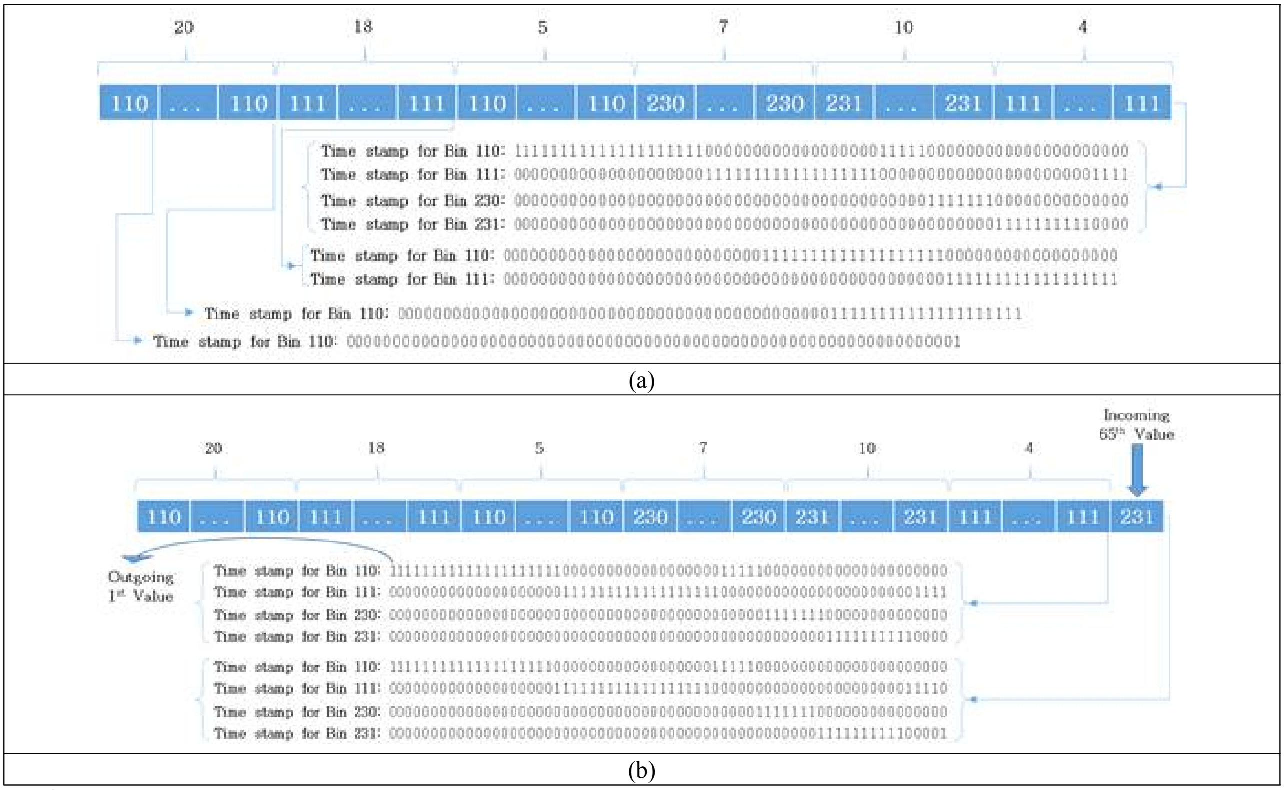


그림 8. 비트연산 타임스탬프 메서드: (a) 빈이 최대 윈도우 크기에 도달하기 전에 채워지는 타임스탬프 비트 업데이트; (b) 윈도우 사이즈가 초과되어 제한된 만료값의 삭제

Fig. 8. Bit Wise Time-stamp method: (a) Update of the timestamp bits as the bins are filling in before the max window size is reached; (b) Deletion of the outdated value as the window size exceeds its limit.

using the histogram of the foreground image. All the non-zero histogram values are counted and then upper two third of these values are set as 1 while the others as 0 (Figure 10 (c)). The reason why only non-zero values are used is that with changed weather and environment, the intensity of foreground blob can vary a lot and a fixed threshold between 0-255 cannot work. This proposed technique works well until it encounters with a situation where there is no foreground object in the frame. In such case, noise is enhanced as foreground object and hence show false ROIs. To eradicate this problem, the threshold determination method is assisted by a temporal array which stores white pixels count of previous few foreground frames. As, the environment signatures doesn't change abruptly in the thermal videos, white pixel count is supposed to be stable in the consecutive frames. But when a case like 'no foreground object in the frame' happens, two-third of the non-zero histogram bins generates a gigantic amount of white pixels in foreground image. Hence, the temporal white-pixel-count array is then used to determine a sensible threshold value which should give similar number of white pixels in the foreground as it were in previous few frames.

5-2 Morphology

To further culminate the quality of foreground object blob, remove noise and connect broken blobs, a series of morphology operations are applied to the images. First of all, isolated pixels are removed from the whole image (Figure 10 (d)). Then dilation is applied which is followed by erosion to remove thin lines-like shapes generated due to noise (Figure 10 (e)). Afterwards, small components comprising less than 10 connected pixels are deleted from the image which helps removing all small noise components and leaves the foreground with only strong candidates for the ROIs blobs (Figure 10 (f)). Finally, dilation is again applied to connect the broken blobs of an object (Figure 10 (g)). This way all the ROI candidates with a distance of less than 10 pixels in between are combined to form one shape.

Ⅵ. Results & Discussions

Evaluation of the median techniques as well as the previous occurrence based upcoming background pixel method has been carried out on the dataset. All results have been analyzed subjectively (Figure 11) for the accurate region of interest (ROI)

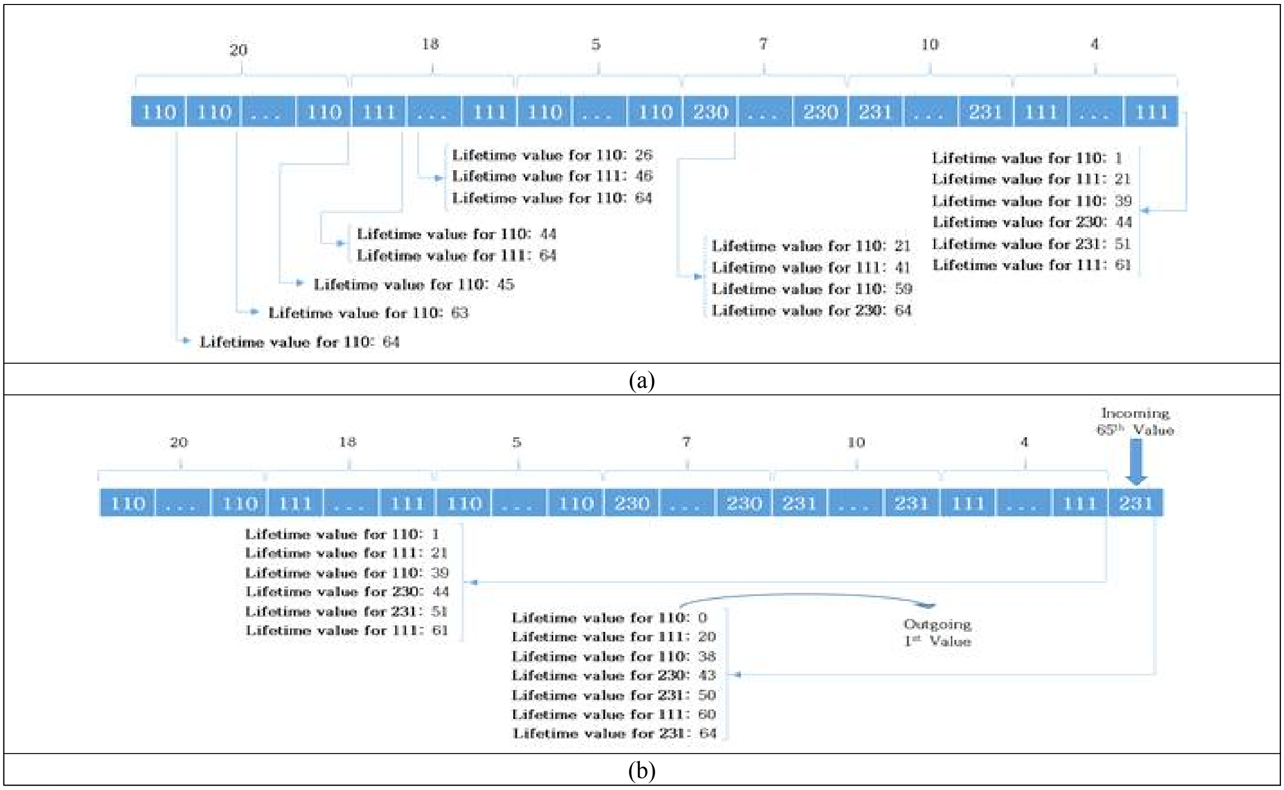


그림 9. 빈 변화에 의한 시간 추적 메서드: (a) 빈이 최대 윈도우 크기에 도달하기 전에 채워지는 라이프타임 값 업데이트; (b) 윈도우 사이즈가 초과되어 제한된 만료값의 삭제

Fig. 9. Bin change based time tracking method: (a) Update of the lifetime values as the bins are filling in before the max window size is reached; (b) Deletion of the outdated value as the window size exceeds its limit.

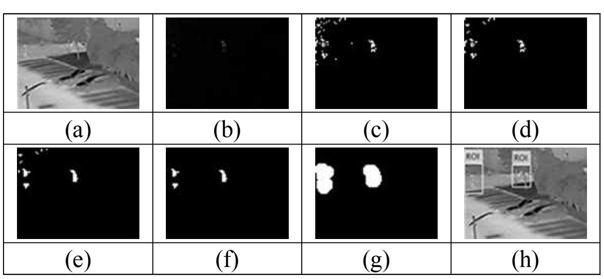


그림 10. (a) 열화상 이미지; (b) 배경 추출; (c) 바이너리 스레시홀드 처리; (d) 모든 고립된 픽셀을 제거한 깨끗한 이미지; (e) 침식에 의한 확장 적용 후 획득 이미지: (f) 10픽셀 미만으로만 된 구송 요소 제거; (g) 손상된 ROI 부분을 결합하는 확장; (h) ROI 경계 표시

Fig. 10. (a) Thermal image; (b) extracted background; (c) after binary threshold; (d) cleaned up by deleting all isolated pixels; (e) image obtained after applying dilation followed by erosion; (f) removal of components comprising less than 10 pixels; (g) dilation to combine broken ROI parts; (h) marked ROI boundary.

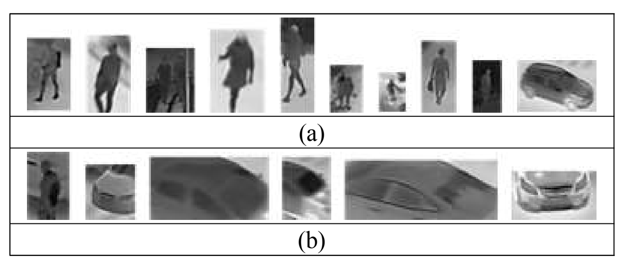


그림 11. 제안된 알고리즘에 의한 ROI 표시된 이미지: (a) 올바르게 표시된 ROI; (b) 일부분만 표시된 ROI

Fig. 11. ROIs marked by the proposed methodology: (a) correctly marked; (b) partially marked

extraction. While assessing the results, emphasis was made on the fact that the foreground objects are not missed out in the background separation or enhancement. Secondly, it was also kept in consideration to detect as less as possible false ROIs, to keep the burden low on classification part.

The second technique (Previous Occurrence based Upcoming Background Pixel) proposed in our paper has shown superiority over temporal median filtering in a way that it is more tolerant to slow moving objects. It can eradicate most of the ghosts or shadows from the background. This way, the ROI boundary gets more accurate with less number of false alarms (Figure 12 (d) and (h)).

In our dataset of 12 clips at five different locations, more than

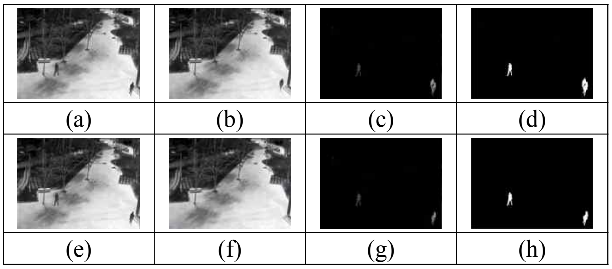


그림 12. 제안된 (이전에 존재되어진 픽셀 기반의 다가오는 배경 픽셀) 것은 천천히 움직이는 대상에 대해 더 좋다: (a)와 (e)는 객체가 확실하게 천천히 움직이는 열화상 영상; (b) 시간 중간값 필터에 의해 획득된 배경의 고스트 이미지; (c)와 (d)는 각각 전경의 에러; (f) 제안된 알고리즘에 의한 배경 추정; (g)와 (h)는 각각 전경 이미지

Fig. 12. Proposed (previous occurrence based upcoming background pixel) is more tolerant to slow moving subjects: (a) and (e) thermal video case in which object moves noticeable slow; (b) ghost image in the background obtained by temporal median filtering; (c) and (d) subsequently error in foreground; (f) background estimated by the proposed technique; (g) and (h) subsequent foreground images.

1050 humans, 70 motorbikes, 110 bicycles and 750 other vehicles were annotated subjectively. Out of these, 8 humans and 68 other vehicles were partially marked (Figure 11) whereas only 4 humans and 4 cars went undetected. Rest of the subjects were marked as ROI successfully using the proposed methodology. Some of the cases can be seen in Figure 13. Whereas in the other datasets [23-26], results obtained were also highly encouraging with successful extraction of almost all the ROIs, as shown in Figure 14.

6-1 Memory Consumption

The memory consumption comparison of the techniques is only carried out for pixel stream representation. Rest of the factors are purely dependent on programming environments and varies from application to application hence, are not considered a part of this comparison.

For simple median filtering, we need one circular ring array and a methodology to sort the arrays. This sorting methodology adds additional burden to the memory but can be freed once the sorting is done and median is evaluated. Hence the major memory consumption would be by one array representing the pixel stream, which in case of a 63 sized window would be 67 bytes (4 bytes for pointer). For the fast-smart implementation of the same technique, two circular ring based arrays are used among which, one array would store the temporal information of when a pixel came in the window and the other array would keep the median information to continuously update the background. Hence, overall memory consumption per pixel for a 63 size window would be: ((window size) x 2) = 63 x 2 = 126 bytes. Adding the pointer of size 4 bytes makes it 130 bytes. Temporal averaging can be performed on any array which naturally can retain temporal property hence, one ring array is enough to represent the whole pixel stream. Overall memory consumption per pixel to calculate is equal to the window size. Hence for a 63 size window, a total of 67 bytes would be required (pointer included). Although this technique is memory efficient as compared to TMF but it leave ghosts and shadows in the background image. Whereas in our proposed selective histogram bins based methods, the memory consumption decreases drastically.

그림 13. 제안된 알고리즘에 의한 데이터베이스 내의 ROI 표시 이미지

Fig. 13. ROIs marked in the dataset by our proposed technique of previous occurrence based upcoming background pixel.

To implement the proposed bit wise time-stamp with either selective sorted histogram bins based median calculation method or previous occurrence based upcoming background pixel method requires a structure containing bin values (v), occurrences (n) and time-stamp (idx) among which v and n requires 1 byte each in memory whereas for a window size of 63, idx requires an 8 byte number. Total bytes consumed to represent all the intensity values of one pixel can be represented by:

$$(nBins \times (idx + v + n)), \quad (3)$$

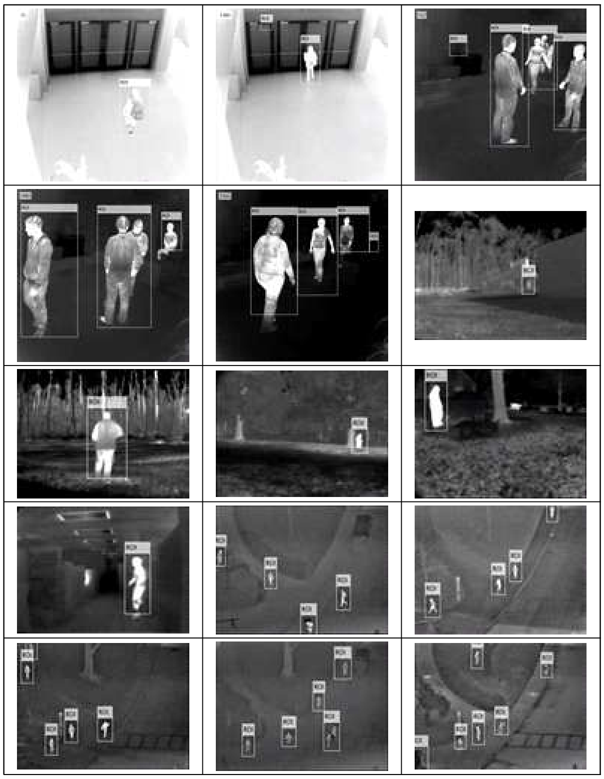


그림 14. 제안된 알고리즘을 적용한 공공데이터 셋의 ROI 표시 이미지

Fig. 14. ROIs marked in the open datasets by our proposed technique of previous occurrence based upcoming background pixel.

where nBins = Average number of bins required for the dataset. Adding the values to the equation, we get: (2.64 x (8 + 1 + 1)) = 26.4 bytes. To access values in structure based implementation, a pointer of 4 bytes is additionally required which elevates the total memory consumption to 30.4 bytes.

Whereas structure based implementation of the bin change based time tracking with either of the selective histogram bins based techniques, bin values (v), occurrences (n), life-time values (ltv), bin values corresponding to ltv (bv) and previous value in the sequence (prevV) are used. In this algorithm, bv and ltv are the arrays which store the intensity values in sequence of every change and their lifetime values respectively. Hence their usage depends on the average number of changes (nIntensityChanges) in the pixel stream. Whereas, prevV is only one number which keeps the last intensity value. Rest of the values are dependent on average number of bins. Hence, the memory consumed by this technique can be calculated as:

$$[(nIntensityChanges \times (bv + 1tv))] + (nBins \times (v+n) + prev\,V]bytes$$
$$[(5.37 \times (1+1)] + (2.64 \times (1+1)) + 1] = 17.02bytes \quad (4)$$

표 1. 각각의 알고리즘에 따른 하나의 픽셀 스트림을 표시한 바이트 소비

Table 1. Number of bytes consumed in each technique to represent one pixel steam.

Methodology	Average Per Pixel Memory Consumption (bytes)
Temporal Median Filtering	67
Temporal Fast Median Filtering	130
Bit wise time tracking with selective histogram bins based proposed methods	30.40
Bin Change Based Time Tracking with selective histogram bins based proposed methods	21.02

With additional 4 bytes for the pointer, this implementation for one pixel stream would cost 21.02 bytes in total.

All our experiments are conducted in Dell Optiplex 9020, with windows 07, 16GB RAM. All the codes are written in Matlab 2013a with mex-implementation. Table 1 compares the memory consumption of the given methodologies with traditional temporal median filtering implementation which shows that our proposed histogram based techniques can reduce the memory for a pixel stream representation to as low as 16.5 % as compared to its existing counterpart. It should also be noticed that the comparisons are only given for window size of 63. If this window size is increased, say doubled, the memory consumption of median filtering would elevate proportionally and get to 256 bytes per pixel. Whereas, our proposed methodology is based on intensity changes hence it cannot get effected much by the window size.

표 2. 모든 논의된 알고리즘들의 배경이미지 모델에 따른 시간 소비 비교

Table 2. Comparison between all discussed techniques for the time consumption to model a background image

Methodology	Average Per Frame Time Consumption (Seconds)
Temporal Median Filtering	0.0541
Temporal Fast Median Filtering	0.0492
Bit wise time tracking with selective histogram bins based proposed methods	0.0718
Bin Change Based Time Tracking with selective histogram bins based proposed methods	0.1681

In terms of time consumption, traditional temporal median filtering is faster than the proposed methods (Table 2). As, our techniques are implemented using structure based methodology and in mex-implementation all the structures are converted back and forth between Matlab and C++ on every iteration, it makes the system slower. In future, further experiments can be conducted to test the whole system in any low level language like C++. Currently to keep the system real-time using the proposed techniques, time dilation was used in which one out of every 10 frames was picked to model the background optimally.

Ⅶ. Conclusions

This research work attempted to devise a memory efficient methodology for effective background extraction in thermal images. In order to perform this task, selective histogram bins are introduced which can compact large series of thermal into small, memory efficient representations. This makes it possible to implement median filtering with minimal memory consumption. Furthermore, selective histogram bins also facilitates to make a new previous occurrence based technique which can overcome the limitations of median filtering. To assist these bins to add temporal property, two timestamping techniques are also introduced.

In order to evaluate the proposed methodology, a self-recorded dataset has been used along with the existing thermal imaging benchmarks. This dataset has been generated by keeping in mind a few hard cases for the thermal images which include: temperature variations, diverse background environments, involvement of extreme heat sources in the surveillance area, etc. Results of the experiments performed on these datasets showed that the proposed methodology is effective in separating out the foreground from thermal images and outperforms the existing methods in terms of memory consumption and accuracy. This paper also provided discussion on the morphological operations necessary to subtract out the noise and enhance the ROI blob to minimize false alarms.

These ROIs and their identification using binary or multiclass classifiers can be performed as a future work to utilize the methodology and examine the new dataset introduced in this paper.

Acknowledgments

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-R2718-16-0011) supervised by the IITP(Institute for Information & communications Technology Promotion)

References

[1] Wong, Wai Kit, Poi Ngee Tan, Chu Kiong Loo, and Way Soong Lim, "An effective surveillance system using thermal camera." in Proceeding of International Conference on Signal Acquisition and Processing, Malaysia, pp. 13-17, Apr 2009.

[2] Teutsch, Michael, Thomas Mueller, Marco Huber, and Juergen Beyerer, “Low resolution person detection with a moving thermal infrared camera by hot spot classification.” in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 209-216, Jun 2014.

[3] Donoser, Michael, and Horst Bischof, “Efficient maximally stable extremal region (MSER) tracking.” in proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York: NY, Vol. 1, pp. 553-560, Jun 2006.

[4] Kumar, K. Senthil, G. Kavitha, R. Subramanian, and G. Ramesh, "Visual and thermal image fusion for UAV based target tracking." MATLAB-A Ubiquitous Tool for the Practical Engineer. InTech, pp. 307-327, 2011.

[5] Torabi, Atousa, G. Massé, and G. A. Bilodeau, "Feedback scheme for thermal-visible video registration, sensor fusion, and people tracking." in Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 15-22, Jun 2010.

[6] Markov, Stefan, and Andreas Birk, "Detecting humans in 2d thermal images by generating 3d models." The Journal of Springer, pp. 293-307, Sep 2007.

[7] Wang, Weihong, Jian Zhang, and Chunhua Shen, "Improved human detection and classification in thermal images." in Proceeding of 17th IEEE International Conference on Image Processing, pp. 2313-2316, Sep 2010.

[8] Li, Jianfu, and Weiguo Gong, "Real Time Pedestrian Tracking using Thermal Infrared Imagery." The Journal of JCP, Vol. 5, no. 10, pp. 1606-1613, Oct 2010.

[9] Leira, Frederik S., Tor Arne Johansen, and Thor I. Fossen, "Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera." in Proceeding of IEEE on Aerospace Conference,Big Sky:

MT, pp. 1-10, Mar 2015.
[10] Hu, Ming-Kuei, "Visual pattern recognition by moment invariants." The Journal of IRE Transactions on Information Theory , Vol. 8, No. 2, pp. 179-187, Feb 1962.
[11] Gupta, Umesh, Maitreyee Dutta, and Mahesh Vadhavaniya, "Analysis of Target Tracking Algorithm in Thermal Imagery." The Journal of Computer Applications, Vol. 71, No. 16, pp. 1-71, Jan 2013.
[12] Zhan, Chaohui, Xiaohui Duan, Shuoyu Xu, Zheng Song, and Min Luo, "An improved moving object detection algorithm based on frame difference and edge detection." in Proceeding of Fourth International Conference on Image and Graphics, Sichuan: CH, pp. 519-523, Aug 2007.
[13] Hung, Mao-Hsiung, Jeng-Shyang Pan, and Chaur-Heh Hsieh, "Speed up temporal median filter for background subtraction." in Proceeding of First International Conference on Pervasive Computing Signal Processing and Applications, Harbin: CH, pp. 297-300, Sep 2010.
[14] Han, Ju, and Bir Bhanu, "Human activity recognition in thermal infrared imagery." in Proceeding of IEEE Computer Socity Conference on Computer Vision and Pattern Recognition-Workshops, San Diego: CA, pp. 17-17, Sep 2005.
[15] Jeon, Eun Som, Jong-Suk Choi, Ji Hoon Lee, Kwang Yong Shin, Yeong Gon Kim, Toan Thanh Le, and Kang Ryoung Park, "Human detection based on the generation of a background image by using a far-infrared light camera." The Journal of Sensors, Vol. 15, No. 3, pp. 6763-6788. Mar 2015.
[16] Beiping, Hou, and Zhu Wen, "Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients." The Journal of JCP, Vol. 6, No. 8, pp. 1597-1604, Aug 2011.
[17] Han, Guang, Xi Cai, and Jinkuan Wang, "Object Detection based on Combination of Visible and Thermal Videos using A Joint Sample Consensus Background Model." The Journal of JSW, Vol. 8, No. 4, pp. 987-994, Apr 2013.
[18] Goubet, Emmanuel, Joseph Katz, and Fatih Porikli. "Pedestrian tracking using thermal infrared imaging." The Journal of International Society for Optics and Photonics, Vol. 6206, pp. 62062C, May 2006.
[19] Davis, James W., and Vinay Sharma, "Robust Background-Subtraction for Person Detection in Thermal Imagery." in Proceeding of CVPR Workshops, pp. 128, Jun 2004.
[20] Davis, James W., and Vinay Sharma, "Robust detection of people in thermal imagery." in Proceeding of International Conference on Pattern Recognition, 2004, Vol. 4, pp. 713-716, Aug 2004.
[21] Lenac, Kruno, Ivan Maurović, and Ivan Petrović, "Moving objects detection using a thermal Camera and IMU on a vehicle." in Proceeding of International Conference on Electrical Drives and Power Electronics, pp. 212-219, Sep 2015.
[22] Qi, Bin, Vijay John, Zheng Liu, and Seiichi Mita, "Use of sparse representation for pedestrian detection in thermal images." in Proceeding of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 274-280, Jun 2014.
[23] Miezianko, Roland, IEEE OTCBVS WS Series Bench, "Terravic research infrared database.", 2005.
[24] Wu, Zheng, Nathan Fuller, Diane Theriault, and Margrit Betke, "A thermal infrared video benchmark for visual analysis." in Proceeding of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 201-208, Jnu 2014.
[25] Bilodeau, Guillaume-Alexandre, Atousa Torabi, Pierre-Luc St-Charles, and Dorra Riahi, "Thermal–visible registration of human silhouettes: A similarity measure performance evaluation." The Journal of Infrared Physics & Technology, Vol. 64, pp. 79-86, May 2014.
[26] Davis, James W., and Mark A. Keck, "A two-stage template approach to person detection in thermal imagery." in Proceeding of Seventh IEEE Workshops on Application of Computer Vision, Vol. 1, pp. 364-369, Jan 2005.

Gwang-Hyun Yu

2018 : MS College of Engineering

2016~2018: Electronics and Computer Engineering from Chonnam National University (CNU), Korea.
2017~: Cofounder of FTEN company, Korea.
※Interest areas : Machine Learning, Deep Learning, Digital Signal Processing, Image Processing.

Muhammd Zaigham Zaheer

2017 : MS College of Engineering

2012: Acquired undergraduate in Computer and Information Sciences from Pakistan Institute of Engineering and Applied Sciences (PIEAS).
2017: in Electronics and Computer Engineering from Chonnam National University (CNU), South Korea. 2018
2018: Pursuing PhD in the department of Computer Software from the University of Science and Technology (UST), South Korea.
※Interest areas: Machine Learning, Artificial Intelligence, Deep Learning, Image Processing.

Jin-Yeong Kim

1986: BS degree, in Dept. of Electronics Eng, Seoul National University
1988: MS degree, in Dept. of Electronics Eng, Seoul National University
1994: Ph.D degree, in Dept. of Electronics Eng, Seoul National University

1993~1994: Full-time lecturer, KT Software Research Center
1995~: Professor, Chonnam National University
※Interest areas : Digital Signal Processing, Image Processing, Speech Signal Processing

Do-Seong Shin

1993 : BS degree, in Dept. of Information and Communication Eng, Dongshin University
1998 : MS degree, in Dept. of Electronics Eng, Chonnam National University
2017 : Ph.D degree, in Dept. of Electronics, Information and Communication Eng, Chonnam National University

2009~2010 : Research professor, in Institute of Information Science and Engineering Research
2010~2016 : Research professor, in Information Technology Research Center
2017~ : Reaearch Manager, in DefenseTech Inc.
※Interest areas: Digital Signal Processing, Image Processing, Machine Learning, Biometrics