

Moving Object Detection with Background Model based on Spatio-Temporal Texture

Ryo Yumiba, Masanori Miyoshi
Hitachi Research Lab, Japan
{ryo.yumiba.xp, masanori.miyoshi.br}@hitachi.com

Hironobu Fujiyoshi
Chubu University, Japan
hf@cs.chubu.ac.jp

Abstract

Background subtraction is a common method for detecting moving objects, but it is yet a difficult problem to distinguish moving objects from backgrounds when these backgrounds change significantly. Hence, we propose a method for detecting moving objects with a background model that covers dynamic changes in backgrounds utilizing a spatio-temporal texture named "Space-Time Patch", which describes motion and appearance, whereas conventional textures describe appearance only. Our experimental results show the proposed method outperforms one conventional method in three scenes: in an outdoor scene where leaves and branches of a tree are waving in intermittent wind, in an indoor scene where ceiling lights are turned on and off frequently, and in an escalator scene beside a window facing outdoors where some passengers are leaning over the hand-rail.

1. Introduction

Background subtraction is a common method for detecting moving objects, and it has been widely used in many surveillance systems. Moving object detection using background subtraction compares an input image with a background model previously prepared, and picks up regions in an input image, which do not match a background mode. Moving object detection using background subtraction has the advantage of not requiring previous knowledge of moving objects such as shapes or movements. On the other hand, background subtraction has a problem in that it cannot discriminate moving objects from backgrounds when these backgrounds change significantly. There are many types of background changes, which we put into two categories; one is global changes in images such as transitions of sunlight or adjustments of ceiling light luminance, and the other is local changes in images such as waving of tree branches or fluctuations of water.

Many kinds of methods have been proposed for generating a background model to cover changes in

backgrounds and robustly detecting moving objects. These methods could be roughly divided into two approaches. Methods of the first approach generate background models of each pixel, which is the minimum component of an image, using statistical distribution based on backward observation [1][2][3]. These pixel-wise background models have an advantage of covering frequent changes in the pixel grayscale in local areas in images, such as waving of tree branches or fluctuations of water. On the other hand, these pixel-wise background models have a defect in that they hardly follow rapid translations of a pixel grayscale well across a global area in images, such as transitions of sunlight or adjustments of ceiling light luminance, because statistical distributions of each pixel grayscale belonging to the background translate substantially in a short time according to that kind of changes.

Methods of the second approach generate background models of each small patch in images using features robust to changes in luminance [4][5]. These patch-wise backgrounds models have an advantage in that features in small patches represent textures robustly within the patch and they can cover global changes in images such as transitions of sunlight or adjustments of ceiling light luminance. On the other hand these patch-wise background models have a defect in that they hardly follow rapid and intensive changes within image patches such as waving of tree branches or fluctuations of water, because textures within an image patch changes substantially according to that kind of changes.

There is a method for generating a background model which uses a combination of a pixel-wise background model and a patch-wise background model [6]. This combinatorial method reduces false positives using conjunction of detection results of a pixel-wise method and a patch-wise method. Meanwhile, this combinatorial method has a redundancy in that it executes two procedures of signal processes for one region in images.

We propose a method for detecting moving objects with a background model that covers dynamic changes in backgrounds using a spatio-temporal texture which describes motion in addition to appearance. This proposed method can cover global changes in images by using appearance information similar to other conventional spatial textures. In addition, it can cover local changes in

images by using motion information, although local changes are difficult to cover for conventional patch-wise models which use only appearance information.

In Section 2 we describe space-time patch (hereafter, ST-Patch in short) features, which are a kind of spatio-temporal texture, used in our proposed method. In Section 3, we describe a method for generating a background model using ST-Patch features, and a method for detecting moving objects using the generated background model. In section 4, we describe the evaluation experiments of our proposed method. In section 5, we conclude this paper.

2. ST-Patch Features

ST-Patch features describe appearance and motion information simultaneously within a small spatio-temporal patch in videos [7]. These ST-patch features have been used for several applications; region detection in videos where human actions occur [7], event detection in crowded scenes [8], abnormal behavior detection [9], human behavior discrimination [10], moving object detection according to direction of movement [11], cut-edit points detection in personal videos [12], shot boundary detection in video streams[13][14], action recognition for head-mounted cameras[15].

ST-Patch features are calculated as statistical values of pixel grayscale gradients within a small patch in videos. At each pixel within a video patch, let (P_x, P_y) be a space axis gradients of gray scale according to the x - and y -axes, let (P_t) be a time axis gradient of grayscale according to the t -axis, and let (P_x, P_y, P_t) be a spatio-temporal gradients, then two dimensional moments of gradients are represented by matrix M in Eq. (1).

$$M = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (1)$$

Matrix M is called a Gram matrix. Elements in the upper left 2x2 of matrix M do not contain any time component, and they represent appearance information within a video patch. Elements in the 3rd row and 3rd column of matrix M contain space components and time components, and they represent motion information. ST-Patch features are composed of elements of matrix M . Considering the symmetry of matrix M , ST-Patch features are denoted by six dimensional vectors in Eq. (2), which are composed of unique elements of matrix M .

$$v = [\sum P_x^2, \sum P_x P_y, \sum P_y^2, \sum P_x P_t, \sum P_y P_t, \sum P_t^2] \quad (2)$$

Table.1 lists example values of ST-Patch features from four patches; (1) is a patch on still tree branches, (2) is a

patch on waving tree branches, (3) is a patch on a road surface when sunlight is stable, and (4) is a patch on a road surface when sunlight is transiting. Note that (1) and (2), and (3) and (4) are located at the same position in images. The patch size is set to 15 pixels by 15 pixels by 5 frames. In Table.1, ST-Patch features $\{v_j\}_{j=1 \sim 6}$ have the following tendencies.

- Appearance components differ between tree and road without regard to motion. (compare $v_{1,2,3}$ between (1) or (2) and (3) or (4).)
- Motion components increase according to temporal change (compare $v_{4,5,6}$ between (1) and (2), v_6 between (3) and (4))
- Motion components differ from transitions of sunlight and waving of tree. (compare $v_{4,5,6}$ between (2) and (4))

Table 1 Example values of ST- Patch features

	v_1	v_2	v_3	v_4	v_5	v_6
	$\sum P_x^2$	$\sum P_x P_y$	$\sum P_y^2$	$\sum P_x P_t$	$\sum P_y P_t$	$\sum P_t^2$
(1) Still Tree	68.3	1.1	38.4	-0.2	0.1	0.6
(2) Waving Tree	66.5	-11.7	52.7	1.8	-1.3	1.7
(3) Road under Stable Sunlight	0.7	-0.1	0.6	0.0	0.0	0.0
(4) Road under Transiting Sunlight	1.4	-0.1	0.7	-0.4	0.0	1.2

Appearance Components
Motion Components

3. Background subtraction using ST-Patch features

In this section, we describe the method for detecting moving objects in dynamic backgrounds using a background model based on ST-Patch features. In Section 3.1, we describe how to generate a background model using Gaussian mixture distributions of ST-Patch features, to cover several states of background changes. In Section 3.2, we describe how to detect moving objects by comparing ST-Patch features of an input video with the background model. In section 3.3, we describe how to gradually update the background model, to follow translations of changes in a background.

3.1. Generation of background model

Some changes in background are constant, and others are not. An example of the former is the movement of an escalator, which keep the same direction and same velocity during operation. Examples of the latter are random movements of waving tree branches according to the wind, and rapid sunlight transitions according to the fast movement of clouds. In addition, a change in backgrounds could be a combination of several kinds of changes in

backgrounds. Hence, our proposed method for generating a background model use Gaussian mixture distributions of ST-Patch features, to cover background changes, which possess several variations. We denote a Gaussian mixture distribution in Eqs. (3) and (4).

$$p(x|C) = \sum_{i=1}^M w_i p(x|C_i) \quad (3)$$

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \sigma_i^{-1}(x-\mu_i)\right\} \quad (4)$$

In Eqs.(3) and (4), $p(x|C)$ is the probability density of probability variable x belonging to a Gaussian mixture distribution C , $p(x|C_i)$ is the probability density of x belonging to i -th normalized distribution represented as C_i , w_i , μ_i , and σ_i , each of which is respectively weight factor, mean, and standard deviation of C_i . We could assume with our proposed method that each feature $\{v_{ij}\}_{i=1 \sim 6}$ independently follows individual Gaussian mixture distribution. Each parameter w_i , μ_i , and σ_i should be calculated previously from examples of background videos using an expectation-maximization (EM) algorithm.

Considering that background changes generally differ according to locations within an image, our proposed method calculates the above parameters of a Gaussian mixture distribution at each block which partitions images evenly by prescribed size.

3.2. Detection of moving objects using background model

When an input video is given, our proposed method extracts ST-Patch features in each block within a detecting area at every frame, and detects moving objects by evaluating the number of blocks that does not follow the background model described in Section 3.1. Below we describe every step of the processes mentioned above, and show an overview in Figure.1.

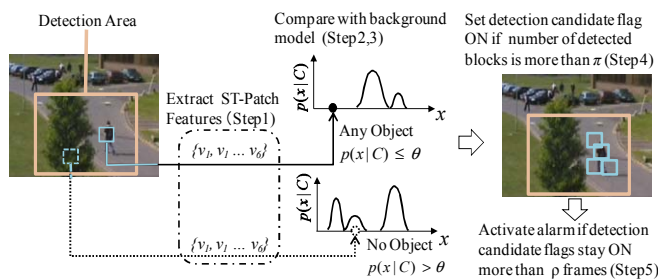


Figure 1 Overview of moving object detection using background model based on ST-Patch features

Step 1: ST-Patch features are calculated at each block of each frame in input videos.

Step 2: A probability density belonging to the background (Eq. (3)) is compared with the threshold θ (Eq.(5)), assuming that each feature of the ST-Patch feature $\{v_{ij}\}_{j=1 \sim 6}$ is independent probability variable x .

$$p(x|C) > \theta \quad (5)$$

Step 3: If every ST-Patch feature $\{v_{ij}\}_{j=1 \sim 6}$ follows Eq. (5), the block is discriminated to belong to the background. If not, the block is discriminated to belong to the moving objects.

Step 4: When the number of blocks is more than the threshold π which was set according to the size of moving objects, the moving object candidate flag is set ON.

Step 5: When the flag in Step 4 stays ON continuously more than the threshold ρ , which was set according to frame-wise detection precision, an alarm is activated against one or more moving objects in the detection area.

3.3. Update of background model

It is difficult to generate a background model in advance that wholly covers changes in a background, because the changes can translate as time passes. For example, in an outdoor scene sunlight intensity changes gradually but widely as time passes and the movement of tree branches can randomly change according to the strength and directions of the wind. Therefore, our proposed method gradually updates background model during moving object detections.

When changes in a background are not constant, some changes might disappear and some new changes might arise as time passes. Also, similar changes are redundant and should be merged into one change on update. Therefore our proposed method applies the updating method below[3], which has conditional branches (Delete / Add / Merge). Figure.2 shows an overview of the updating method.

Step 1: It is discriminated whether or not ST-Patch features belong to the background using the method of Step.2 in Section 3.2.

Step 2a: When ST-Patch features belong to the background, a normalized distribution nearest to them is updated with a pre-set updating ratio α using Eqs. (6) and (7).

$$w_i' = (1 - \alpha)w_i + \alpha \quad (6)$$

$$\mu_i' = (1 - \eta)\mu_i + \eta x$$

$$\sigma_i'^2 = (1 - \eta)\sigma_i^2 + \eta(x - \mu_i)^T(x - \mu_i)$$

$$\eta = \alpha p(x | C_i) \quad (7)$$

Step 2b: When ST-Patch features do not belong to the background model and the number of normalized distributions M are less than M_{max} , a new normalized distribution is added whose mean is x , weight and standard deviation are pre-set constants w_o and σ_o .

Step 3: If whichever condition below is satisfied, a redundant or minute normalized distribution is deleted.

- The means of two distributions are close. (Delete one with smaller weight.)
- Weight is less than threshold w_{min}

Step 4: Each weight $\{w_{ij}\}$ is normalized so that the sum of weights is 1.

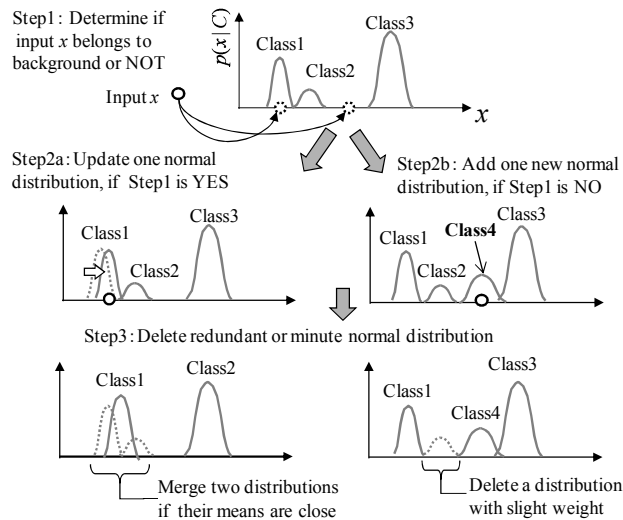


Figure 2 Overview of updating background model based on ST-Patch feature in [3]

4. Experimental Results

We describe the experimental results of our proposed method below in series; outline of experiments, evaluation videos in three scenes, example results, two procedures of quantitative evaluations.

4.1. Outline of experiments

To evaluate our proposed method, we conducted experiments using three scenes. To compare the

performance with a conventional method, we simultaneously evaluated a method applying grayscale version of normalized distance, which is an ordinary method using appearance features within a patch in images [5]. A method for applying normalized distance has equal conditions with our proposed method; block partitioning in images and subsequent processes after the block-wise detection. (see Steps 4. and 5. in Section 3.2.) In addition, the method applying normalized distance created a background image from an average image of the same video used for a method applying ST-Patch to generate a background model, and update a background image at the updating rate α (Eq.(7)) using infinite impulse response (IIR) filter. Parameters were set as follows; input videos were at 320 pixels height, 240 pixels width and a 30 fps frame rate, the patch size was set for the method applying ST-Patch at 15 pixels width x 15 pixels height x 5 frames, for the method applying normalized distance at 15 pixels width x 15 pixels height. The updating ratio α (Eq.(7)) was set to 0.01. There were three normalized distributions for generating a background model, and a maximum of 10 on updating. (see M_{max} in Step 2b in Section 3.2)

The evaluation indexes were recall and precision calculated with Eq. (8) using summary math at each frame of the following values; TP (True Positive: a true alarm for one or more objects), FP (False Positive: a false alarm for no object), FN (False Negative: a miss alarm for one or more objects), and TN (True Negative: no alarm for no object). We also added F-measure to the evaluation indexes, which is calculated by the harmonic mean between recall and precision.

$$\text{Recall} = \frac{\sum \text{TP}}{\sum \text{FN} + \sum \text{TP}} \quad (8)$$

$$\text{Precision} = \frac{\sum \text{TP}}{\sum \text{FP} + \sum \text{TP}}$$

4.2. Evaluation videos

4.2.1 Evaluation video: outdoor scene

The first evaluation video was an outdoor scene of a published video from PETS 2001[16]. There was a tree in the middle of the image, and intermittent wind gusts which caused inconstant waving of leaves and branches. There were also rapid movements of clouds, and they caused frequent and intensive transitions of sunlight. The waving of leaves and branches of a tree was a local change in the background and transitions of sunlight were global changes in the background.

This video had two parts, training and testing. We generated a background model using the training part and evaluated performance of moving object detection using the testing part. To generate a background model we used 1179 frames extracted from the training part of the video in which there were no pedestrians. In Figure.3, we show sample frames of the video used to generate the background

model. During evaluation of moving objects detection, there were a total of 6255 with 1359 frames with efficient size of one or more pedestrians. We regarded these 1359 frames as frames with moving objects and the remaining frames as frames without moving objects. We set detection area around the tree where the most intensive change occurred in the scene, and we regarded passing pedestrians near the tree as moving objects.

Note that the magnitude of the waving of the tree was generally weaker in the training part than in the testing part. In the former most of the frames with waving were limited to leaves, whereas in the latter some frames with waving involved branches in addition to leaves.



Figure 3 Example frames of outdoor scene used to generate background model

4.2.2 Evaluation video: ceiling lights scene

The second evaluation video was a published video in which there was frequent turning on and off of ceiling lights [17]. There were several ceiling lights, and one or more of them was changed on or off at one rotation of them. The rotation of ceiling lights caused rapid changes of the pixel grayscale in the images. In this video, the entire luminance of the room changed according to the number of ceiling lights turned on, and the luminance of each three dimensional structure in the room changed according to locations of ceiling lights turned on and the surface direction of the structure. The luminance changes of each structure were local changes in the background and the luminance changes in the entire room were global changes in the background.

The anterior 655 frames of the video with no people were used to generate a background model. In Figure.4, we show sample frames of this video for generating a background model. We set the detection area around the rod-shaped structure where the most intensive change occurred in the scene, and we regarded passing people near the structure as moving objects.



Figure 4 Example frames of ceiling light scene used to generate background model

In evaluating moving objects detection, we used the posterior 1981 frames of which 99 frames with efficient size of people. We regarded the 99 frames above as frames with moving objects and the remaining as frames without moving objects.

4.2.3 Evaluation video: escalator scene

The third evaluation video was a laboratory escalator scene in which certain parts of passengers' bodies protrude outside the hand-rail when the passenger leaned over. These body parts moved according to the velocity of the escalator. There were large windows facing outside, and sunlight illuminate the escalator and regions around it. The sunlight through the windows casted shadows of passengers onto the metallic region outside the hand-rail, and these shadows moved according to the velocity of the escalator. The transitions of sunlight caused changes in brightness and unevenness of the metallic region outside the hand-rail. On this metallic region, moving shadows of passengers were local changes in the background and the entire brightness and unevenness in the transit caused by sunlight were global changes in the background.



Figure 5 Example frames of escalator scene used to generate background model

A segment of the video was used to generate a background model, in which a passenger not leaning over the hand-rail. In Figure.5, we show example frames of this video segment used to generate a background model.

We set the detection area on the metallic region outside the hand-rail, and we set body parts of passengers outside the hand-rail as moving objects.

In evaluating moving objects detection, there were a total 3250 of which 215 frames contained efficient size of such protruding body parts. We regarded the 215 frames above as with moving objects and the remaining frames as without moving objects. The body parts of passengers were heads, arms, and upper bodies, and there were one or more passengers such riding escalator at a time.

4.3. Example results

4.3.1 Example results from outdoor scene

Figure.6 shows example results from the outdoor scene described in Section 4.2.1. In each image of columns (b) and (c), a rectangle around the tree in the middle of the

image indicates the detection area, and the small blocks within this detection area indicate the detected blocks. In the upper left corner of each image in these columns, an evaluation of alarm (False Negative, True Positive, and so on) of the frame is indicated.

The images in each column show the following frames; Row (1) shows a frame in which the waving of leaves was becoming stronger according to occurrence of the wind gusts. Row (2) shows a frame when the waving of branches was becoming stronger than the one shown during the generation of the background model. Row (3) shows the frame in which strong waving of branches kept 289 frames after (2). Column (4) shows a frame when sunlight rapidly weakened. Note that there was no moving object (no pedestrian) in any of these rows.

The method applying normalized distance ended up with four false positives in rows (1) to (4). On the other hand, the method applying ST-Patch achieved three true negatives in rows (1), (3) and (4), though it gave one false positive in row (2). The method applying ST-Patch, the number of false positive blocks gradually declined as the frames passed from (2) to (3) and the occurrences of false positive frames also gradually declined. This decline in false positive frames could be an effect of updating the background model described in Section 3.3.

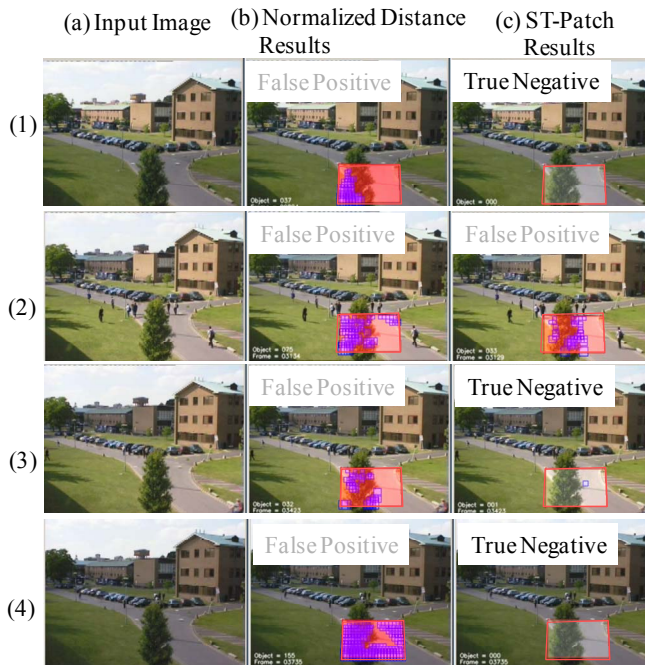


Figure 6 Example results in outdoor scene

Other tendencies in the experiment results of the first video were described below. The method applying ST-Patch tended to give false negatives when pedestrians walked beside the tree. The method applying normalized

distance tended to give false negatives when pedestrians walked beside the tree, and when the grayscales of pedestrians were close to that of the road surface.

4.3.2 Examples results from ceiling lights scene

Figure.7 shows example results in the ceiling light scene described in Section 4.2.2. In each image of column (b) and (c) of Figure.7, a rectangle around the rod-shaped structure in the middle of the image indicates the detection area.

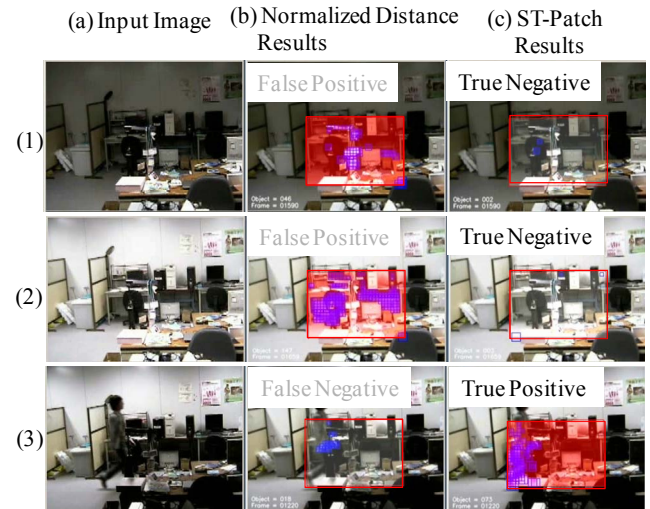


Figure 7 Example results in ceiling lights scene

Images at each row show the following frames; Row (1) shows a frame just after the lights, other than the ones illuminating the center area of the image, were turned off. Row (2) shows a frame just after all lights that had been turned off in row (1) were turned on. Row (3) shows a frame in which a person with low contrast passed under low illumination.

The method applying normalized distance ended up with two false positives in rows (1) and (2) and one false negative in row (3). On the other hand, the method applying ST-Patch achieved two true negatives in rows (1) and (2) and one true positive in row (3).

Other tendencies in the experiment results of the second video were described below. The method applying normalized distance gave five successive false positives out of eight after any ceiling light was turned on or off. On the other hand, the method applying ST-Patch gave two false positives out of eight during just a few frames after any ceiling light was turned on or off.

4.3.3 Example results from escalator scene

Figure.8 shows example results from the escalator scene described in Section 4.2.3. In each image of columns (b) and (c) of the figure, a belt-like quadrangle indicates the detection area, which was located along the metallic region

outside the hand-rail and stretching vertically from the bottom to the middle of the image.

The images in each row show the following frames; Row (1) shows a frame in which several passengers were on the escalator and did not lean over the hand rail. Row (2) shows a frame before a passenger leaning over the hand rail reached the detection area. Row (3) shows a frame in which an arm of a passenger protruded outside the hand-rail.

Note that forms of shadows in rows (1) and (2) were different from those in Figure.5 used to generate the background model, due to the differences in the number or posture of passengers. And note that in row (3), the grayscale of the skin on the arm and the metallic background region were similar, and the contrast between them was low.

The method applying normalized distance ended up with two false positives in rows (1) and (2) and one false negative in row (3). On the other hand, the method applying ST-Patch achieved two true negatives in rows (1) and (2) and one true positive in row (3).

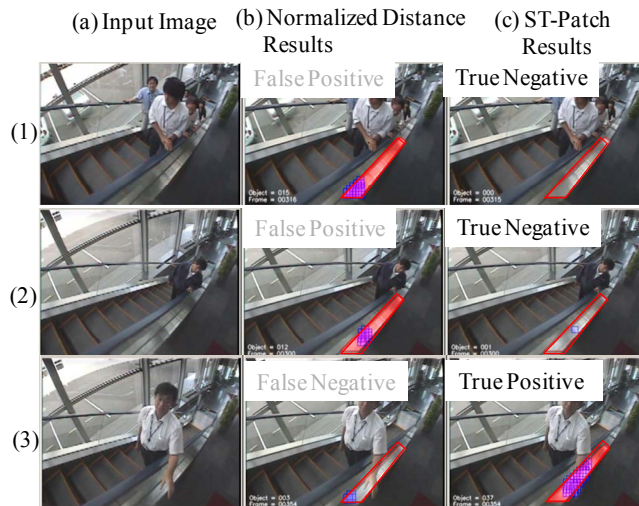


Figure 8 Example results in escalator scene

4.4. Quantitative evaluations

4.4.1 Evaluation indexes from examples of results

Table.2 shows 3 evaluation indexes (precision, recall, F-measure) calculated from experimental results in three scenes described in Section 4.3. In Table.2, the method applying ST-Patch achieved better indexes in any scene out of three than the method applying normalized distance especially in recall and F-measure. (In Table.2 from top to bottom the former method surpassed the latter method 38, 68, 46% by precision and 32, 55, 32% by F-measure.)

Considering there were several local and global changes in backgrounds through three scenes, these comparison of

indexes show that the former methods had much better performance than the latter method as to moving object detection under dynamic backgrounds.

Table 2 Evaluation indexes from example results

	Normalized Distance (%)			ST-Patch (%)		
	Recall	Precision	F-Measure	Recall	Precision	F-Measure
Outdoor	76	41	53	92	79	85
Ceiling Light	64	26	37	91	94	92
Escalator	78	34	47	90	70	79

4.4.2 Evaluations of precision recall curves

To comprehensively evaluate both methods applying ST-Patch and normalized distance, we calculated precision recall curves from experimental results in three scenes described in section 4.3. Precision recall curve is a characteristic curve that connects several operating points on a graph which possess recall-axis and precision-axis. The more a precision recall curves goes upper right in graphs, they show the better performances with regard to both precision and recall.

In section 4.4.2, we changed operating points adjusting a block-wise detection threshold respectively for the method applying ST-Patch (θ in Eq. (5)) and for that applying normalized distance. Besides, we changed frame-wise detection threshold referring to the number of detected blocks (π in Step4 in Section 3.2).

Three precision recall curves are shown in Figure.9. In each graph in the figure, a curve of the method applying ST-Patch is located highly upper right to that applying normalized distance over a wide range from precision approximately 100% to recall approximately 100%. These tendencies of precision recall curves show that in the three scenes the method applying ST-Patch had much better performance over wide range of operating points than that applying normalized distance.

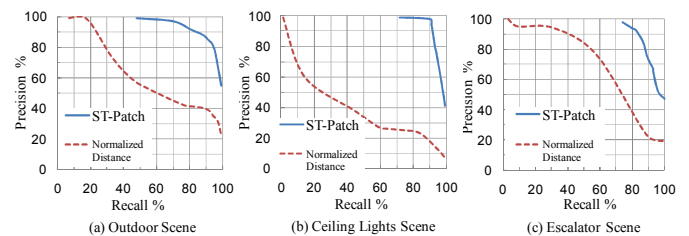


Figure 9 Precision recall curves for three scenes

5. Conclusion

We proposed a method for detecting moving objects with a background model that covers local and global changes in backgrounds using a spatio-temporal texture called ST-Patch features, which describe motion and appearance at the same time. Our proposed method performed better than the method applying normalized distance, which is an example of conventional features that describe appearance only, in three scenes.

However, our proposed method only determines whether or not there is any moving object within a detection area, and cannot extract any information about moving objects. For future work, we will enhance our proposed method to extract such information as position, size or trajectory from moving objects in dynamic backgrounds.

References

- [1] C.Stauffer, W.Grimson. Adaptive background mixture models for real-time tracking. In Proc. CVPR, 1999.
- [2] A.Elgammal, D. Harwood, L. Davis. Nonparametric Model for Background Subtraction. In Proc. ECCV, 2000.
- [3] A. Shimada, D. R. Taniguchi. Increment and Decrement of Gaussians in Adaptive Mixture-of-Gaussian Background Models. In Proc. MIRU, pp.746-751, 2006
- [4] T. Sato. S. Kaneko. Y. Niwa. Y. Kazuhiko. Robust Object Detection by Radial Reach Filter. In IECE Trans, J86-DII, 616-624, 2003.
- [5] S. Nagaya. T. Miyatake. T. Fujita. H. Ueda. Moving Object Detection by Time-Correlation-Based Background Judgment Method. In IECE Trans, J79-D-II, (4), 568-576, 1996.
- [6] T. Tanaka. A. Shimada. R. Taniguchi. T. Yamashita. D. Arita. Adaptive Background Modeling Considering Spatio-temporal Features for Object Segmentation. In Proc. MIRU, 127-134, 2009
- [7] E. Shechtman. M. Irani. Space-Time Behavior Based Correlation. In Proc. CVPR, 2005.
- [8] Y. Ke. R. Sukthankar. M. Hervert. Event Detection in Crowded Videos, In Proc, ICCV2007, pp.8-15, 2007
- [9] Y. Murai H. H. Fujiyoshi. M. Kazui. Abnormal Motion Detection at Escalator Scene based on Spatio-temporal Features, In Proc. PRMU, pp.247-254, 2008
- [10] M. Kazui, M. Miyoshi, S. Muramatsu, H. Fujiyoshi, Incoherent Motion Detection using a Time-series Gram Matrix Feature. 19th International Conference on Pattern Recognition(ICPR), pp.1-5, Dec. 2008
- [11] Y. Murai. H. Fujiyoshi. T. Kanade. Classifying Direction of Movement and Segmenting Regions of Objects by Using Space-Time Patches. In Proc. CVIM, (1) ,pp.21-31, 2008
- [12] T. Furukawa, H. Fujiyoshi, A. Nomura, A Method for Estimating Cut-Edit Points in Personal Videos, In Proc. ICME 2010
- [13] Y. Murai, H. Fujiyoshi, "Shot Boundary Detection Using Co-occurrence of Global Motion in Video Stream", In Proc. ICPR 2008
- [14] Y. Murai, Y. Murai, H. Fujiyoshi, Shot Boundary Detection using Global and Local Changes, In Proc. CVIM (159), pp. 217-224, 2007
- [15] T. Furukawa, H. Fujiyoshi, Action Recognition using ST-patch Features for First Person Vision, In Proc. PRMU, pp.53-58, 2010
- [16] Published video data: <ftp://pets.rdg.ac.uk/PETS2001/>
- [17] Published video data: <http://limu.is.kyushu-u.ac.jp/dataset/>