Anomalous Trajectory Detection in Surveillance Systems Using Pedestrian and Surrounding Information

Doan Trung Nghia[1], Sunwoong Kim[1], Vo Le Cuong[2], and Hyuk-Jae Lee[1]

[1] Inter-university Semiconductor Research Center, Department of Electrical and Computer Engineering, Seoul National University / Seoul, South Korea {nghiadt, ksw, hyuk_jae_lee}@capp.snu.ac.kr
[2] School of Electronics and Telecommunications, Hanoi University of Science and Technology / Hanoi, Vietnam cuong.vole@hust.edu.vn

* Corresponding Author: Hyuk-Jae Lee

Received August 1, 2016; Accepted August 28, 2016; Published August 30, 2016

** Regular Paper*

Abstract**:** Concurrently detected and annotated abnormal events can have a significant impact on surveillance systems. By considering the specific domain of pedestrian trajectories, this paper presents two main contributions. First, as introduced in much of the work on trajectory-based anomaly detection in the literature, only information about pedestrian paths, such as direction and speed, is considered. Differing from previous work, this paper proposes a framework that deals with additional types of trajectory-based anomalies. These abnormal events take places when a person enters prohibited areas. Those restricted regions are constructed by an online learning algorithm that uses surrounding information, including detected pedestrians and background scenes. Second, a simple data-boosting technique is introduced to overcome a lack of training data; such a problem particularly challenges all previous work, owing to the significantly low frequency of abnormal events. This technique only requires normal trajectories and fundamental information about scenes to increase the amount of training data for both normal and abnormal trajectories. With the increased amount of training data, the conventional abnormal trajectory classifier is able to achieve better prediction accuracy without falling into the over-fitting problem caused by complex learning models. Finally, the proposed framework (which annotates tracks that enter prohibited areas) and a conventional abnormal trajectory detector (using the data-boosting technique) are integrated to form a united detector. Such a detector deals with different types of anomalous trajectories in a hierarchical order. The experimental results show that all proposed detectors can effectively detect anomalous trajectories in the test phase.

Keywords**:** Pedestrian detection, Pedestrian tracking, Anomalous trajectory, Superpixel classification, Trajectory Features, Neural network (NN)

1. Introduction

Currently, the increasing number of cameras used for surveillance creates a new challenge in terms of data analysis, owing to the fact that storing the information captured by those cameras is not the only pivotal task. As these cameras are installed in public places, manual monitoring and validation of the human behavior recorded in those systems are often beyond the abilities of human observers. In reality, it is hard for a camera observer to rapidly react to anomalous trajectories captured by a surveillance system, because the observer has to manage a great amount of information from many cameras at the same time. Consequently, automatic analysis of video-based data as a support tool for monitoring personnel is required. For instance, potentially abnormal trajectories that are detected and annotated by such systems can have a significant impact on the monitoring of surveillance cameras, because only suspicious cases trigger a warning, and the reaction time is obviously shortened. In addition, cameras installed inside vehicles can be a good field for applications where computer vision solutions assist drivers with early warnings about abnormal movements of pedestrians, such as crossing a road against the traffic lights.

The definition of abnormal trajectories is understood in two ways. First, if one or more regions that are prohibited to access by humans exist in a video frame, any trajectories that move through such areas are, by definition, abnormal trajectories. Second, a track is also assumed to be abnormal if it is considerably different from the dominant trajectories in direction and/or speed. For instance, the dominant trajectories are the sets of tracks that are parallel to the main walking path, whereas the rare trajectories are the sets of tracks in which people cross the main walking path. To deal with those abnormal trajectories, two frameworks that make use of pedestrian and surrounding information are introduced in this paper.

The rest of the paper is organized as follows. Section 2 presents the related work in the literature, and Section 3 explains the proposed frameworks in detail. Sections 4 and 5, respectively, give experimental results and conclusions from this paper.

2. Related Work

Analysis of anomalies in object trajectories has been conducted for decades in many works. Owing to the fact that unusual behaviors rarely occur and strongly depend on specific contexts, many approaches employ previously recorded information about scenes where the directions of moving objects are already determined. However, there are other works that successfully overcome the existent problems by building adaptive learning systems that can be applied to various scenes. Extensive surveys of related research that investigate and make use of trajectory information for video surveillance systems have been published [1-3].

Hu *et al.* [4] were among the first to address the problem of detecting abnormal object trajectories by applying an unsupervised learning algorithm. A fast and accurate *fuzzy* K-means algorithm is the core of the tracking algorithm proposed in their work. Besides, a framework that automatically learns motion patterns by using spatial and temporal information was introduced, and each motion pattern was then represented as a chain of Gaussian distributions. This framework not only annotates the unusual events by analyzing the statistical data of motion patterns, but it is also able to predict behaviors for early warning.

Jung *et al.* [5] proposed a framework that includes four-dimensional (4-D) histograms and trajectory clustering. At the initial stage, sample trajectories are grouped into main clusters based on mixture of Gaussians, followed by a process of removing outliers. In each refined cluster, the position and velocity of each tracked object are arranged to form the 4-D histogram that represents the local of trajectories within each cluster. At the test step, the 4-D histogram having the same format as that built in the training phase is created for each new trajectory. With the newly calculated histogram, the coherence of the test track is evaluated against those from the training step in order to detect abnormal events.

A method for online learning and sequential anomaly detection in trajectories was introduced by Laxhammar and Falkman [6], in which a parameter-light algorithm called Sequential Hausdorff Nearest-Neighbor Conformal Anomaly Detector (SHNN-CAD) was investigated. This algorithm does not suffer from the difficulties of tuning a lot of parameters, of over-fitting, and of poorly calibrated alarm rates during a supervised online-learning process.

Having a slightly different approach while modelling the trajectories, H. Jeong *et al.* [7] formulated a track regarding the temporal–spatial property of the overall path, as opposed to the conventional similarity measures proposed in almost all other related papers. To represent a trajectory, position information is extracted by quantizing the location of trajectory sub-parts into cells sized at 10×10 pixels. Every point in the original track is assigned to a word (a cell), and a complete track is represented by a vocabulary that is the combination of many words. A temporal model of a trajectory is then formulated using a hidden Markov model to detect anomalous events.

To the best of our knowledge, for both clustering-based and non-clustering-based anomalous approaches, all related work in the literature only considers trajectories from the perspectives of direction and/or speed, while additional cues (such as background information) are not used effectively.

3. Proposed Frameworks

The first framework tries to warn about a person who enters prohibited areas for a number of consecutive frames. The motivation for the first framework is that cameras used for surveillance systems often capture scenes where the regions indicated for humans to appear have a considerably different appearance than the remaining regions in which humans are less likely to walk through. Therefore, if those distinguished regions are correctly separated, annotations abut humans entering such restricted areas are helpful to camera observers. Notice that to yield these kinds of abnormal event annotations, none of the directional properties of pedestrian trajectories are taken into account, since the only required task is to compare the positions of the detected human with that of the restricted area; therefore, fast annotations are possible. In the second framework, the direction and speed of a trajectory are considered. As a traditional approach, a supervised learning model is used to train the classifier with negative and positive trajectory samples. Notice that those samples are created by a data-boosting technique, where only normal trajectories and basic prior information about the scenes are involved. Furthermore, by assuming the average speed of human movement, deviations determine the ranges of slow, normal, and high speeds, so the speed of a given track in the test phase is compared with the defined speed thresholds in order to categorize that track into slow, normal, or fast. Note that inputs to the proposed frameworks are the trajectories derived from proposed human tracking algorithms [8, 9] where they used human detection algorithms [10, 11]. Ultimately, a hierarchical abnormal trajectory detector that makes use of the above two frameworks is introduced as a new approach

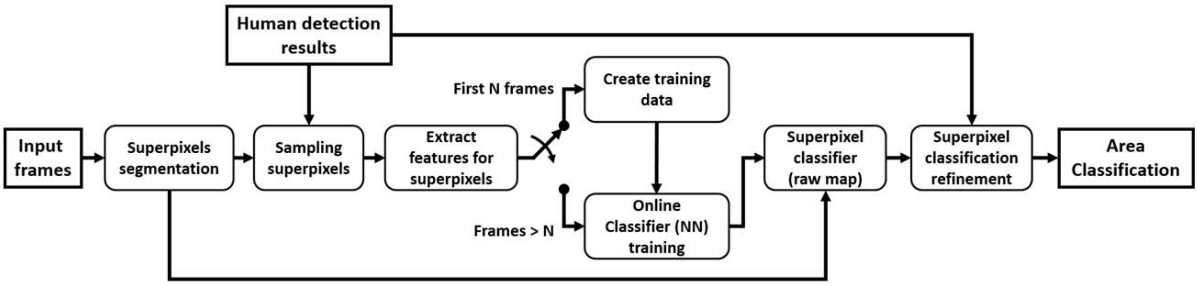


Fig. 1. The flow of the proposed framework to classify human-like and non-human-like areas.

for real-life applications, where it effectively deals with separate aspects of anomalous trajectory-based events.

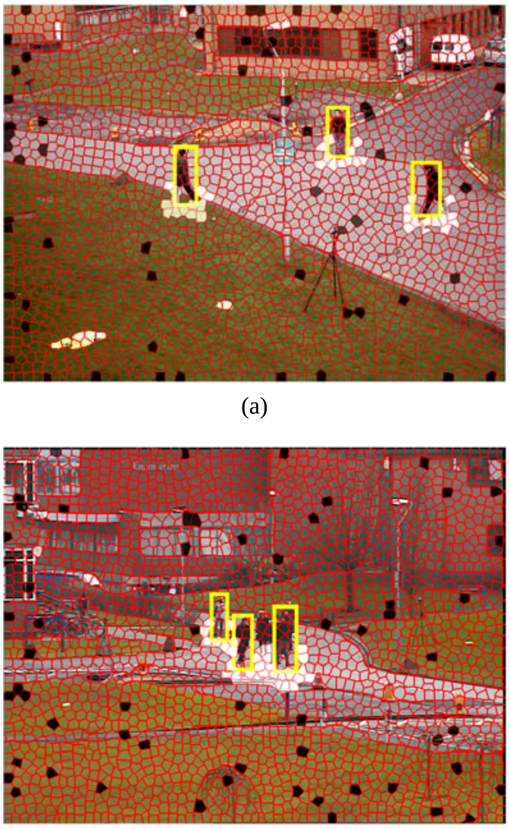
3.1 Anomalous Trajectory Detection in Prohibited Areas

Fig. 1 illustrates the flow of the abnormal trajectory detector, which tries to provide annotations when a person enters prohibited areas. Those restricted regions are either defined in an explicit manner, such as hand labelling, or a learning algorithm automatically describes the restricted areas. In the proposed framework, the second approach is chosen, since it can be adapted to many circumstances in real applications.

The main idea of the proposed framework is to make use of existing human detection results in order to build a map that represents two kinds of area. One of them indicates the regions in which humans can appear, whereas the other points out the regions in which human are less likely to appear or that are prohibited areas. To separate those regions, an image is initially segmented into sub-regions; in particular, the choice of segmentation methods could strongly affect the overall framework. Consider the simple grid segmentation, by which a given video frame is split into multiple sub-regions having the shapes of small squares. It is obvious that the boundaries of objects that exist in the image frame could be broken if a sparse grid segmentation is used; hence, the information contained in those sub-regions could end up having different appearance properties. On the other hand, using a dense grid-segmentation creates more computational demand, which is also not suitable for real-time applications. Thanks to superpixels [12, 13], such a segmentation method splits the whole frame into smaller parts while preserving the boundaries among objects that appear in the frame. Besides, all pixels inside a superpixel often contain similar appearances, since they are grouped based on small gradient differences. It turns out another advantage for using superpixel segmentation is where spatial redundancy often existed. Therefore, applying feature engineering to construct the important features for a superpixel could remarkably reduce the complexity of the learning model without affecting the classification accuracy. Human detection results, such as human bounding boxes, are used to indicate the labels of sampling superpixels by considering the relative location between a human bounding box and surrounding superpixels. The superpixels locating around the footage areas of a human bounding box are labelled a human-like area, while some random superpixels located distant from the human bounding boxes are assumed to be a non-human-like area. Figs. 2(a) and (b) give illustrations about the above superpixel sampling step, where the yellow rectangles are the human bounding boxes. The lighter superpixels surrounding the human bounding boxes are labelled as human-like area superpixels, whereas the darker superpixels randomly located far from human bounding boxes are marked as non-human-like area superpixels. Once positive (human-like-area) and negative (non-human-like-area) superpixel samples are completely taken from the segmented frame, each superpixel then goes through the features extraction step before an online learning algorithm is taken into account. As location and appearance information of a superpixel are pivotal cues, the feature vector of the superpixel is defined as a set of p_x, p_y, c_R, c_G, c_B, and c_{Grey}. The parameters $\{p_x, p_y\}$ are the average corresponding horizontal and vertical positions of all pixels inside a sample superpixel in the image frame, which aims at describing the global location information of a superpixel. On the other hand, $\{c_R, c_G, c_B, c_{Grey}\}$ define the appearance characteristics of a superpixel by exploiting the average values of four-color channels (red, green, blue, and grey) of all the pixels.

At the first N-frames of the video sequence, all superpixel features and their labels are stored to make sure that the amount of training data is large enough prior to the training process. This preparation step is illustrated as the Create training data step in Fig. 1. In this step, it is necessary to guarantee that people are assumed to appear only in the allowed areas. When the number of frames exceeds the pre-determined value of N, the online neural network (NN) classifier is trained by those prepared features and labels. The superpixel NN model consists of two hidden layers, where the number of neurons are 10 and 6 for the first and second hidden layers, respectively. Note that from the (N+1) frame, the processes of segmenting a new frame into superpixels and computing their features with another step that removes old superpixel feature vectors are required for the online training classifier. In this paper, the value of N is set to 200, which gives acceptably empirical results for the test datasets. Note that an online classifier is able to deal with changes in terms of illumination of the objects, such as light and weather conditions. Obviously, this aspect is a benefit from the proposed framework when it is applied to real applications.

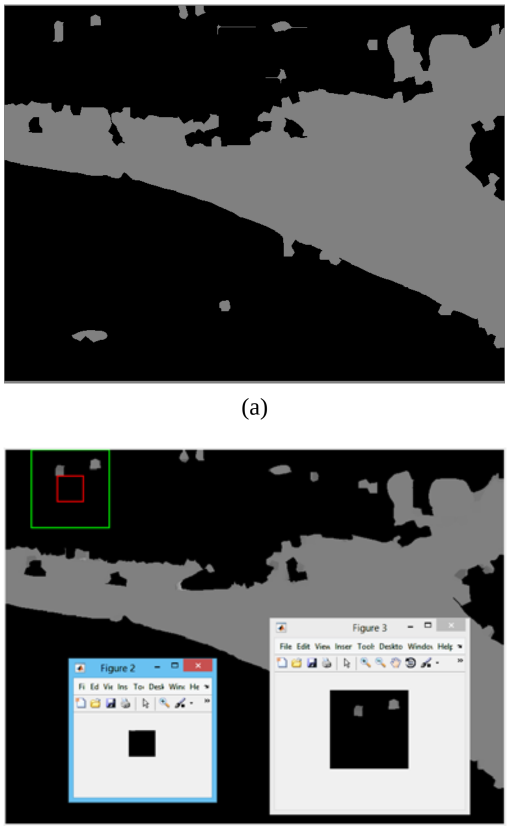
By using the already trained classifier, every



(a)

(b)

Fig. 2. Results from the superpixel segmentation step (a) PETS09-S2-L1-VIEW-001, (b) PETS09-S2-L1-VIEW-004.



(a)

(b)

Fig. 3. Human-like (grey) and non-human-like (black) area maps of the PETS09-S2-L1-VIEW-001 dataset (a) raw map, (b) refining process by using the moving window approach.

segmented superpixel of the original frame is then categorized into two groups: non-human-like-area superpixels and human-like-area superpixels. This very first classification results in a raw map where mis-classified superpixels can exist. Fig. 3 shows the superpixel classification results in which human-like and non-human-like superpixels are represented by grey and black regions, respectively. The respective grey and black regions represent the non-prohibited and prohibited areas for humans. Fig. 3(a) shows raw superpixel classification of a given frame, where all superpixels in that frame are classified by the trained classifier, one by one. As clearly seen from Fig. 3(a), there exist wrongly classified superpixels where some non-human-like superpixels are treated as opposite. To handle that problem, a moving window approach is proposed, the main purpose of which is to detect the wrongly classified superpixels. The refinement step using the proposed window approach is illustrated in Fig. 3(b). In this step, two different windows with the same center recursively scan the whole image, and the content in the inner window is then compared with that of its outer window. It is obvious that the human-like area cannot be completely covered by the prohibited area, since a person cannot suddenly appear in any area of the image without his/her trajectories. This leads to the fact that such an area was mis-classified, and it is to be corrected to a non-human-like area by changing the label of all wrongly detected human-like area superpixels to non-human-like area superpixels. Although a case exists where the refinement process causes a side effect when it also removes correct human-like area superpixels, the experimental results show that this negative aspect still does not strongly affect the accuracy of the current framework.

Another problem occurs when the online classifier is used; not all the superpixels surrounding human bounding boxes are human-like superpixels. Therefore, only the "safe" superpixels are chosen to train the classifier. The safe superpixels are the ones that belong to humans appearing in non-prohibited areas, which are determined by the current classifier. Once the human-like area map is completed, the detection process turns out to be fairly easy. If the footage area of any person mostly contains the black area for consecutive frames, then this person is assumed to have entered a prohibited area.

3.2 Anomalous Trajectory Detection in Terms of Direction and Speed

The second type of abnormal trajectory detection differentiates between dominant trajectories and rare ones, where the unusual trajectories contain significantly different expressions regarding direction of movement and speed. The whole process of the proposed framework is shown in Fig. 4, where only the direction is taken into

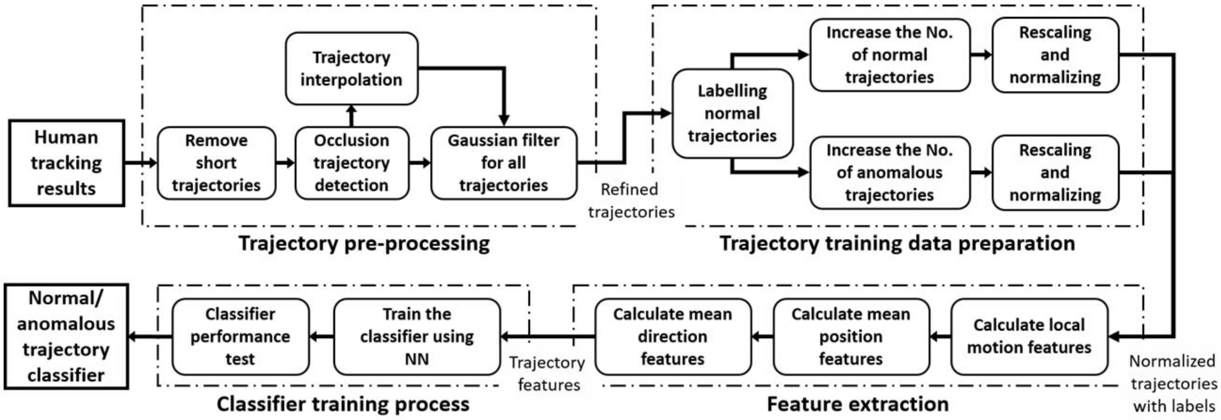
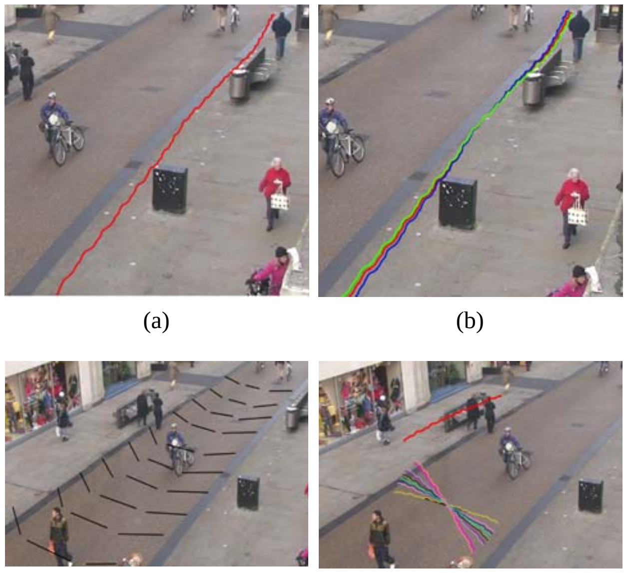


Fig. 4. The flow of the proposed framework to detect an anomalous pedestrian trajectory in which only the direction is considered.

account. Initially, given the raw human trajectories of a human tracking algorithm, those raw tracks enter a pre-processing step that eliminates erroneous tracking data as well as refines tracks affected by noise and/or occlusions. All short trajectories are removed, since their information is not sufficient for further steps; hence, only the ones having a length greater than or equal to M frames are kept. In this paper, the value of M is set to 50, which achieves good results. Occlusions often exist in crowded pedestrian scenes, which results in discontinuous trajectories. In particular, trajectories with a discrete expression lead to more difficulties in the steps that re-sample or extract features of trajectories. This negative impact of occlusion is resolved by using both linear and n-tap interpolation techniques, with weighted values used to combine the results of those two methods. The non-discrete tracks are then smoothed using a Gaussian filter to remove the noise caused by the tracking algorithm itself, before entering the trajectory training data preparation step.

In the trajectory training data preparation step, positive (normal) and negative (abnormal) samples are created. The proposed framework is different from others in that only normal trajectories and basic prior information about the given scenes are needed to build the training data for both abnormal and normal samples. The motivating idea is that in anomalous trajectory detection applications, unusual trajectories occur at a significantly lower frequency, in comparison with normal ones. It leads to the fact that a greater training dataset for trajectories in general, and rare trajectories in particular, have to be built to achieve high accuracy from the detection algorithm. First, based on the labelled normal tracks, a new set of normal tracks (which are slightly rotated from the original ones) is created. Fig. 5(a) shows a normal trajectory; Fig. 5(b) depicts additional trajectories formed by rotating the track in Fig. 5(a) using small rotation angles. Note that the rotation angles are constrained to completely preserve the general direction of the normal ones. Rescaling (up-scaling and down-scaling) the tracks plays a pivotal role in ensuring that the classifier is persistent with a wide variety of pedestrian speeds in real applications. Besides, instead of considering the complete path of a person, the smaller parts of the whole track are split and used to extract track features. The reason for that is that an abnormal track often contains a small part that is assumed to be abnormal, whereas the remaining part is still normal; therefore, it is more accurate to evaluate the track by considering its sub-parts only, not the complete track. A normalizing step is then performed to subsample the whole track into shorter tracks with the same length, and the sub-tracks partly overlap each other. In addition, there is a little difference when the same measure is applied to abnormal tracks to increase the amount of anomalous training data. With anomalous trajectories, prior information is required to create such samples for the training process. For instance, in the test sequence TownCentre, the dominant trajectories are the ones that pass alongside the pavement, whereas the rare trajectories occur when people cross the main road. Then, the abnormal trajectory is created as follows. First, the



(a) (b)

(c) (d)

Fig. 5. Examples of normalized trajectory generation from the TownCentre dataset (a) a normal track, (b) additional normal tracks, (c) base lines of rare tracks, (d) a normal trajectory (red color) and abnormal trajectories created from the normal one (other colors).

base lines that represent the directions of rare tracks are manually created, as shown in Fig. 5(c). The black lines represent the crossing movements alongside the road. Second, for each base line of the rare tracks, additional abnormal tracks are created by rotating the normal track around that base line. Taking the normal track as a seed track for the creation of rare trajectories preserves the physical properties in terms of curve, fluctuation ratio, speed, etc. Fig. 5(d) shows the abnormal tracks created from the normal one by using the above technique. At the end of this step, for each newly created anomalous track, several smaller parts are split from the complete track, and all sub-tracks are normalized in the same manner as explained for the normal tracks.

The next feature extraction step is applied to the normalized tracks that already contain either normal or abnormal labels. Because each track is represented by a group of consecutive positions, the global positions, the global directions, as well as the local motions (between each pair of successive points) of a trajectory are elemental to representing important properties of that track. Specifically, normalized track t with length n is the set of n successive points in a two-dimensional plane, which is represented as

$$t_n = \{(x_1, y_1), (x_2, y_2), ..., (x_{n-1}, y_{n-1}), (x_n, y_n)\} \quad (1)$$

A complete feature vector of t_n comprises five elements $\{D_x, D_y, P_x, P_y, \alpha\}$, which are expressed as follows:

$$\begin{cases} D_x = mean\{(x_2 - x_1), (x_3 - x_2), ..., (x_n - x_{n-1})\} \\ D_y = mean\{(y_2 - y_1), (y_3 - y_2), ..., (y_n - y_{n-1})\} \end{cases} \quad (2)$$

In (2), D_x and D_y give the average values of horizontal and vertical local motion, respectively, between all consecutive positions in the normalized trajectory. On the other hand, P_x and P_y illustrate the respective global horizontal and vertical positions of the whole trajectory, as written in (3):

$$\begin{cases} P_x = mean\{x_1, x_2, ..., x_{n-1}, x_n\} / X_{\max} \\ P_y = mean\{y_1, y_2, ..., y_{n-1}, y_n\} / Y_{\max} \end{cases} \quad (3)$$

In (3), the parameters X_{max} and Y_{max} represent the positions having the maximum values in the horizontal and vertical directions of the frame, respectively. Ultimately, the last element of the trajectory feature vector is denoted as α, which expresses the average direction of the normalized track:

$$\alpha = mean(\tan^{-1}\frac{y_2 - y_1}{x_2 - x_1}, \tan^{-1}\frac{y_3 - y_2}{x_3 - x_2}, ..., \tan^{-1}\frac{y_n - y_{n-1}}{x_n - x_{n-1}}) \quad (4)$$

Next, the trajectory feature vectors are transferred to the classifier training process step. In this step, an NN is trained by using all trajectory features, as well as their labels, followed by a classifier evaluation process using the test data. Note that the same NN configuration and training methods as explained in Section 3.1 are used for the second framework. The only difference is the number of neurons in the input layer, which is six for the trajectory feature data.

When speeds are taken into consideration, pedestrians who move at fast and slow speeds are also assumed to contain some abnormal movement. In the training process, after converting centroid positions of pedestrians in the image plane to the world plane by using a given camera model of the scene, all human trajectories are represented in the world coordinate system. A trajectory speed is defined as fast, slow, and medium, as presented by Jung *et al.* [5]. If the current trajectory is classified as slow or fast, an annotation is shown on top of the corresponding human bounding box to illustrate such anomalous trajectories.

```
FOR i=1:maxTrackIndex
 t=Tracks[i];
 isInProhibitAreas =
        Abnormal_Track_Detector_1(t);
 If isInProhibitAreas
        Abnormal_Track_Notation(t);
 ELSE
       [isAbnDirection, isAbnSpeed] =
       Abnormal_Track_Detector_2(t);

        If isAbnDirection
         Abnormal_Track_Notation(t);
        END IF

        If isAbnSpeed
         Abnormal_Speed_Notation(t);
        END IF
 END IF
END FOR
```

Fig. 6. The pseudo code of the hierarchical abnormal trajectory detector.

3.3 Hierarchical Abnormal Trajectory Detector

The two proposed frameworks deal with different types of trajectory-based anomalous events. Regarding the main characteristics of each introduced framework, it is possible to combine them in order to construct a united and consistent detector that effectively utilizes the advantages of both frameworks. Specifically, the first type of abnormal trajectory detection is fairly simple, where one of the most complicated computational processes is segmenting a frame into superpixels, which is required only once for the first background scene. The remaining online training classifier used to build the human-like and non-human-like map (as depicted in Fig. 1) is not taken into account as the frequency of every new frame. Instead, this process is scheduled for every 50 frames regarding the current configuration of the implementation program. Furthermore, the classifier training process in general can be either pipelined or parallel to the detection process. This assumption holds for both proposed frameworks, which is also a key aspect for designing a real-time system in surveillance-related applications. Using the human-like-area map to evaluate whether a trajectory is anomalous or not only requires simple comparisons. Besides, it is clear that there is no need to consider a track in the prohibited area as normal or abnormal in terms of direction and speed. Therefore, a trajectory is first classified in this manner to accelerate the computing time of the whole system, since all processes needed for the second framework are terminated early if the examining track lies inside the restricted areas. It turns out that the second proposed framework becomes involved when the position of a trajectory is detected as located inside human-like areas. This exclusive detector is able to detect rare trajectories by considering direction and speed.

Fig. 6 illustrates the main idea of the combined detector. Given a human-like-area map, the hierarchical abnormal-trajectory detector examines all trajectories (t) existing in the current frame; Abnormal_Track_Detector_1 is brought out to tackle the simple cases, where a track's path moves through a prohibited area. If current track t is found in such restricted areas, a notation for the person whose track is t is set, and the process is finished for current track t. Otherwise, the second classifier is picked to evaluate the normality of t in terms of speed and direction, which is denoted as Abnormal_Track_Detector_2 in Fig. 6. Finally, with respect to each unusual property of t, the notifications for abnormalities from the direction and/or speed aspects are noticed for the corresponding persons. If t is a normal track, the two detectors are tested for t before the whole process is applied to the next track in the current frame.

4. Experimental Results

All proposed frameworks were implemented in Matlab version R2015b for the Windows operating system. Two datasets, PETS09-S2-L1-VIEW-001 and PETS09-S2-L1-VIEW-004 [14], were used for the first proposed framework, in which a person enters a restricted area. On the other hand, the TownCentre [15] dataset was evaluated for the second proposed framework, where rare tracks have different directions and speeds, compared to the dominant tracks. Another dataset named SNUCafe1 was recorded to evaluate the effectiveness of the hierarchical detector, as explained in Section 3.3.

Fig. 7 shows the experimental results of the first proposed framework. Figs. 7(a) and (b) illustrate the original input frames, whereas Figs. 7(c) and (d) depict the refined human-like area map, which is the output of the online NN classifier followed by a refinement process, where the proposed moving window approach presented in Section 3.1 is involved. In Figs. 7(c) and (d), the roads in which almost all people in the datasets appear correspond to the grey part of the human-like area map. In contrast, the grass areas are classified as black parts, where humans are not supposed to appear. Note that the human-like area map is the result of an online learning method; hence, it was adaptively changed, according to illumination conditions.

The annotations for persons who enter restricted areas are based on the human-like area maps, as shown in Fig. 8. In particular, if the track of that person lies in the black areas over a predetermined number of consecutive frames,

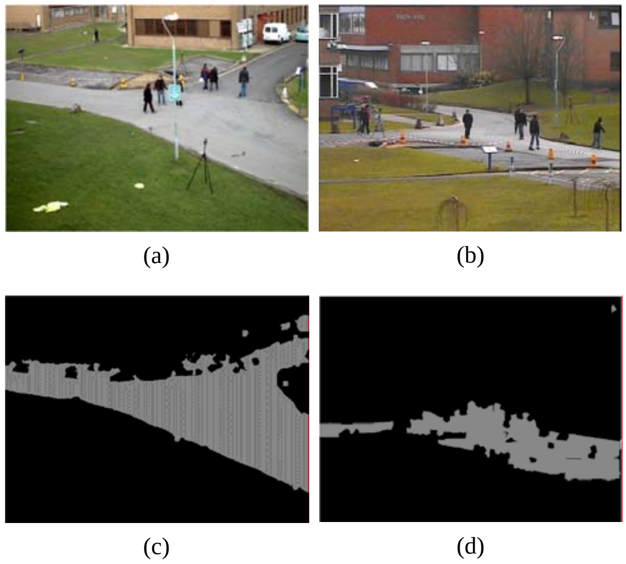


Fig. 7. Results of the superpixel classification refinement step (a) original frame of the PETS09-S2-L1-VIEW-001 dataset, (b) original frame of the PETS09-S2-L1-VIEW-004 dataset, (c) refined human-like area maps of (a), (d) refined human-like area maps of (b)

then this person is assumed to have entered a prohibited area. The notifications for such trajectories are denoted by the large red bounding boxes of the humans with indices 3 and 4. When a person with index 3 moves out of the grass area (a prohibited area) the notification is turned off, as shown in Fig. 8(d).

In the second proposed framework, the classifier was evaluated in the last step, as depicted in Fig. 4, and the results show that 98.2% of the trajectories in the test data were correctly classified. With labelled normal trajectories, the accuracy of the trained classifier is 98.8%, and the figure for the labeled anomalous tracks is still significantly high at 96.6%. Fig. 9 shows results in which the second proposed framework was applied to the TownCentre test sequence. If a person has an abnormal trajectory in direction, a large red bounding box is used to indicate that trajectory. On the other hand, if a person moves with either a slow or fast speed, a speed classification sign ("Fast" or "Slow") is placed on top of his/her bounding box. Figs. 9(a) and (b) capture abnormal trajectories when people are crossing the road, and all people in those two frames are moving at normal speeds. In contrast, Figs. 9(c) and (d) only depict anomalous movements with unusual speeds but normal directions. As shown in those figures, two persons are moving at high speed, compared to the average speed calculated in the training phase. Finally, Figs. 9(e) and (f) depict anomalous trajectories where speed and direction are both abnormal. The man in Fig. 9(e) stays in the same position, which results in a slow speed and an anomalous direction. The woman whose track crosses the road in Fig. 9(f) moves in an unusual direction, since the dominant tracks are parallel to the road. Besides, the speed of this woman is also considered high, as indicated by the notification on top of her bounding box.

Figs. 10 and 11 depict the experimental results of the hierarchical detector built based on the two proposed frame-

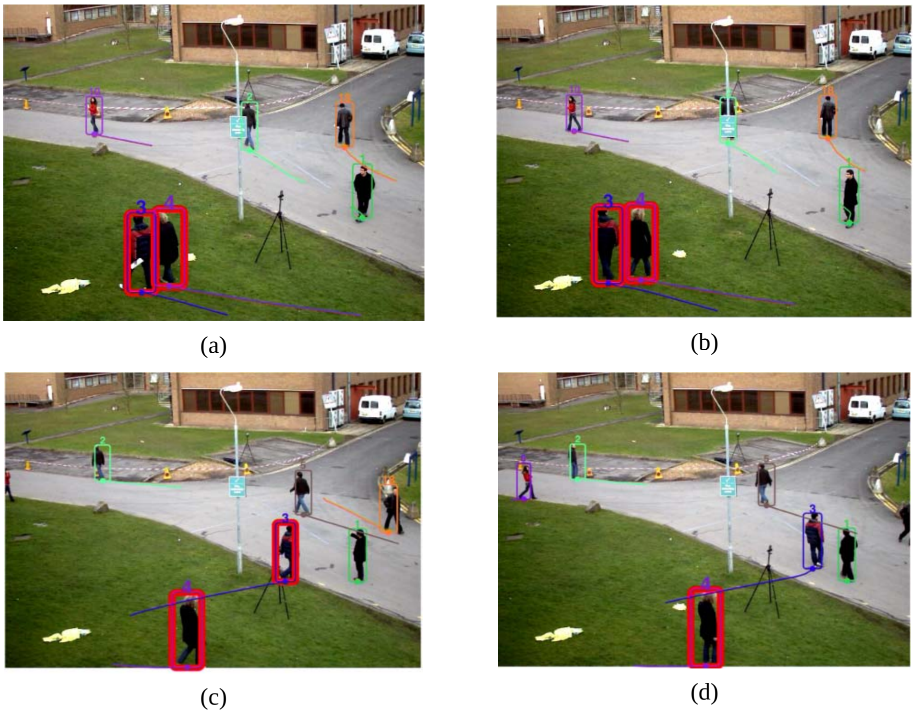


Fig. 8. Prohibited trajectory annotations of the PETS09-S2-L1-VIEW-001 dataset: (a, b, c) two people with indices 3 and 4 are detected, (d) one person with index 4 is detected.

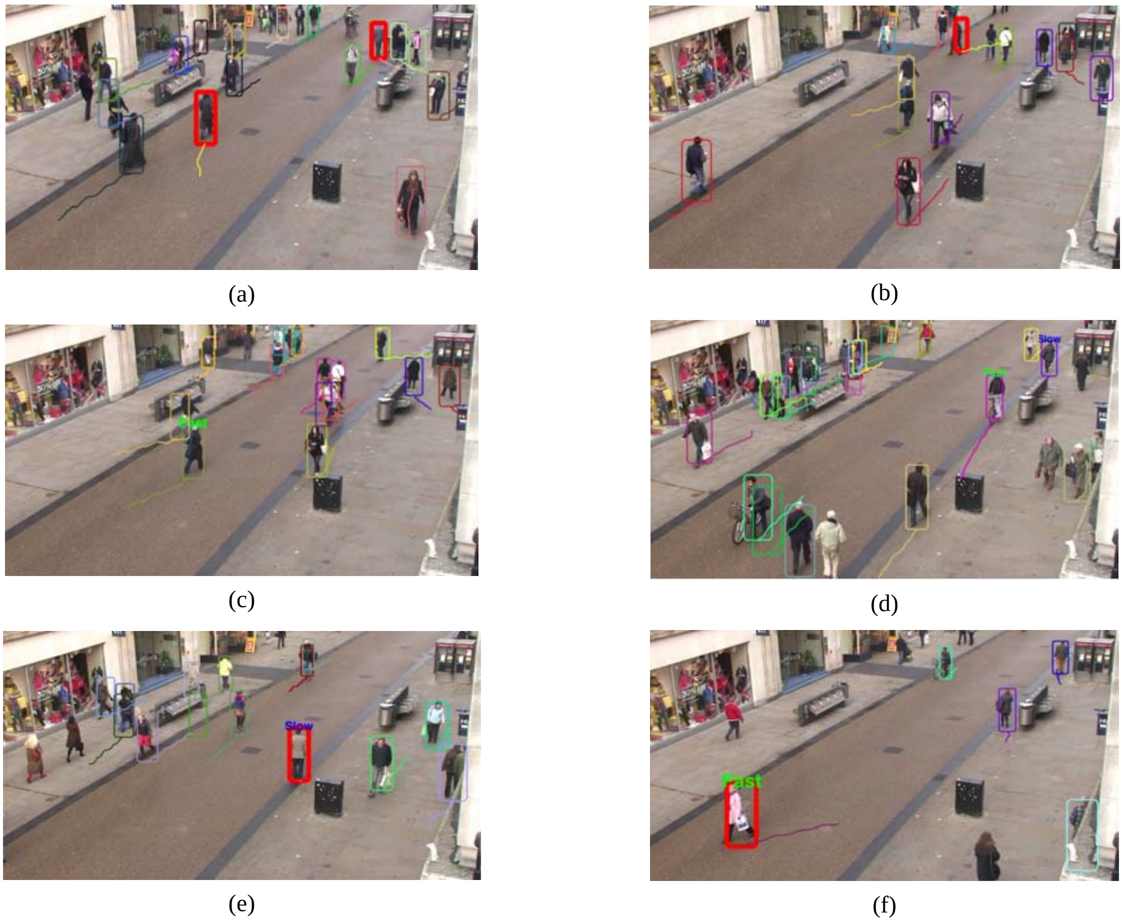


Fig. 9. Rare trajectory detection in terms of speed and direction from the TownCentre dataset (a, b) humans crossing the road at normal speed but in anomalous directions, (c, d) normal moving directions but with anomalous speeds, (e, f) anomalous trajectories in both speed and direction.

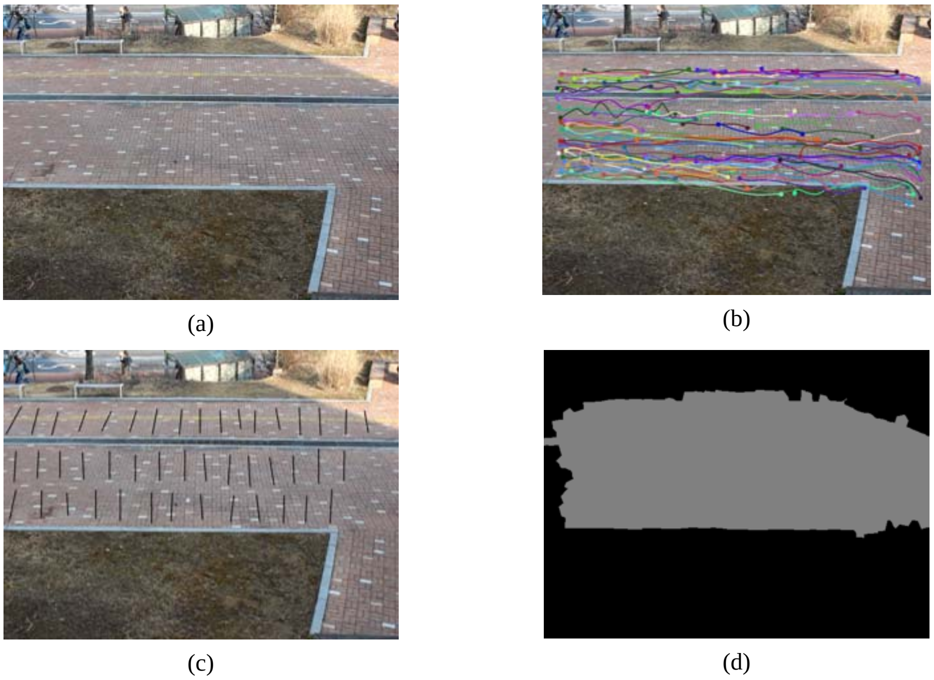


Fig. 10. The results of creating sample rare trajectories and a superpixel classification map (a) a background scene, (b) normal trajectories, (c) base lines of rare trajectories, (d) the corresponding human-like and non-human-like areas of the background scene in (a).

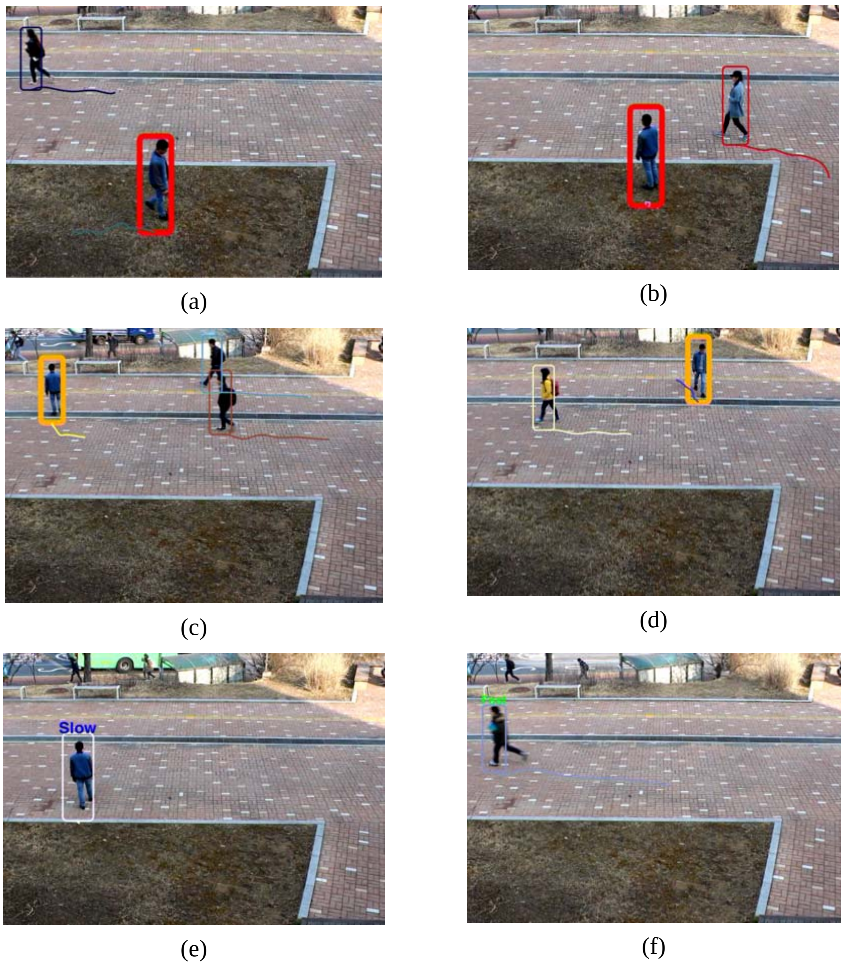


Fig. 11. Rare trajectory detection results from the hierarchical detector conducted on the SNUCafe1 dataset (a, b) anomalous trajectories detected in prohibited areas, (c, d) anomalous trajectories in terms of direction, (e, f) anomalous trajectories in terms of speed.

works. Fig. 10(a) shows a background scene, whereas the majority of pedestrian trajectories with directions relatively parallel to the main moving path, are depicted in Fig. 10(b). Base on that assumption, rare track base lines are created in the directions that are relatively perpendicular to the main moving path, as shown in Fig. 10(c). Note that the base lines do not need to conform exactly to the main moving path at a specific angle, since a significant number of rare tracks will be created by rotating a normal track within a range of rotation angles alongside the base lines. Fig. 10(d) illustrates a human-like-area map constructed during the online learning process of the first framework. The main moving path in Fig. 10(a) corresponds to the grey area of Fig. 10(d). In contrast, other areas in Fig. 10(a) where humans are not detected, such as the soil area located in the bottom left corner of the background scene, result in the black areas in Fig. 10(d).

Fig. 11 shows the notifications for the detected abnormal trajectories of the hierarchical detector. If a trajectory appears in a prohibited area, then a large red bounding box represents the warning sign. In this situation, owing to the fact that the man is standing inside the soil area that corresponds to the non-human-like areas based on the created map, as depicted in Fig. 10(d), notations are given as shown in Figs. 11(a) and (b). On the other hand, if a trajectory is located in the allowed moving areas and is classified as an unusual track in terms of direction, a large bounding box in orange is used to indicate the event. Those examples are described in Figs. 11(c) and (d), when a man crosses the main moving path where other tracks behave as normal ones. Finally, regarding the trajectory speeds, slow or fast notifications are placed on top of the human bounding boxes to signal anomalous events, shown in Figs. 11(e) and (f), in which one man is moving at a slow speed, whereas the other is running.

5. Conclusion

Two frameworks that detect anomalous trajectories are introduced in this paper, where notifications are employed to indicate the tracks of humans entering restricted areas and moving in unusual directions and at unusual speeds. For a superpixel in the first framework, information about the global position and the average values for R, G, B, and grey color channels are used to form the features. For the second framework, a complete track given by the pedestrian tracking algorithms is initially refined in the pre-processing step, where noise and/or occlusion effects are mitigated. In particular, a proposed data-boosting technique is carried out to significantly increase the amount of training data for both anomalous and normal trajectories. This method takes the refined tracks of the previous block and creates more training data using fundamental information about the background scenes. The newly generated tracks are then re-scaled to deal with speed variances in real applications before being normalized into shorter parts. These normalized tracks are represented in terms of track features such that a complete feature vector includes information about global position, motion, and direction. Besides, two trajectory speed thresholds are calculated from the training data. The speed of a test track is compared with those thresholds to indicate whether this track contains an anomalous (fast or slow) moving speed or not. The frameworks make use of NN-based classifiers. Finally, the experimental results were conducted with PETS09-S2-L1-VIEW-001, PETS09-S2-L1-VIEW-004, and TownCentre datasets, along with a self-recorded SNUCafe1 dataset to evaluate the effectiveness of the proposed abnormal trajectory detectors. It was observed that the empirical results achieved high accuracy when almost all of the anomalous trajectories were captured and given notifications, as explained in Section 4.

Acknowledgement

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1008) supervised by the IITP (Institute for Information & communications Technology Promotion)

References

[1] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *Circuits and Systems for Video Technology, IEEE Transactions on* 18.8 (2008): 1114-1127. Article (CrossRef Link)

[2] P. Remagnino, S. A. Velastin, G. L. Foresti, and M. Trivedi, "Novel concepts and challenges for the next generation of video surveillance systems," *Machine Vision and Applications* 18.3 (2007): 135-137. Article (CrossRef Link)

[3] M. Shah, "Understanding human behaviour from motion imagery," *Machine Vision and Applications* 14.4 (2003): 210-214. Article (CrossRef Link)

[4] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," Pattern Analysis and Machine Intelligence, *IEEE Transactions on* 28.9 (2006): 1450-1464. Article (CrossRef Link)

[5] C. Jung, L. Hennemann, and S. R. Musse, "Event detection using trajectory clustering and 4-D histograms," *Circuits and Systems for Video Technology, IEEE Transactions on* 18.11 (2008): 1565-1575. Article (CrossRef Link)

[6] R. Laxhammar and G. Falkman, "Online learning and sequential anomaly detection in trajectories," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36.6 (2014): 1158-1173. Article (CrossRef Link)

[7] H. Jeong, H. J. Chang, and J. Y. Choi, "Modeling of moving object trajectory by spatio-temporal learning for abnormal behaviour detection," *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE, 2011.

Article (CrossRef Link)
[8] G. Führ and C. R. Jung, "Combining patch matching and detection for robust pedestrian tracking in monocular calibrated cameras," *Pattern Recognition Letters* 39 (2014): 11-20. Article (CrossRef Link)
[9] A. Milan, L. Leal-Taixé, K. Schindler, and I. Reid, "Joint tracking and segmentation of multiple targets," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. Article (CrossRef Link)
[10] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36.8 (2014): 1532-1545. Article (CrossRef Link)
[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.9 (2010): 1627-1645. Article (CrossRef Link)
[12] X. Ren and J. Malik, "Learning a classification model for segmentation," *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003. Article (CrossRef Link)
[13] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on.* Vol. 2. IEEE, 2004. Article (CrossRef Link)
[14] J. Ferryman and A. Ellis, "PETS2010: Dataset and challenge," Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on. IEEE, 2010. Article (CrossRef Link)
[15] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. Article (CrossRef Link)

Doan Trung Nghia received his B.S. degree in the Electrical-Electronics Engineering Department from Posts and Telecommunications Institute of Technology, Hanoi, Vietnam in 2014. He is currently working toward the M.S. degree in the Electrical and Computer Engineering Department at Seoul National University, Seoul, Korea. His research interests are hardware-aware algorithm optimization, hardware design for videos/images, and computer vision applications.

Sunwoong Kim received B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2010, 2012, and 2016, respectively. He is currently working as a postdoctoral researcher at Seoul National University. His research interests are hardware design for multimedia systems, computer architecture, and memory systems.

Cuong Vo Le received the Ph.D. degree in ultrahigh-speed and ultrahigh-sensitivity camera field from Kinki University, Osaka, Japan, in 2009. From 2009-2011, he worked as a senior research engineer to develop Software Defined Radio devices at Viettel Group. Since 2011, he has been an Assistant Professor at Hanoi University of Science and Technology. From 2014-2015, he worked as a research associate in the Korea Global Frontier Project with Center for Integrated Smart Sensors at KAIST to develop object detection and structured light deep extraction. His research topics are applications of deeplearning, real-world person re-identification and image sensor design.

Hyuk-Jae Lee received B.S. and M.S. degrees in Electronics Engineering from Seoul National University, Korea, in 1987 and 1989, respectively, and obtained a Ph.D. degree in Electrical and Computer Engineering from Purdue University at West Lafayette, Indiana, in 1996. From 1998 to 2001, he worked at the Server and Workstation Chipset Division of Intel Corporation in Hillsboro, Oregon as a senior component design engineer. From 1996 to 1998, he was on the faculty of the Department of Computer Science of Louisiana Tech University at Ruston, Louisiana. In 2001, he joined the School of Electrical Engineering and Computer Science at Seoul National University where he is currently working as a Professor. His research interests are in the areas of computer architecture and SoC design for multimedia applications.