

# Performance Benchmark in Febrile Mass Screening Detection

Siti Sofiah<sup>1</sup>, Kamarul Hawari<sup>2</sup> and Sabira Khatun<sup>3</sup>

<sup>1</sup>Universiti Malaysia Pahang, Kampus Pekan, Pahang, Malaysia

<sup>2,3</sup>Universiti Malaysia Perlis, Pauh Putra Main Campus, Arau, Perlis, Malaysia

<sup>1</sup>sofiahradzi@yahoo.co.uk, <sup>2</sup>kamarul@ump.edu.my, <sup>3</sup>sabira@unimap.edu.my

## Abstract

Recently, detecting and tracking people using infrared sensors in public surveillance has been given attention by many researchers since the global outbreak of severe acute respiratory syndrome, and machine vision is bound to play an important role. In machine vision, sliding window approach has appeared as most promising approach. Since Papageorgious et al who proposed the first sliding window detectors until Dalal et al who have come out with a large gains detector based on histogram adoption, the ideas of sliding in human detections continue to serve as a root for modern detectors. While there is increasing research in thermal spectrum, very little research focus on the effectiveness of state of art detectors in febrile mass screening application. The exposed area of head-to-shoulder is an important body part for region of interest detection priorly before the temperature of febrile person is measured by the thermal camera. In addition, some of the thermal datasets used in previous research are more fit to surveillance or safety applications where the targets are mostly far-scales. One of the challenges in detecting the pedestrian in thermal images is the nature and quality of image in infrared spectrum as well as the real crowd situation in public area that cause occlusion. Therefore in this paper, we are interested to evaluate the top three detectors' performance on thermal images taken during fever screening in Kuala Lumpur International Airport. We also analyze the best approach to be adopted in the detection using a new context of training and evaluation.

**Keywords:** Thermal, detection, HOG, Haar, LBP, mass screening, evaluation, data set, fever, KLIA, crowd

## 1. Introduction

Since the outbreak of SARs, various countries has set up thermal imaging system in airport to find potential suspect with elevated body temperature as a first mass screening before further examined by medical doctors. Hence, a surge of interest in researching the best methods and the feasibility of the thermal camera installed for screening has been carried out [1]–[9]. While the debate rages on differences in finding the best area of human head that best represent core body temperature [1], [3], [6], [7], it is undeniably that the exposed area of head-to-shoulder is an important body part for target detection; a step prior to the febrile mass measurement. Fewer researches in the last decade have used non-visible spectrum sensors, such as thermal imaging. The breathtaking progress in computer vision research has mostly been associated with visible-light sensors. This is because initially, low cost infrared cameras had poor spatial resolution and cameras with better image quality were prohibitively expensive for many researchers. Due to different nature of images produced by visible-light and thermal cameras, this raise questions about the effectiveness of current state-of-art detectors used in visible dataset to be applicable in thermal dataset. In F.Suard et Al [10], HOG-SVM framework was used in detecting pedestrian applied to infrared images. Using HOG, they have evaluated using variety of parameters with accuracy of detection nearly 99%. Another research by Bertozzi et al [11] has been incorporating subsystem tetra-vision based for pedestrian detection. This system has use of both visible and far infrared cameras to detect people especially at night. Apart of their system, HOG-SVM detector based has been applied with very good results 91% of recognition rate. In [12], the nighttime pedestrian detection with in infrared thermal HOG is combined in their last step after pre-processing step to enhance the recognition and precisely locate the detected object.

These previous algorithm performances using infrared sensors signify positive impact in their context, but they mainly focus on specific applications such driving assistance safety, surveillance and military to such a degree make the camera setup and dataset characteristics different to our purposes. In addition, some of the thermal datasets used in previous research [13]–[15] are more fit to surveillance applications where the targets are far-scale. Nonetheless, their approach is based on the popular non-template framework, or using background subtraction, which requires two frames (the background and current frame). Thus it is less practical in real time application as the system need to frequently update the background image whenever the angle of camera or mountings adjusted. To sum up, more challenging thermal datasets are essential and to inspire new applicable ideas especially in adapting the challenging real time thermal subjects. Our contribution is threefold (1) In this research, we introduce, KLIA#1; a small-scale thermal dataset using high resolution infrared camera, taken from real stream people in Kuala Lumpur International Airport (KLIA) corresponds to application of fever mass screening in airports. The pedestrians vary in appearance, pose and scale; furthermore, occlusion information is annotated (see Fig.4). These statistics are more representative of real world applications and allow for in depth analysis of existing algorithms. (2) In this research, we are interested on template methods, where the examples of all possible targets to be detected are trained and independent of any previous/background frame. The three distinctive framework HOG, Haar and LBP are selected to detect the region of interest (ROI) in crowds using the thermal camera. We also take a different perspective by training these features using new local context and combining with Gentle Adaboost (GAB) for better performance. (3) We highlight best detector among the three distinctive detectors for situations of practical interest under which existing methods fail and identify future research directions.

## 2. Method

### 2.1. Dataset collection

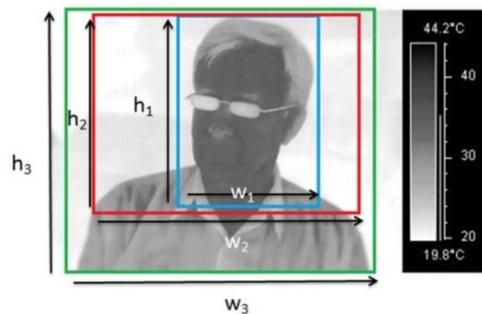
A few of thermal infrared dataset have been published in the past, e.g., the OTCBVS Benchmark 1, the LITIV Thermal-Visible Registration Dataset [16], the AIC Thermal Visible Night-time Dataset ,[17] and the ASL Thermal Infrared Dataset [18]. Typically these datasets focus on specific biometric applications or involve thermal-visible multimodal systems and imply a close-up view of the objects in the scene. The thermal camera (TIV) dataset collected by high-resolution high-speed cameras (FLIR SC8000, FLIR Systems, Inc., Wilsonville, OR) is the largest infrared dataset.[19] However, for the specific task such as the thermal health screening, there is still lack of complexity of visual events in realistic, challenging environments especially in the airport for human operators to manually monitor and detect febrile humans in crowd.

#### 2.1.1. Training Dataset

We split the database into training/testing data and specify our evaluation methodology. For training dataset, we train our detectors on Terravic Facial Infrared Database [13](see Fig.1). The database contains total no. of 20 classes (19 men and 1 woman) of 8-bit gray scale JPEG thermal faces. Size of the database is 298MB and images with different rotations are left, right and frontal face images also available with different items like glass and hat. In our training, random 102 images are selected. The reason we train on Terravic dataset, because, it contains a large variation of infrared poses of upper body part of human with good in resolution, 240x320 pixels, centered and non-overlap. Inspired by [21], we train a detector on instances in Terravic dataset that contain a person's entire head, neck and part of shoulder. As illustrated in Figure 1 our choice of training area is different from area trained in [20] [21], where the red box of size  $h_2 \times w_2$  shown the local context of our training instances. Viola at al trained on fixed face  $h_1 \times w_1$  and Kruppa et al train from head to to upper part of body. Intuitively, this choice of context promises to hold an important cue for the presence of our ROI, especially in highly occluded crowd in thermal images where temperature of human clothes normally exhibit similarly to the ambient temperature that can make the lower body parts merges into each other and lose the shape cues.



**Figure 1 .Terravic Facial Infrared Database for training database, random 102 images are selected for training.**



**Figure 2.The red box of size  $h_2 \times w_2$  show the local context of our training instances from head-to-shoulder. Blue box  $h_1 \times w_1$  is used as bounding box (BB) in Viola Jones [20] and green box  $h_3 \times w_3$  is used in Kruppa et. al.[21]**

### 2.1.2. KLIA Dataset and Ground Truthing

We collected approximately 156 seconds of 15 Hz thermal videos (~Total 2340 Frames and 1810 Frames with Humans walking) taken from a mounted thermal camera in a fixed indoor environment of KLIA arrival terminal to scan the arrival passengers (camera setup shown in Figure 3(a) and (d)). The scanner used is P640 FLIR system (FLIR Systems, 2009), a focal plane array, uncooled microbolometer 640 x 480 pixels with a thermal sensitivity of 30 mK at 30°C spectral range of 7.5–13  $\mu\text{m}$  and measurement accuracy at  $\pm 2\%$  of the real-time reading. The input of the system is the form of sequence image of temperature values, X0 Celsius degree as the intensity level is captured under controlled and fixed ambient temperature (25-260Celcius) in arrival area of Kuala Lumpur International Airport (KLIA). The camera is placed fixedly mounted and changed to three angles positions to capture the flow of moving subjects on the walkalator.

As for our testing dataset, apart from the video that we recorded, we extracted random of 160 images and named as KLIA#1. For each frame in KLIA#1, the annotator marks two types of bounding boxes (BB) for further analysis. The first is BB-Human that indicates the full extent of human in order to get distribution of height,  $h_f$  as in Fig. 4a and occlusion amount in Fig 4b. The second BB-R indicates the full extent of area from head-to-shoulder of human as shown in Fig 4c. The aspect ratio of BB-R is obtained as in Fig 4d. A histogram of the aspect ratios, using logarithmic bins, is shown in Fig. 3f, and indeed the distribution is lognormal. The log-average aspect ratio is 1.02, meaning that typically  $w_R \approx 1.02h_R$ , the width of BB-R is greater than height BB-R. However, to define near, medium and far scales, instead of  $h_R$ , we used the height of BB-Human,  $h_f$ , as this is more accurate and consistent to reflect the ratio of true height of human. Another reason, we annotate two type of BB; BB-Human and BB-R is to analyze occluded pattern of both parts. For both BB human and BB-R, the other BB is used to delineate the visible regions; BB-Hvis for BB-Human and BB-Rvis for BB-R, see Figure 4a

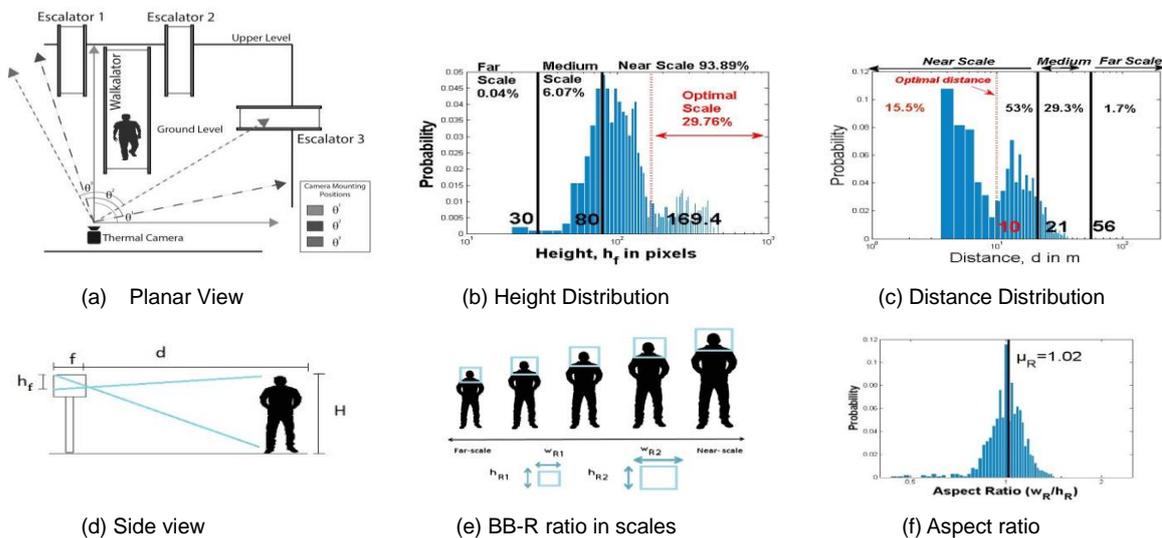
and 4c. This is useful to estimate the location of hidden parts. The summary of the dataset statistics is summarized in the next section.

### 2.1.3. Dataset Statistics

A summary of the database is given in Fig.3b and 3c. Below, we give detailed analysis of the distribution of pedestrian scale, occlusion and location. This will serve as a foundation for establishing the requirements for a real world system.

**Scale:** We group pedestrians by their image size (height in pixels) into four scales: near (80 or more pixels), medium (between 30-80 pixels) and far (30 pixels or less) and optimal scale (over 169 pixels). The value of optimal scale is calculated from pinhole camera model equation. Initially division into three scales (far, medium and near) is motivated by work in [22] as comparison. However our purpose is for fever screening requirements we add another scale called optimal scale. In a mass-screening survey, a long distance is required for a crowd of people. Research by [23] suggest that optimal distance of 10m or less is the best condition for the screening test and based on initial three scale division, this lie under near scale.

In Fig. 3b, we histogram the heights of the 956 BB-Humans in our KLIA#1 database using logarithmic sized bins. In Fig 3b we mark the height based on measurements of  $h_f$ , where a pedestrian's pixel height  $h_f$  is inversely proportional to distance to the camera. Note that only ~6.1% of the pedestrians lie in the medium scale, and nearly 94 % lie in near scale. The cutoffs for the near/far scales correspond to about  $\pm 1$  standard deviation from the mean height (in log space). The decrease at the other end, below 30 pixels, is due to the crowded people that are very small and difficult to be annotated. They are also overlapping and decrease in resolution quality. In Fig 3c distance distribution calculated from  $h_f$  values are plotted in logarithmic bins. Optimal distance for thermal screening is at 10m or less. The camera is mounted still on the tripod while passengers arrive from plane is entering the arrival section in airport is walking forward near to the thermal camera while being scanned. This is mainly the reason why most people lie in near scale.



**Figure 3. (a) Planar view of camera setup in KLIA airport and images are captured at three different angles. (b) Distribution of height,  $h_f$  in pixels of persons who walk through the camera. Based on [22] the division is defined into three scales; far (less than 30 pixel) medium (between 30 to 80 pixels) and near scale (over 80 pixels). In this work, another division named ‘optimal scale’ is added as this is relevant to the febrile detection (c) The distribution of distance,  $d$ , of people where  $d$  is inversely proportional to  $h_f$ . For febrile screening optimal distance for detection is at 10m. Based on division adopted from [22], our distance of interest lies in near scale category. (d) The side view based on pinhole camera to show the relationship of  $h_f$  and  $d$  based on  $h_f/f=H/d$ . (e) The BB-R is the bounding box of training from head-to-**

**shoulder. The aspect ratio of BB-R is relatively changed to the scales. (f) Distribution of BB-R aspect ratio with  $wR \approx 1.02hR$**

In contrast to [22], detection in the near scale is more reliable for febrile screening applications. Due to the large volume of tourists who arrive at the airport each day, a less time-consuming and reliable method to screen body temperature is needed. In this study, we chose a camera setup that mirrors expected febrile screening applications: The P640 visual camera includes matching Field Of View (FOV) lenses, so IR and visual images are shown at similar long distances using the same FOV. The focal length is 38mm with FOV of  $24^\circ \times 18^\circ$  and the detector is a focal plane array, uncooled microbolometer 640 x 480 pixels with a thermal sensitivity of 30 mK at 30°C spectral range of 7.5–13  $\mu\text{m}$  and measurement accuracy at  $\pm 2\%$  of the real-time reading. The mean temperature of a skin surface is measured from the field of view of a thermal imager with an appropriate adjustment for skin emissivity and it may vary from site to site in the range of 0.94–0.99 (0.98 is used here). The skin temperature is lower than the normal 37°C body temperature because of well-studied heat evaporation, conduction and convection principles.

$$AFOV(^{\circ}) = 2x \tan^{-1}(w/f) \tag{1}$$

$$h_f \approx H \times f/d \tag{2}$$

Using (1) focal length in pixels is  $f \approx 1129$ , given the angular field of view (AFOV) at  $24^\circ$  and  $w$  is horizontal dimension or width of image resolution is 480. Based on pinhole camera model relationship (2) and Fig. 3d, an observed pixel height of object,  $h_f$  is inversely proportional to the distance  $d$  to the camera: where  $H$  is the actual object height. Assuming 1.5m tall pedestrians, we can obtain an estimate of the distance to a pedestrian of observed pixel height  $h_f$ .

Although for febrile detection the best distance is at 10m and less, detecting coming people at medium scale may leave for far scale pedestrians are less relevant as temperature is cannot be measure accurately. Moreover, most people are observed at the near scale as human performance is quite good in the near scales but reduces noticeably at the far scale. Therefore, we shall use the near and medium distinction throughout this work.

**TABLE 1  
 Training Parameters for Three Detectors**

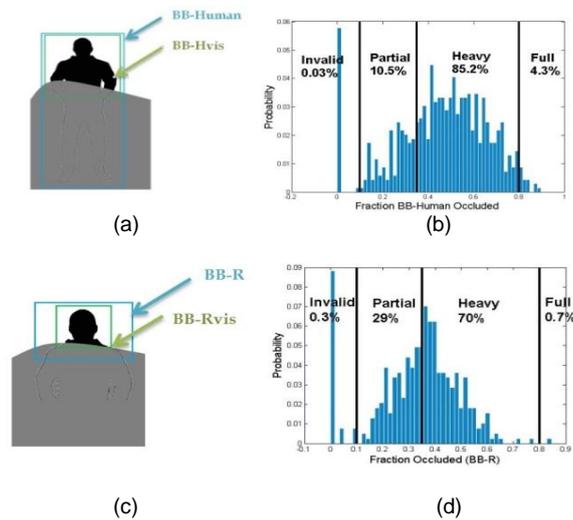
Features/ Boosting Parameters	Haar	HOG	LBP
Positive Samples Per Stage	105	105	105
Negative Samples Per Stage	210	210	210
Stages	6	7	10
Min hit rate	0.9950	0.9950	0.9950
Max False Alarm Rate	0.2	0.2	0.2
Width	37	37	37
Height	32	32	32

Comparison of training parameters using Gentle AdaBoost for three distinctive detectors. The training for the detectors halt and return cascade detector to certain stages due to limited samples.

Occlusion: Little previous work has been done to quantify detection performance in the presence of occlusion (using real data) especially in thermal datasets. Inspired by work for pedestrian detection in visible dataset in [22][24], we adopt their way to calculate fraction of occlusion. As described, occluded pedestrians are annotated with two types of BBs; BB-Human and BB-R that denote the head-to-toe and head-to-shoulder extent and each type of

BBs, we annotate their full and visible parts. In Fig 4a and 4c, the annotator labels visible parts for BB-Human and BB-R differently for the same person. For BB-Human, anything that hide the area between head-to-toe is consider as occlusion, likewise for BB-R, anything that hide area between head-to-shoulder is an occlusion.

In Fig 4 occlusion amount, i.e., for each pedestrian we measure the fraction of occlusion for each person and compare which parts are mostly occluded. To calculate the fraction of occluded, one minus the visible area over total area which is obtained from the visible and full parts of BBs. Aggregating, we obtain the histograms in Fig. 4b and Fig. 4d. Most of occlusions are generally caused by overlapping persons in crowd or any structure in the airports hide the visibility from the camera point of view. For BB-Human in Fig. 4(b), over 85% occlusion typically indicates heavy occlusion, while 0% is used to indicate that a BB could not represent the extent of the visible region. Compared to Fig. 4d the heavy occlusion in BB-R is less than BB-Human, which is 70%. This meant that head-to-shoulder parts are lessly occluded in crowds compared to the body and torso parts. We further subdivide the occlusion into partial occlusions, for BB-Human, only 10.5% is occluded and for BB-R, 29% of humans are partially occluded. This shows that probability of occlusion is non-uniform, detection using BB-R, more cues could be detected as this area is overlapped.



**Figure 4. Occlusion statistics (a) The annotate label head-to-toe BB, BB-Human for measuring heights. The exposed part within BB-Human is considered as BB-Hvis (b) The fraction of BB-Human that is occluded (c) The BB-R label from head-to-shoulder and the BB-Rvis, is what is visible within BB-R**

**TABLE 2  
Training Algorithm**

Gentle Adaboost
<ol style="list-style-type: none"> <li>1. Given <math>N</math> examples <math>(x_1, y_1) \dots (x_N, y_N)</math> with <math>x \in \mathbb{R}^k</math> <math>y_i \in \{-1, 1\}</math></li> <li>2. Start with weights <math>w_i = 1/N, i \dots N</math></li> <li>3. Repeat for <math>m = 1 \dots M</math> <ol style="list-style-type: none"> <li>a) Fit the regression function <math>f_m(x)</math>, by weighted least-squares <math>y_i</math> to <math>x_i</math> with weights <math>w_i</math></li> <li>b) Set <math>w_i \leftarrow w_i \exp(-y_i f_m(x_i)), i = 1 \dots N</math>, and renormalize weights so that <math>\sum_i w_i = 1</math></li> </ol> </li> <li>4. Output the classifier <math>sign[\sum_{m=1}^M f_m(x)]</math></li> </ol>

Gentle Adaboost Training Algorithm

## 2.2. Learning approach and implementation details

To measure performance we evaluated the effectiveness on three distinct detection frameworks. The employed detectors' framework is a modified version of the HOG, Viola-Jones and LBP cascade detectors and available through the Open Computer Vision Library and Matlab Computer Vision System Toolbox™ [20]. These are based on the idea of a boosted classifier cascades [20]. Features derived from HOG, Haar and LBP and offers different boosting variants for learning. In this paper, the main contribution is to evaluate three main features using an efficient classifier that is built by selecting number from huge library of potential features using Adaboost [25]. The multiscale sliding-window paradigm is employed. This subsection summarizes the most essential implementation details regarding features, learning algorithm and training parameters. The aim is to evaluate the most effective features on our KLIA#1 thermal dataset.

### 2.2.1. Viola Jones (VJ)

Viola and Jones built introduce integral images for fast feature computation and a cascade structure for efficient detection, and utilizing AdaBoost for automatic feature selection. The “Integral Image” allows very fast evaluation for Harr-wavelet type features, known as rectangular filter. These ideas continue to serve as a foundation for modern detectors. The VJ detection framework in [20] is conceptually straight-forward and later Haar features are extended in [26].

### 2.2.2. Cascaded Histogram Oriented Gradients (CHOG)

Based on adoption of gradient-based features, Dalal and Triggs popularized the HOG features based on evaluating a dense grid of well-normalized local histograms of image gradient orientations over the image windows. The distribution of local intensity gradient or edge directions can present the local object appearance and shape. In their work, the HOG features are combined with linear SVMs [27]. In visible object detection, HOG has shown great success in object detection and recognition. [27]–[30]. HOG has been widely accepted as one of the best features to capture the edge or local shape information. HOG has been successfully adopted for numerous object detection tasks and the HOG is almost the most frequently used descriptor in PASCAL challenges. Zhu et Al [29] has extended Dalal's framework and improve the speed of detector by combining HOG features in Adaboost cascaded classifier.

### 2.2.3. Cascaded Local Binary Patterns (CLBP)

While no single feature has been shown to outperform HOG, additional features can provide complementary information such as Local Binary Patterns (LBP). LBP was first presented by Ojala et al. [31], for the purpose of texture classification. LBP are feature vectors extracted from a gray-scale image by applying a local texture operator at all pixels and the using the result of operators to form histograms. Extending their earlier work, they present a gray-scale and rotation invariant texture operator based on LBP [32] and they derive an operator that is invariant against any monotonic transformation of gray scale.

### 2.2.4. Gentle Adaboost (GAB)

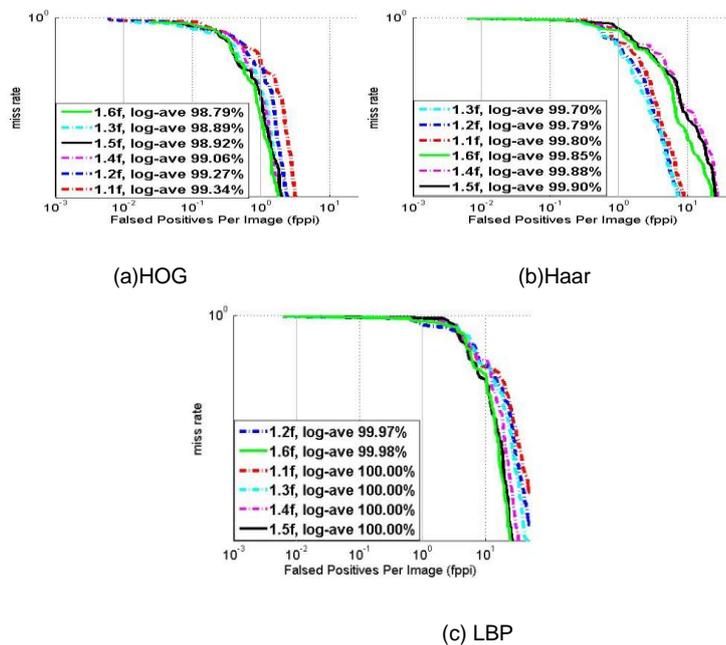
AdaBoost is an algorithm for constructing strong classifiers as linear combination of simple weak learners, which often outperforms SVMs and Neural Networks. The total number of features derived from Haar, HOG or LBP within any image sub-window is very large and Ada-boost is able to do feature selection resulting in relatively simple classifier. The cascade classifier consists of stages, where each stages an ensemble of weak learners. The weak learners are simple classifiers called decision stumps. Each stage is trained using a technique called boosting. Boosting provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners. As a result each stage of the boosting process, which selects a new weak classifier, can be viewed as an efficient feature selection process. Different variants of boosting are known such as Discrete Adaboost, Real AdaBoost, and Gentle AdaBoost are identical with respect to computational complexity,

but differ in their learning algorithm (see Table 2 for GAB algorithm). For face detection tasks, it is empirically shown in [22] that variant in Gentle Adaboost outperforms other type of Adaboost both in accuracy and speed. Thus in this paper, we adopt Gentle Adaboost (GAB).

### 3. Results

We first analyse performance under 3 conditions on testing data KLIA#1. Fig. shows the performance for overall, on near and medium scales, under no and partially visible pedestrians. We plot miss rate versus false positives per image (lower curves or lower percent of miss rate indicates better performance and use of log-average miss rate as a common reference value for summarizing performance. Legend entries display and ordered by the highest performance.

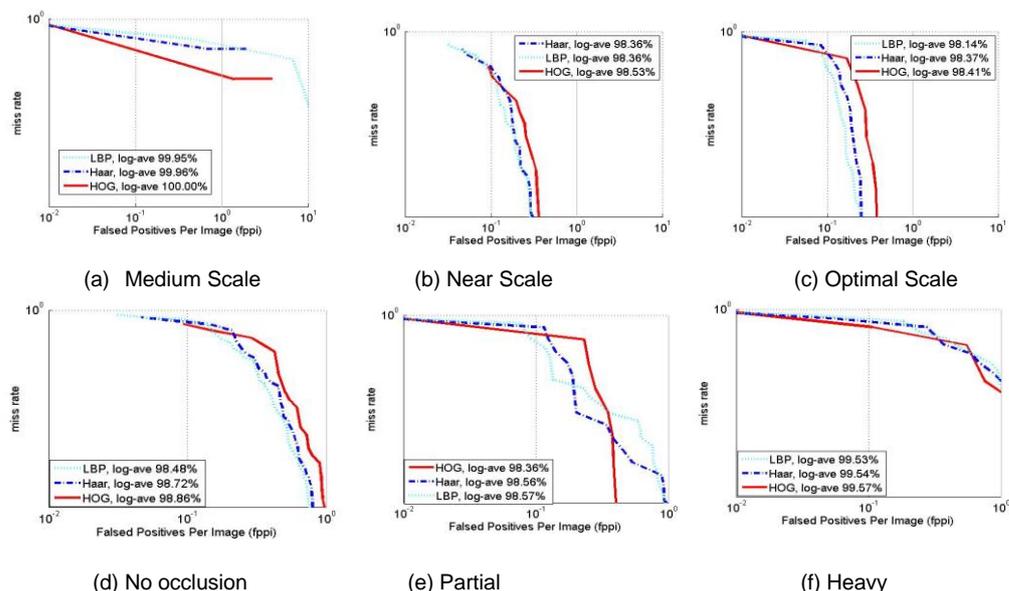
Overall: Fig. 5 plots performance on the entire test set with variable parameters, using a scalar minimum neighbour=6, the scale factors for multiscale scanning. Minimum neighbour is a parameter to specify how many neighbours in each candidate BB should have retain in the post-processing step. If minimum neighbour is 0, the function does not any grouping at all and returns all the detected candidate rectangles, which may be useful if the user wants to apply a customized grouping procedure. The scale factor is the factor by which the search window is scaled between the subsequent scans, if scale factor 1.1 means increasing window by 10 percent. HOG slightly outperforms the other detectors, with Haar a close second. In Fig. 5a HOG, the best scale factor is at 1.6. In Fig. 5b LBP and Fig. 5c Haar both best scale factor at 1.3. However, absolute overall performance is poor, with a log-average miss rate of over 98 percent. To understand where the detectors fail, we examine performance under various conditions in Fig with similar multiscale parameters, minimum neighbor at 6 and scale factor at 1.6.



**Figure 5. Overall performance on KLIA#1 using minimum neighbourhood 6 for (a) HOG (b) Haar (c) LBP. Overall detector is HOG performs better with lowest log-average miss rate**

Scale: Fig. 6 (a) to 6(c) shows the results for three scale (medium, near and optimal) corresponding to heights division of at least 80 pixels (see Section 2.3). LBP performs best with a log-average miss rate of only 98.14 percent; numerous other detectors still achieve slightly similar log-average miss rates around 98 percent. On the medium scale, which contains approximately 6 percent of ground truth humans in BB-Human annotations (see Fig.

3b) performance degrades slightly over 99 percent. LBP and Haar achieve the best relative performance, but absolute performance on HOG in medium scale is very poor with 100 percent log-average miss rate. Moreover, the three performing detectors on far scale pedestrians degrade most. By plotting the miss rate graphs, we can see this trend clearly.



**Figure 6. The performance of three detectors under six different conditions (a) far scale (b) medium scale (c) near scale (d) without occlusion (e) partially occluded and (f) heavy occlusion. The scale is measured through hf of BB-Human while occlusion fraction is based on BB-R and BB-Rvis (see Figure 4)**

**Occlusion:** The impact of occlusion on detecting humans is shown in Figs. 6d to 6e. As discussed in Section 2.3, we classify pedestrians as unoccluded, partially occluded (1-35 percent occluded), and heavily occluded (35-80 percent occluded). Performance drops significantly under heavy occlusion, leading to a log-average miss rate over 99 percent for three detectors. Surprisingly, performance of in partial detectors is substantially better than performance without occlusions.

**Scale-Occluded:** Because the performance in partial occlusion is slightly better than subjects without occlusion, the other further analysis is done under this type of occlusion to examine whether the scale size has influence the result in Fig.7. For near scale subjects (over 80 pixels height) that are partially occluded the detection is performing better than near scales that are not occluded. In Fig. 7c, for subjects that are partially occluded at near scale, the performance is higher with HOG produced the better performance at 97.98 percent. Surprisingly performance in Fig. 7d for optimal scale partially occluded increase with best results for overall detectors with HOG is the best. For near and optimal scales in Fig. 7e and Fig 7f, performance is slightly worst with the presence of heavy occlusion as the log-average miss rate is over 99 percent

This study was carried out to assess the state of the art in ROI detection by remote-sensing thermal camera that could be used as a proxy for core temperature recognition especially in moving subjects. Automatically detecting potential region in humans could have considerable effective impact to help personnels to detect the hot spot region correctly and fastly especially when passengers are arriving. Optimal scale is the best to detect temperature yet medium and near scale is essential to earlier scanning .In this result, we can observe that results in near is generally better than medium or far scales. In this research results of all far scale categories are the worst with log-average miss rate 100 percent. From most of the plots, the LBP and HOG have outperformed Haar features.

We make three main contributions: a new thermal data set taken in real setup, taking a different perspective by training these features using new local context and combining with Gentle Adaboost (GAB) for better performance and an assessment of performance based on scale, degree of occlusion for the three main detectors that is adapted to our application.

#### 4. Conclusion

The challenges in thermal camera are different compare to visible spectrum. One of them is the quality of images which is blurred and has poor contrast compared to visible images that are richer and colorful in presentation. As thermal images is a captured of temperature information and not color of the object, the shape is becoming an important feature. Thermal shape detection is often successfully exploited as normally humans usually emit more heat than other objects. The challenging part arise when the other hot artifacts such as lamps, heated buildings, and heat reflections can be misclassified with the real object of interest as they might share similar temperatures.

There is bigger room for improvement in ROI detection in thermal imaging. Overall performance is far from perfect especially as the miss rate is mostly over 96%. At medium scale, over 99 percent of log average miss rate and the performance degrades catastrophically when heavy occlusion occurs. This means that the current detectors need to be improved in order to cope with sudden mass passengers.

This research indicates the need of further research in our dataset:

1. Medium and near scales: Better performance is needed in the range of greater than 80 pixels and most importantly in optimal scale for distance of less than 10m (scale size greater than 169 pixel).
2. Occlusion: Performance degrades poorly under even heavy occlusion, but better for the partially occluded.
3. Global vs Local Context: Training could be done by combining local and global context especially when the overlapping degree is high. In visible image, when the two bodies are overlapping the temperature of the attires are as similar as ambient temperature thus erasing the borders of each other. Contrariwise in visible images the different of two overlapping parts could be bordered by the color of the cloth. Intuitively even in thermal spectrum images or in overlapping persons, normal human eyes could guess the global shape.
4. Novel features: From the performance results, HOG and LBP frequently and substitutely be the top of the ranks. Therefore, by manipulating the texture based or edge orientation based this can be a good feature in thermal images. Improving the features could achived additional gain.
5. Dataset: Detectors were trained on Terravic Thermal and the selected data we train in this research is smaller than the complete set of Terravic data. Training using larger Terravic data set or larger KLIA dataset is potentially to upgrade performance. Besides, the training data in Terravic perfect from any occlusion. Another study can be done to see the relationship between size of data and different type of thermal dataset that include overlapping humans in training dataset.

#### References

- [1] T. Bourlai, R. R. Pryor, J. Suyama, S. E. Reis, and D. Hostler, "Use of thermal imagery for estimation of core body temperature during precooling, exertion, and recovery in wildland firefighter protective clothing.," *Prehosp. Emerg. Care*, vol. 16, no. 3, pp. 390–9, 2012.
- [2] L. S. Chan, J. L. F. Lo, C. R. Kumana, and B. M. Y. Cheung, "Utility of infrared thermography for screening febrile subjects," vol. 19, no. 2, pp. 109–115, 2013.
- [3] L. Chan, G. T. Y. Cheung, I. J. Lauder, and C. R. Kumana, "Screening for Fever by Remote-sensing Infrared Thermographic Camera," pp. 273–279, 2004.
- [4] B. S. W. Eddie Y.K. Ng, Wiryani Muljo, "Study of Facial," *IEEE Engineering In Medicine and Biology Magazine*, no. June, pp. 68–74, 2006.

- [5] R. K. S. John, A. King, D. De Jong, M. Bodie-collins, S. G. Squires, and T. W. S. Tam, “Border Screening for SARS,” vol. 11, no. 1, 2005.
- [6] E. Y. K. Ng, G. J. L. Kaw, and W. M. Chang, “Analysis of IR thermal imager for mass blind fever screening,” *Microvasc. Res.*, vol. 68, no. 2, pp. 104–9, Sep. 2004.
- [7] A. V. Nguyen, N. J. Cohen, H. Lipman, C. M. Brown, N. A. Molinari, W. L. Jackson, H. Kirking, P. Szymanowski, T. W. Wilson, B. a Salhi, R. R. Roberts, D. W. Stryker, and D. B. Fishbein, “Comparison of 3 infrared thermal detection systems and self-report for mass fever screening,” *Emerg. Infect. Dis.*, vol. 16, no. 11, pp. 1710–7, Nov. 2010
- [8] J. Wang, Y. Zhang, J. Lu, and Y. Li, “Target Detection and Pedestrian Recognition in Infrared Images,” *J. Comput.*, vol. 8, no. 4, pp. 1050–1057, Apr. 2013.
- [9] Y. M. Wu, “Stop outbreak of SARS with infrared cameras,” vol. 5405, pp. 98–105, Apr. 2004
- [10] F. Suard, A. Rakotomamonjy, and A. Bensrhair, “Pedestrian Detection using Infrared images and Histograms of Oriented Gradients,” pp. 206–212, 2006.
- [11] M. Bertozzi, a. Broggi, M. Del Rose, M. Felisa, a. Rakotomamonjy, and F. Suard, “A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier,” 2007 IEEE *Intell. Transp. Syst. Conf.*, pp. 143–148, Sep. 2007.
- [12] S. Chang and S. Chen, “Nighttime Pedestrian Detection Using Thermal Imaging Based on HOG Feature,” 2011, no. June, pp. 694–698
- [13] J. W. Davis and V. Sharma, “Background-Subtraction in Thermal Imagery Using Contour Saliency,” *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 161–181, Jun. 2006.
- [14] J. W. Davis and V. Sharma, “Background-subtraction using contour-based fusion of thermal and visible imagery,” *Comput. Vis. Image Underst.*, vol. 106, no. 2–3, pp. 162–182, May 2007.
- [15] J. W. Davis and V. Sharma, “Robust detection of people in thermal imagery,” *Proc. 17th Int. Conf. Pattern Recognition*, 2004. *ICPR 2004.*, pp. 713–716 Vol.4, 2004.
- [16] A. Torabi, G. Massé, and G.-A. Bilodeau, “An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications,” *Comput. Vis. Image Underst.*, vol. 116, no. 2, pp. 210–221, Feb. 2012.
- [17] C. O’Conaire, N. O’Connor, E. Cooke, and a. Smeaton, “Comparison of Fusion Methods for Thermo-Visual Surveillance Tracking,” 2006 9th *Int. Conf. Inf. Fusion*, pp. 1–7, Jul. 2006.
- [18] J. Portmann, S. Lynen, M. Chli, and R. Siegwart, “People Detection and Tracking from Aerial Thermal Views,” pp. 1794–1800, 2014
- [19] Z. Wu, N. Fuller, D. Theriault, and M. Betke, “A Thermal Infrared Video Benchmark for Visual Analysis,” 2014 *IEEE Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 201–208, Jun. 2014.
- [20] P. Viola, O. M. Way, and M. J. Jones, “Robust Real-Time Face Detection,” vol. 57, no. 2, pp. 137–154, 2004.
- [21] H. Kruppa, E. T. H. Zurich, and C.- Zurich, “Fast and Robust Face Finding via Local Context 1 . Introduction and Related Work 2 . Local Context vs . Object-centered Detection.”
- [22] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: an evaluation of the state of the art,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–61, Apr. 2012.
- [23] M. F. Chiang, P. W. Lin, L. F. Lin, H. Y. Chiou, C. W. Chien, S. F. Chu, and W. T. Chiu, “Mass screening of suspected febrile patients with remote-sensing infrared thermography: Alarm temperature and optimal distance,” *J. Formos. Med. Assoc.*, vol. 107, no. 12, pp. 937–944, 2008.
- [24] P. Doll, C. Wojek, and P. Perona, “Pedestrian Detection : A Benchmark,” pp. 304–311, 2009.
- [25] R. E. S. Yoav Freund, “A Short Introduction to Boosting,” vol. 14, no. 5, pp. 771–780, 2009.
- [26] R. Lienhart, A. Kuranov, V. Pisarevsky, and M. R. L. T. Report, “Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection,” 2002.
- [27] N. Dalal, B. Triggs, and D. Europe, “Histograms of Oriented Gradients for Human Detection.”
- [28] N. Dalal, “Finding People in Images and Videos,” *Institut National Polytechnique De Grenoble*, 2006.
- [29] K.-T. C. Qiang Zhu, Shai Avidan, Mei-Chen Yeh, “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,” 2006 *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 2*, vol. 2, pp. 1491–1498, 2006.
- [30] P. Felzenszwalb, D. McAllester, and D. Ramanan, “A discriminatively trained, multiscale, deformable part model,” 2008 *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, Jun. 2008.
- [31] D. H. T. Ojala, M. Pietikainen, “Study of Texture Measures with Classification based on Featured Distributions,” *In Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [32] T. M. Timo Ojala, Matti Pietikainen, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.