Probabilistic Method to Determine Human Subjects

for Low-Resolution Thermal Imaging Sensor

Yongwoo Jeong, Kwanwoo Yoon, KyoungHo Joung
Applied Thermal Imaging Lab
Samsung S1 Corporation
Seoul, South Korea

***Abstract*— In this work, we present a method of determining human subjects via a low-resolution thermal imaging sensor. Since the image quality of the low-resolution thermal imaging sensor could be suffering from heat signatures and recognizable patterns of human subjects are unable to be determined due to resolution issues, it is recommended to employ a probabilistic method. This paper presents how human subjects can be expressed in terms of pixel size, standard deviation, label movement, vector tracking, label lifetime and a rewarding system based on those. Various pre and post-image processing methods will be covered including background collection, Gaussian filtering, segmentation, local/global adaptive threshold and background learning.**

Keywords*— *probabilistic; patterns; human; thermal; low; resolution; segmentation; recognition; adaptive; threshold

I. INTRODUCTION

Thermal imaging sensors have been adopted in various environments to detect human subjects and heat signatures [1-2]. The modern method to detect human subjects is based on feature detection. Thi Thi Zin *et al* have suggested a method to detect human subjects with both a thermal imaging sensor and a visual camera; in this work, they were able to detect human features, such as head size, subject height, location of torso and legs, from both infrared and visual cameras [7]. Weihong Wang *et al* also proposed the shape context based Adaboost cascade method, which proved to be useful in detection of moving pedestrians, for human detection in thermal images [3, 8].

Thermal imaging cameras used in their works were high performance thermal imagers and there is no doubt that a resolution of at least 160 by 120 pixels (below LD scale) was used to perform the reported feature extraction methods [7-8].

However, when considering cost-effectiveness of in-door monitoring solutions, it is preferable to use 'extremely' low resolution thermal imaging sensors instead of employing below LD to HD scaled ones for short-ranged, spot based surveillance systems; refurbished, below LD scaled thermal imager could still cost up to multiple thousands of dollars.

For the purpose of human detection, since many heat sources are involved in human activities, it is required to analyze how a heat signature can determine a human subject. As the complexity of an image is increased, one may have to utilize a probabilistic method and/or any alternative method(s) that could improve the performance of the surveillance system. Previously, Stolkin, R. *et* al had proposed a Bayesian based rapid background adaptation method to track human subjects in the foreground [9]. This work gave a method of improving the ability to detect a subject with both visual and thermal information in which a probabilistic method based on Bayes' rule was employed to determine target-like pixels [9]. As a result, a cross reference between color and thermal information was key to the contribution of the overall performance [9]. However, if no other cross reference is available, the single sensor system is a liability and heavily depends on its performance and cost. As mentioned in other authors' previous works in [7-8], even below LD scaled thermal imagers are able to extract the feature of human subjects but extremely low-resolution thermal imagers require a different approach. Here we propose the idea of a probabilistic method with multiple pre and post image processing techniques which affect the performance of detection of human subjects using heat signatures from a low resolution thermal imaging system.

II. SENSORS AND PREPROCESSING

A. SensorModule

The thermal imaging sensor we used for this work is a Panasonic Grid-EYE 8 by 8 sensor which come in two versions of differing gain; temperature accuracy of the high gain version is typically ±2.5°C, detection distance is max. 5 m, and viewing angle is 60°. The low gain version of the Grid-EYE may be good for fire detections but since human body temperature is usually around or just above an ambient temperature, the high gain version was used for this project. The high gain product used for this work requires 5.0 V DC operating voltage and a temperature range of 0°C to 80°C (32°F to 176°F). The frame rate is typically 10 frame/sec. The developed sensor module is capable of Ethernet communication for firmware based control, storing the number of human detections and the time of detection for a month. The MCU is STM32F205RC; it has a 256k flash ROM, 96k SRAM, and runs at 120 MHz clock speed.

978-1-4799-2179-9/14/$31.00 ©2014 IEEE

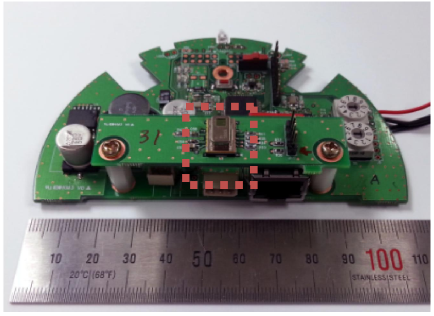


Fig. 1. The developed sensor module; the width is about 11 mm. Top board contains the Panasonic Grid-EYE sensor (red dot line box).

Since the raw thermal signal from the sensor is not like below LD to HD scaled thermal imaging sensors, we cannot use the same techniques as reported previously by other authors [2-3 and 7-8]. The raw image does not report any human features and, even worse, it inevitably contains a temperature fluctuation caused by temperature inaccuracy which creates massive temperature waves in the image. Therefore, the raw image must be processed through image preprocessing and segmentation, which will be discussed in the next section.

B. Image Preprocessing

First of all background information, which is to be subtracted from the current scene [1, 9], is collected for 100 frames (10 sec) after a 29 by 29 pixel sized interpolation process. It is then fed into a process which calculates a running average of every three frames. For these steps in the simulation process, the nVidia CUDA enabled GPU was used; both background and current raw data were fed into GPU. The final product however will not employ the GPU computing since it must run in a firmware mode.

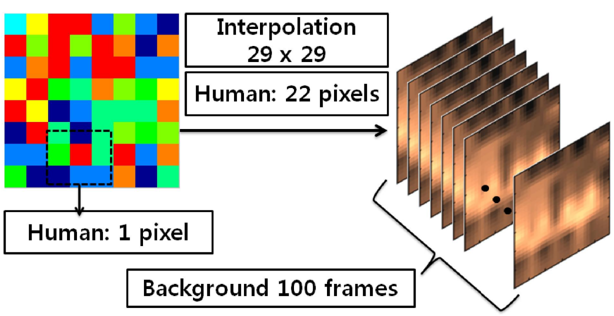


Fig. 2. Interpolation and background collection. A human subject can only occupy 1 pixel if the sensor is mounted on 3 m height ceiling; without interpolation, it is almost impossible to detect any human features.

The processing gain for a moving average over 3 frames is set to detect the current scene clearly while not losing the first and second ones as well, which is shown in equation (1),

$$D = \sum_{i=1}^{3}(C_i + G_i \sum_{k=n-2}^{n} R_k) \quad (1)$$

where D is the image data after the moving average calculation, C is the current data, G is a predefined processing gain array of values [0.15, 0.35, 0.5], and R is the raw data over 3 frames. After that, the current raw data is unloaded from GPU memory and converted to a 29 by 29 pixel sized frame via interpolation. Since the temperature fluctuation due to temperature inaccuracy is noise of a high frequency, in order to remove it Gaussian filtering is used on the interpolated image. The filtering window size was 10 by 10 pixels and the suppression level (sigma) was 3 for the Gaussian filtering process. The collected 100 frame background is then subtracted from the filtered image. The foreground image is selectively labeled with global & local adaptive thresholds that measure standard deviation, maximum, minimum, and average values of the current scene. If the standard deviation of the filtered image is too low, it is to be considered as thermal noise and subjected to removal. In this case, the adaptive threshold is globally applied to remove thermal noises. The global adaptive threshold (AT) is as appears in equation (4),

$$W = 0.025 + \frac{0.85}{(\mathrm{Max}-\mathrm{Mean}+1)} \quad (2)$$

$$ASD = SD \times 0.7 \quad (3)$$

$$AT = (\mathrm{Max} - \mathrm{Mean}) \times W + \mathrm{Mean} - ASD \quad (4)$$

where W is weight, Max and Mean are the maximum and the mean value of the current scene respectively, SD is standard deviation of the segment and ASD is adjusted SD. A local adaptive threshold is also applied using a probabilistic method (to be discussed in next section) to detect a hidden label that has previously been reported as human but is no longer visible in the current frame.

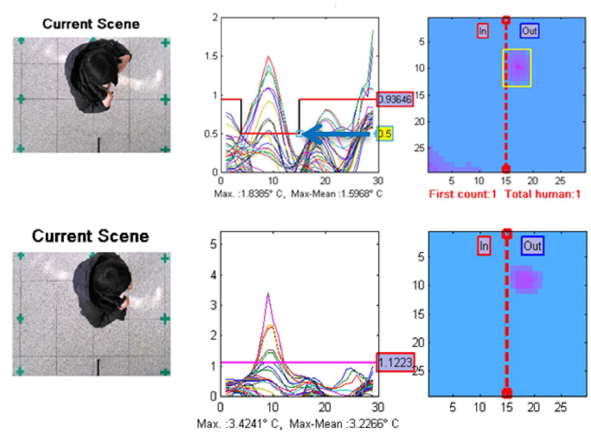


Fig. 3. Top row shows an example of the local adaptive threshold with global adaptive threshold and the bottom row shows the same scene without local adaptive threshold. The human label has not been detected at the 125th frame and is still undetected at 127th frame (bottom row); left column: current scene, center column: heat signal with the global (red line) and local (red line indicated with blue arrow) adaptive thresholds, right column: processed heat signature.

The local threshold level is applied and adjusted based on where the heat signature is observed at the previous frame. It is then able to track the hidden heat signature over frames as long as the heat signature is captured by the local adaptive threshold, as shown below in figure 3.

C. Segmentation

The segmentation process includes generating the temporary segments, overlapping & using a vector tracking calculation between old and temporary segments, and keeping segment history over 10 frames. Whenever the new heat signature that passes preprocessing has been generated, it becomes a temporary segment which is used to determine whether it is connected to any old segments that already exist. Its first observed position is used to track whether the movement has the characteristics of a human subject; a human subject cannot *move* erratically between frames, rather as it moves it leaves a residual path over multiple frames. If the label disappears, then a vector tracking idea is employed to track the previously visible label.

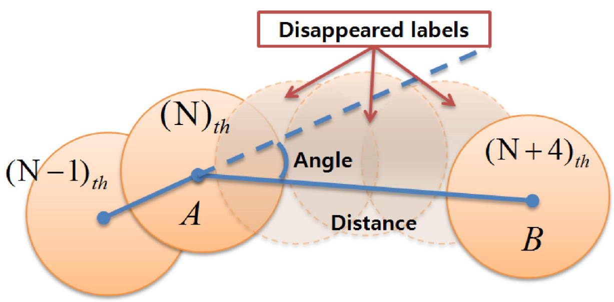


Fig. 4. Segment connectivity calculations; whether the new segment and old segment are connected via a certain path is computed, and a new segment is overlapped over the old segment.

Figure 4 shows the vector tracking and pixel overlapping calculation to determine whether the old segment is connected with a new segment. Even if a new segment is found but the overlapping condition is not satisfied, the vector tracking method will decide whether the new segment is related to previous ones by calculating whether it is on a viable path. The maximum angle of the vector required to determine this tracking performance is 60° and the minimum pixel distant is 10 pixels; for example if a new label *B* appears in the next frame located 20 pixels from the label *A* in figure 3, it isn't considered to be connected. Also, if a label appears and disappears intermittently and it still keeps appearing in a succeeding path, it could be the same label that is connected through frames.

It is important to keep the history of segments. If a segment is created, its characteristics, such as pixel count, center position, maximum, minimum, average temperature, standard deviation, and probability to be a human subject are stored in the segment structure and used for computation whenever a new segment is created in the next frame. The newly created segment will inherit the parental information if it is considered as a connected segment. Merging and separation of segments are monitored via this parental information. Figure 5 depicts how parental information is inherited.

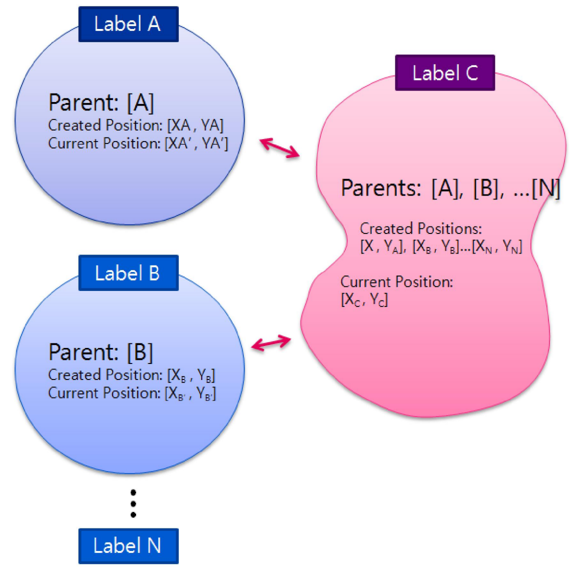


Fig. 5. Segment separation and merging; segment parent information, created positions and current positions are inherited to daughter segments.

III. PROBABILISTIC METHOD AND POSTPROCESSING

A. Probabilities

As discussed previously, a raw thermal signal does not contain significant characteristics to uniquely identify a human subject. But one can clearly see the differences between electric heat sources and a human; human body temperature does not exceed 40°C and has a gradually increasing and decreasing heat pattern over the body. On top of this, a human subject has a certain pixel size, creates a route when it moves, and its body temperature does not fluctuate significantly between frames.

Therefore, here we propose the idea of using a probability calculation to determine if the segment is a human subject. The probability of each segment being a human subject is based on the characteristics mentioned above [4, 5]. The total positive probability consists of sub-probabilities based on pixel size, standard deviation, center position movement over 2 frames, vector angle of movement, and lifetime.

$$P_{px} = 20 \times \frac{\text{pixel count}}{\text{normal human pixel size}} \quad (5)$$

$$P_{SD} = 30 \times \frac{\text{SD of Segment}}{\text{normal SD of human label}} \quad (6)$$

$$P_M = 5 \times \frac{\text{measured movement}}{\text{normal human moved distance per 2 frames}} \quad (7)$$

$$P_V = 40 \times \frac{\text{3 frame vector angle}}{\text{normal human moving path angle}} \quad (8)$$

$$P_L = 5 \times \frac{\text{Life time}}{\text{normal human segment life time}} \quad (9)$$

$$TPP = P_{px} + P_{SD} + P_M + P_V + P_L + P_{TC} \quad (10)$$

where TPP stands for the total positive probability, P_{px} is a sub-probability of pixel size, P_{SD} is a probability based on the standard deviation, P_M is a probability calculated from the distance moved over 2 frames, P_V is a probability based on the vector angle measured over 3 frames, and P_L is a probability of the lifetime of labels.

Maximum values of sub-probabilities per frame are set 20, 30, 5, 40, and 5 respectively; The TPP will therefore have a maximum value of 100 per frame. Besides these sub-probabilities, if the segment shows a temperature consistency for 5 frames it will add an extra sub-probability P_{TC}, which will have a maximum value of 10. This is because thermal noises from the convection have a similar standard deviation and heat signal as human subjects but their temperature is not consistent over 5 frames.

A negative probability is also made to resolve the issue caused by any possible heat sources moving around the border of a scene. If the heat signature label does not move for multiple frames, it will also receive a negative probability. This leads us to a total negative probability

$$TNP = NP_B + NP_M \quad (11)$$

where TNR is the total negative probability, NP_B is the negative probability for the segment at border of the observed area, and NP_M is the negative probability for the segment that appears to be stationary. For each created segment, if the total probability (the sum of TPP and TNP) is equal to or greater than 100, it is considered to be a human subject.

The probabilistic calculation mentioned above is quite similar with the reinforcement learning algorithm that gives rewards based upon how the artificial intelligence behaves. However, this proposed idea does not use the *discount factor* concept of reinforcement learning. Instead, it utilizes the information collected through the segmentation. Connectivity between temporary segments and old segments are the key to deciding how the reward, which is being proposed as a probability of being human, can be inherited to the new, temporary segment. Once the new segment is found to be connected over multiple frames, it will inherit the probability. If its probability is greater than or equal to 100 then it won't calculate the probability again, otherwise it will continuously add the current probability to the inherited one.

Therefore, this proposed method is quite similar to reinforcement learning based on unsupervised training; we collect every single feature of segments then reward them based upon how they are observed in consecutive frames. Then cumulative reward (here we call it total probability) is used to make a decision.

B. Background Learning as Image Postprocessing

Once the probabilities of segments have been calculated and reported, their size and location are merged in the background at different ratios so that new background is generated based on what is observed in the current scene; the old segment with a high enough probability to be regarded as a human subject won't lose its old probability calculated at the last frame while the probability of new segment will be calculated based on the newly generated background [5, 6]. As a result, possible heat noises from electric equipment or convection will be merged to the background quickly. Figure 6 describes how the background learning algorithm works.

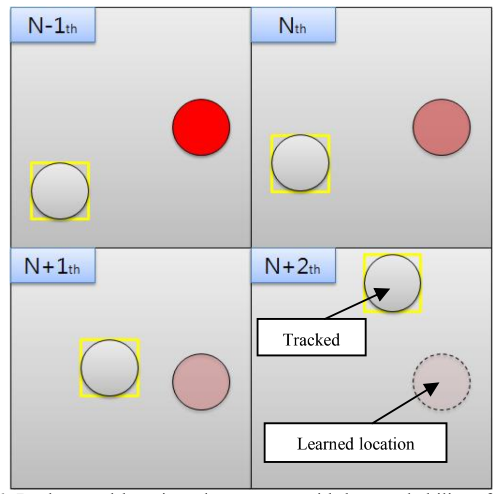
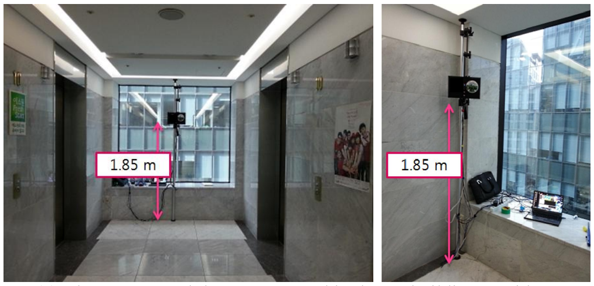


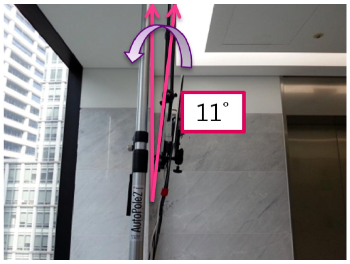
Fig. 6. Background learning; the segment with low probability of being a human subject will be merged to the background quickly while another segment known as a human subject is monitored and tracked.

IV. EXPERIMENTS AND RESULTS

A sensor system was deployed to prove the proposed idea of detecting only human subjects. The sensor was mounted at 1.85 m height and the angle was tilted toward the ground at 11° to observe a 1.75 m human subject. It was installed in the middle and at the corner of the S1 building corridor. A CMOS camera was also deployed on top of the thermal imager to capture the current scene as shown in figure 7. The corridor is entirely covered with marble stones and reflective materials except the ceiling; a heat signature from any heat source easily reflects on fine surfaces and it is easily noticed by thermal imagers as shown in figure 8(a). This scene was taken with a FLIR E40 thermal camera. If any heat source is located near the metal or marble wall, it reflects on surfaces of the wall, which creates a heat mirroring effect. The mirroring effect of the visual scene causes a significant performance drop of CMOS or CCD camera based intellectual surveillance systems. However, false detections due to the heat mirroring effect can be avoided with the proposed probabilistic method as depicted in figure 7(c) and 7(d).

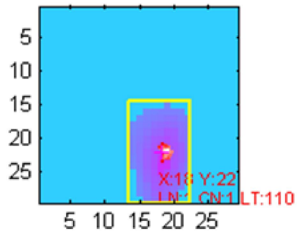


(a) The sensor module was mounted in the S1 building corridor



(b) The tilt angle was 11° to detect human subject.

(c) Captured scene



(d) Processed scene

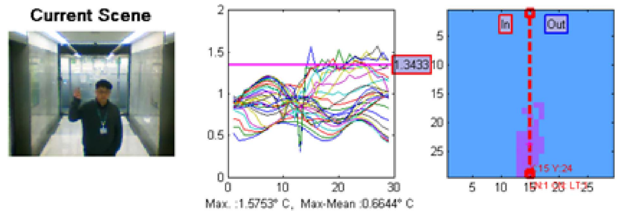
Fig. 7. Probabilistic method detects only the human's body heat signature and removes other segments caused by heat reflection (A) and hot cup (B).

The result shows only the human subject was detected while mirrored heat signature and other heat sources were subtracted and identified as non-human subjects.

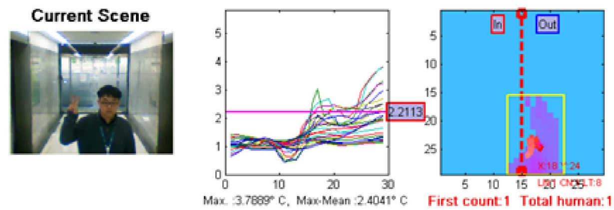
The sensor detection range is determined by whether the human subject is coming to the sensor or going further from it, and how much high ambient temperature is observed. If a human subject is moving out of the office room where ambient temperature is quite high compared to human body temperature, the human can be determined only at 2 m distance as depicted in figure 8(b) and 8(c). Before the tester was standing at 2 m distance, the human subject does not have a high enough probability.

However, when the human subject is moving further from the sensor, its maximum detection range, 5 m in the datasheet, could be increased a bit by using the local adaptive threshold as shown in the figure 8(d). The human subject was standing 5.3 m away from the sensor but was still detected and reported as a human subject thanks to the local adaptive threshold. This is due to the proposed probabilistic method that keeps its probability over segments; if the human label once detected and determined as a human, even smaller label can be determined as a human thanks to the local adaptive threshold.

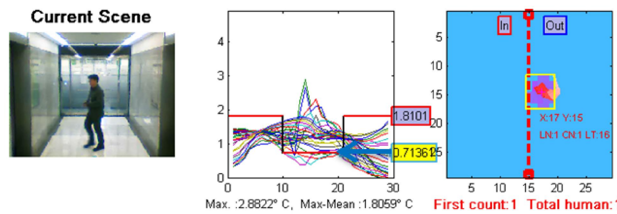
(a) Heat reflection on the wall and glass door; taken with FlIR E40



(b) The tester was at 3 m from the sensor



(c) The tester was at 2 m from the sensor

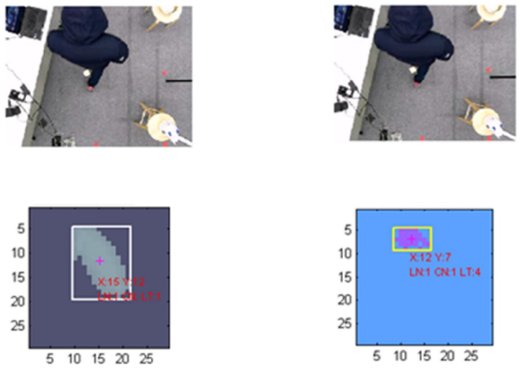


(d) The tester was moving further from the sensor

Fig. 8. Top: FLIR camera E40 (160 by 120 pixel resolution) was used to check the heat reflection on the marble wall and glass door. Second: the tester was coming toward the sensor (at 3 m). Third: the tester at 2 m distance to the sensor; the sensor is able to determine him as a human only at 2 m distance. Bottom: local adaptive threshold helped to increase the detection range when the tester is moving further from the sensor.

The proposed method was also tested to prove the ability to determine a human subject that appeared and disappeared intermittently but keeping the same viable path. In this experiment, the tester wore a thick jacket with hood and moved across the small heat sources.

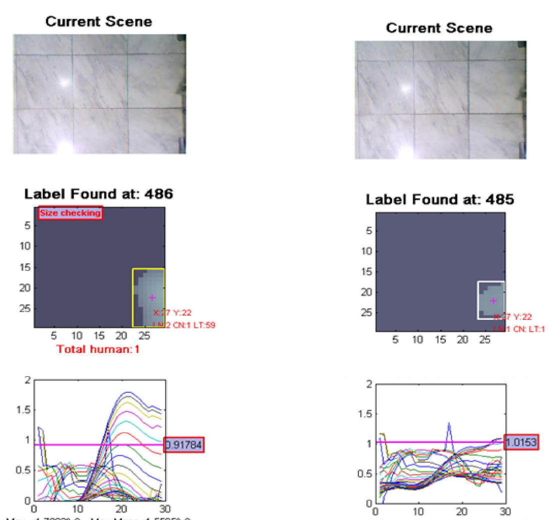
The figure 9 shows 2 columns of images. The top images indicate the live video feed captures and the bottom images are processed images. The vector tracking method keeps checking labels whether any label under probability of 100 is creating a viable path. When some label is appeared to be connected through frames, even though it disappeared and appeared intermittently, its probability is increased. Even though a lower adaptive threshold was applied at figure 9(a), which resulted in a larger label, it is not able to determine the tester as a human subject (white rectangle). However, thanks to the vector tracking method, it can detect the human subject despite of a relatively higher adaptive threshold as shown in figure 9(b) (golden rectangle).



(a) (b)

Fig. 9. Vector tracking method; (a) Without vector tracking (b) With vector tracking.

In order to test the background learning algorithm, the sensor module was placed on the ceiling. The background learning algorithm is supposed to remove the heat noises caused by convection. The experiment was conducted in winter when a heavy warm air circulation was expected near the front gate. As depicted in figure 10, the sensor was placed where the convection was expected; without background learning, the heat signature caused by the convection was reported and incorrectly identified as a human subject – highlighted by the golden rectangle boundary in figure 10(a).



(a) (b)

Fig. 10. Background learning off and on comparison; (a) Without background learning. (b) With background learning.

However, the background learning algorithm helped to remove the heat signature and the segment was not reported as a human – shown by the white rectangle boundary in figure 10(b). The threshold levels shown at the bottom row of the figure 10 indicate how it was trained with the background learning method.

V. Conclusions

The proposed idea showed a probabilistic method with multiple pre and post image processing techniques for a low resolution thermal imaging sensor to determine whether detected heat signatures are human subjects. Preprocessing and segmentation provide the base structure then a probabilistic method calculates how likely a heat signature is to be a human subject. The probability of being a human subject that is calculated can be inherited as long as new segment's characteristics are considered to be connected. Even if a human segment temporarily disappears, the proposed idea is able to track it with the local adaptive threshold. Ambient heat sources such as warm air circulation can also be removed by using background learning.

References

[1] J. Davis; V. Sharma, "Robust background-subtraction for person detection in thermal imagery," IEEE Int. Wkshp. On Object Tracking and Classification Beyond the Visible Spectrum, 2004.

[2] P. Viola; M. Jones; D. Snow, "Detecting pedestrians using patterns of motion and appearance," In Proc. IEEE CVPR, New York, NY, volume 1, pp. 734741, 2003.

[3] Ju Han; Bhanu, B., "Human Activity Recognition in Thermal Infrared Imagery," Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on , vol., no., pp.17,17, 25-25 June 2005.

[4] W. Wang; J. Zhang; C. Shen, "Improved human detection and classification in thermal images," Image Processing (ICIP), 2010 17th IEEE International Conference on, vol., no., pp.2313,2316, 26-29 Sept. 2010.

[5] Pierard, S.; Lejeune, A.; Van Droogenbroeck, M.,"A probabilistic pixel-based approach to detect humans in video streams," Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on , vol., no., pp.921,924, 22-27 May 2011.

[6] Dawei Li; Lihong Xu; Goodman, E., "Real-Time Statistical Background Learning for Foreground Detection under Unstable Illuminations," Machine Learning and Applications (ICMLA), 2012 11th International Conference on , vol.1, no., pp.468,472, 12-15 Dec. 20122003.

[7] Thi Thi Zin; Hideya Takahashi; Takashi Toriu; Hiromitsu Hama, (2011). Fusion of Infrared and Visible Images for Robust Person Detection, Image Fusion, Osamu Ukimura (Ed.), ISBN: 978-953-307-679-9, InTech, Available from:

http://www.intechopen.com/books/image-fusion/fusion-of-infrared-and-visible-images-forrobust-person-detection

[8] Weihong Wang; Jian Zhang; Chunhua Shen, "Improved human detection and classification in thermal images," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.2313,2316, 26-29 Sept. 2010

[9] Stolkin, R.; Rees, D.; Talha, M.; Florescu, I., "Bayesian fusion of thermal and visible spectra camera data for mean shift tracking with rapid background adaptation," Sensors, 2012 IEEE , vol., no., pp.1,4, 28-31 Oct. 2012