Night-time pedestrian classification with histograms of oriented gradients-local binary patterns vectors

G.Preethi Monisha, V.Pushpalakshmi,.C.Vanitha

1. Assistant Professor , Department of EEE Hosur Institute of Technology
& Science , Krishnagiri
2. Assistant Professor, Department of ECE, T.S.M Jain College Of Technology,
Melur
3. PG Student, ME Communication System, T.S.M Jain College Of Technology
Melur
kavivino1994@gmail.com

Abstract: The use of night vision systems in vehicles is becoming increasingly common, not just in luxury cars but also in the more cost sensitive sectors. Numerous approaches using infrared sensors have been proposed in the literature to detect and classify pedestrians in low visibility situations. However, the performance of these systems is limited by the capability of the classifier. This paper presents a novel method of classifying pedestrians in far-infrared automotive imagery. Regions of interest are segmented from the infrared frame using seeded region growing. A novel method of filtering the region growing results based on the location and size of the bounding box within the frame is described. This results in a smaller number of regions of interest for classification, leading to a reduced false positive rate. Histograms of oriented gradient features and local binary pattern features are extracted from the regions of interest and concatenated to form a feature for classification. Pedestrians are tracked with a Kalman filter to increase detection rates and system robustness. Detection rates of 98%, and false positive rates of 1% have been achieved on a database of 2000 images and streams of video; this is a 3% improvement on previously reported detection rates.

1 Introduction

Safety is an extremely important factor in the design of automotive technology. In recent years, the area of automotive safety has expanded to incorporate the safety of road users outside of the vehicle. These are often referred to as vulnerable road users (VRUs) and are most commonly pedestrians and cyclists. Consumers are also becoming more safety conscious: in a study carried out by the European New Car Assessment Program (Euro NCAP) [1], 94% of respondents listed safety in vehicles as a major concern [2].

A disproportionate number of road fatalities occur during the hours of lowest visibility (at night). More than half (51%) of fatalities on European roads occur during the hours of darkness [3], despite the fact that the traffic densities are at their lowest at night. The use of advanced driver assistance systems (ADAS) capable of automatically determining the location of pedestrians in the vicinity of the vehicle contribute to reducing the number of pedestrian road fatalities when utilised in conjunction with other factors such as more stringent driver training and better road infrastructure. ADAS are being developed for use in lane detection [4], vehicle detection [5], vehicle navigation [6] and parking assistance systems [7] to create a safer enviroment for all road users.

Numerous techniques exist using visible spectrum cameras to detect VRUs during daylight hours, but these are not generally suitable for object detection at night, except for vehicle detection which can be done through detection of the head and tail lights [8]. However, detection of pedestrians at night is more difficult because of the lack of a visible natural light source. To compensate for this lack of ambient illumination, an infrared (IR) sensor can be employed to detect the IR heat signature generated by the human body. Passive-IR sensors do not require an illumination source and rely purely on heat signatures from the environment to produce a grey-scale image. Thermal radiation from humans peaks in the $8-15$ µm far-IR (FIR) spectral band, which makes FIR sensors suitable for pedestrian detection at night. FIR has also been shown to yield better driver response times than active near-IR sensors [9].

Image processing techniques used within the visible spectrum for pedestrian detection are generally not applicable in the FIR spectrum [10], because of the inherent differences between images generated by cameras in the IR and visible spectra. Hanqvist [9] states that pedestrians are generally warmer than the background environment and thus appear brighter in FIR imagery than other background scenery. High intensity thresholds can, therefore be used to

segment the bright regions from the darker background. Binary thresholding is commonly used in infrared pedestrian detection algorithms to isolate bright regions which could potentially contain a pedestrian [11–13]. Once isolated, these regions of interest (ROI) can be classified as containing a pedestrian or non-pedestrian object based on their features.

Accurate classification is vital to avoid non-detection of a pedestrian. Some of the more common pedestrian classification technologies used in automotive imagery include: support vector machines (SVMs) [14], artificial neural networks [15] and Adaboosting [16]. Histograms of oriented gradients (HOG) features have been successfully used in pedestrian detection systems for both day-time [17] and night-time [18] environments when used in conjunction with a SVM.

Local binary patterns (LBP) have also been used for feature extraction in training pedestrian detection systems [19], and have demonstrated high detection rates. LBP have also been used for texture classification [20], facial detection in the visible spectrum [21] and near-IR spectrum [22]. This paper describes the use of LBP features for FIR spectrum-based pedestrian detection.

The fusion of HOG features with LBP features to form a feature vector for SVM training has demonstrated improved detection rates in day time visible spectrum images when compared with a purely HOG trained classifier [23]. Wang *et al.* [23] implement a HOG–LBP-based classifier for day-time pedestrian detection that can also compensate for partial occlusion caused by objects concealing regions of a pedestrian.

Automotive IR video footage has a high degree of continuity, where paths taken by pedestrians can be reasonably deterministic over a short period of time; therefore tracking of detected pedestrians between frames can make detection more robust as the future position of a pedestrian can be predicted if detection fails temporarily because of the factors such as partial occlusion. A Kalman filter is used in the proposed system to track pedestrians between IR frames. Kalman filters have been widely used to implement tracking in pedestrian detection algorithms in both daytime [24] and night-time environments [12]. When tracking is used in conjunction with a SVM classifier, pedestrian detection rates have been shown to increase in FIR video [12].

This paper presents a novel method of classifying pedestrians in FIR automotive imagery using a combination of HOG and LBP feature vectors. This technique has not been previously reported for detection of pedestrians in FIR images. Results show that HOG–LBP features improve upon previous works utilising either HOG or LBP feature vectors (but not both), and a comparison of the HOG, LBP and combined HOG–LBP features for pedestrian detection in FIR images is presented in this paper. The system was tested with footage captured using an FIR camera mounted on a car in urban scenes with a range of pedestrian and non-pedestrian objects present in each scene. Classification of pedestrians has been performed on a database of 2000 FIR images, and streams of FIR video (~15 000 frames in total). Target tracking was implemented for HOG–LBP detection in streams of FIR video which resulted in a higher detection rate and a lower rate of false positives.

The structure of this paper is as follows: Section 2 discusses how ROIs are identified in frames of FIR video. Section 3 describes the extraction of HOG and LBP features from segmented ROIs and the process of training a SVM classifier. The tracking of pedestrians between frames is presented in Section 4. The results are described in Section 5, and Section 6 presents the conclusions, benchmarks the performance of the proposed system with the known literature and discusses the direction of potential future work in this area.

2 ROI isolation

To ensure accurate classification, ROIs that could potentially contain a pedestrian are segmented from the image using a feature-based, seeded region growing method. This has an advantage over sliding window approaches which are commonly used in pedestrian detection algorithms in the automotive domain. The main disadvantage with sliding windows is that a pedestrian cannot be detected if the image of the pedestrian does not fit within the window dimensions. It is also difficult to detect pedestrians that only occupy a small proportion of the window, so trade-offs must be established. Previous research has described the use of multi-scale windows to resolve this problem, however, this method can be computationally expensive as a large number of window sizes must be used in order for the system to be able to extract all possible variations of pedestrian sizes that will be encountered due to differing heights and distances from the IR sensor. A single classifier can also be implemented with a region growing approach rather than cascades of classifiers which are typically used for multi-resolution detection to aid sliding window-based approaches [25].

An IR pedestrian detection system relies on the fact that the temperature of a pedestrian is generally higher than the surrounding environment. In real-world environments, temperatures fluctuate throughout the year and pedestrians dress in appropriate clothing for each season which can distort the shape of a pedestrian (particularly when heavily insulating clothing is worn in cold weather). O'Malley *et al.* [26] refer to this as 'clothing distortion' and remove it from the image using a morphological closing operation. Examples of this distortion can be shown in Fig. 1*a* where two pedestrians are situated in close proximity with clothing distortion present in their torso regions. A circular kernel is typically used in closing operations, however, this would cause groups of bright regions in close proximity to one another to merge as shown in Fig. 1*b*, this would subsequently cause the classifier to falsely classify the ROI as a non-pedestrian object. To prevent merging of pedestrians in close proximity, a vertically orientated kernel is used to remove the clothing distortion, as shown in Fig. 1*c*. A vertically oriented kernel preserves the vertical shape of a pedestrian which aids segmentation of ROIs from the background imagery.

After removing the clothing distortion from the image, a region growing method is used to grow 'seed' regions within the image to accurately segment ROIs from the background [26]. A high intensity threshold is applied to the image in Fig. 2*a* that separates high intensity seeds from the rest of the infrared image, as shown in Fig. 2*b*. These seeds are then grown by applying thresholds of lowering intensities until two stopping conditions are met. The first stopping condition is the aspect ratio of the bounding box which encompasses the ROI

$$\text{Aspect ratio} = \frac{w}{h} \tag{1}$$
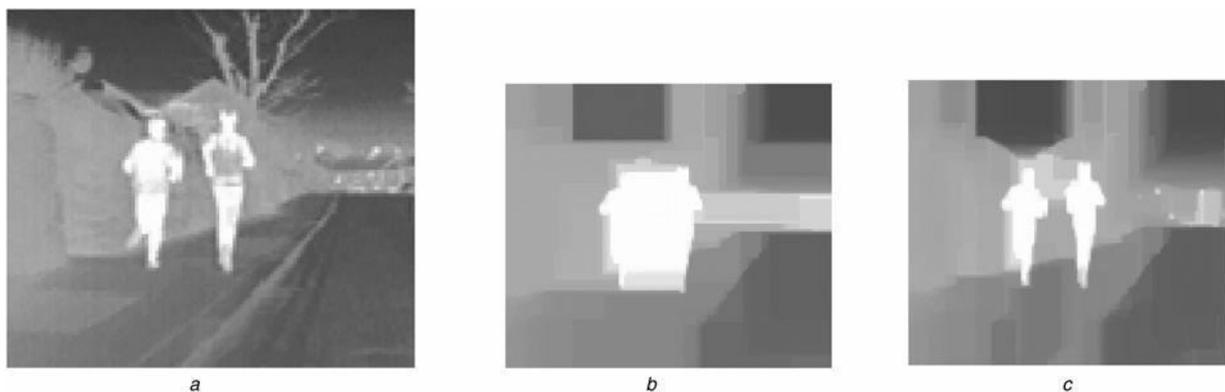
Fig. 1 *Images from clothing distortion compensation stage*

a Original FIR Image
b Image closed with a circular kernel, pedestrians have merged
c FIR image is closed with a vertical kernel, no merging of pedestrians
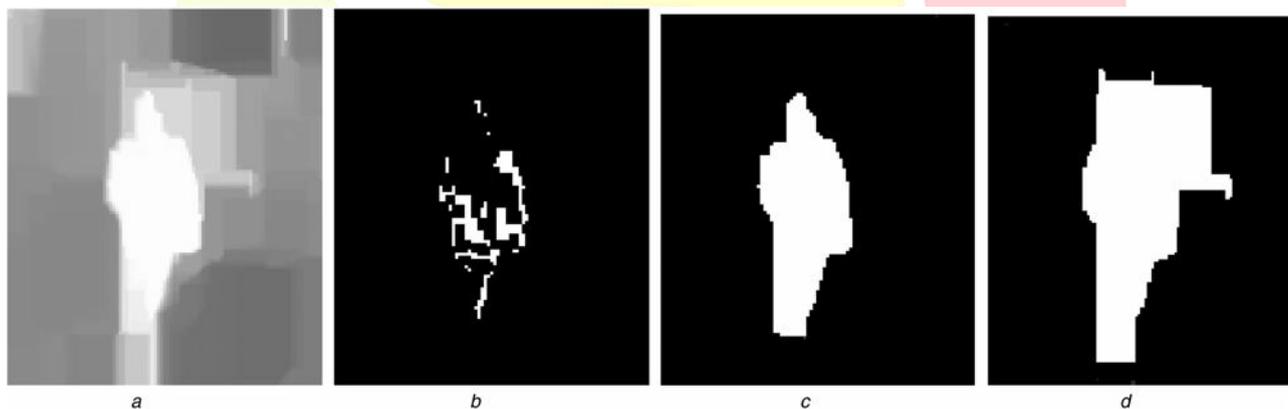
Fig. 2 *Steps involved in seeded region growing*

a Image segment closed with vertical rectangle kernel
b High intesity threshold segments seeds (threshold = 0.98)
c Seed is grown to the limits of the parameters set out in Table 1 (threshold = 0.823)
d Merged with background and filtered by extent and aspect ratio boundaries (threshold = 0.706)

where w is the width of the bounding box and h is the height of the bounding box. This is a measure of the proportion of the pedestrian's height to its width. Pedestrians in an image from a camera mounted on a vehicle will generally appear taller than they are wide (this assumes that the pedestrians are not seated, crouched or lying down). The second stopping condition is the extent of the region

$$\text{Extent} = \frac{a}{wh} \qquad (2)$$

bounding boxes within the FIR frame. A pedestrian situated far from the sensor will have a bounding box with an upper vertex located close to the horizon and the height and width of the ROI will be quite small. Pedestrians located close to the sensor will have an upper vertex situated further from the horizon and will have a taller, wider bounding box. Based on these observations, threshold values were empirically derived and used to define an area of interest (AOI) within the image. ROIs which have an upper vertex located outside of this AOI are discarded. ROIs with an upper vertex located inside of the AOI are analysed and

where a is the area of the bounding box. A pedestrian's shape is quite compact and will generally fill a large proportion of its bounding box. The upper and lower limits for the extent and aspect ratio parameters are shown in Table 1. When a pedestrian region merges with the background scenery it typically breaks the extent or aspect ratio limits (Fig. 2*d*). When this occurs, the co-ordinates of the ROIs bounding box (determined in the previous threshold, Fig. 2*c*) are recorded for classification. A sample of outputs from the different stages of the region growing process are shown in Fig. 2.

The number of ROIs for classification can be further reduced based on the position and relative size of their

based on the location of the upper vertex relative to the height of the ROI it is either discarded or stored for classification. The values used to filter the ROI are scaled, based on the location of the upper vertex within the AOI.

Table 1 Limits of extent and aspect-ratio for ROI

segmentation	Parameter	Value
	extent	0.50–0.95
	aspect ratio	0.20–0.49

If the bounding box of an ROI falls outside these limits it is not counted as a valid ROI for classification
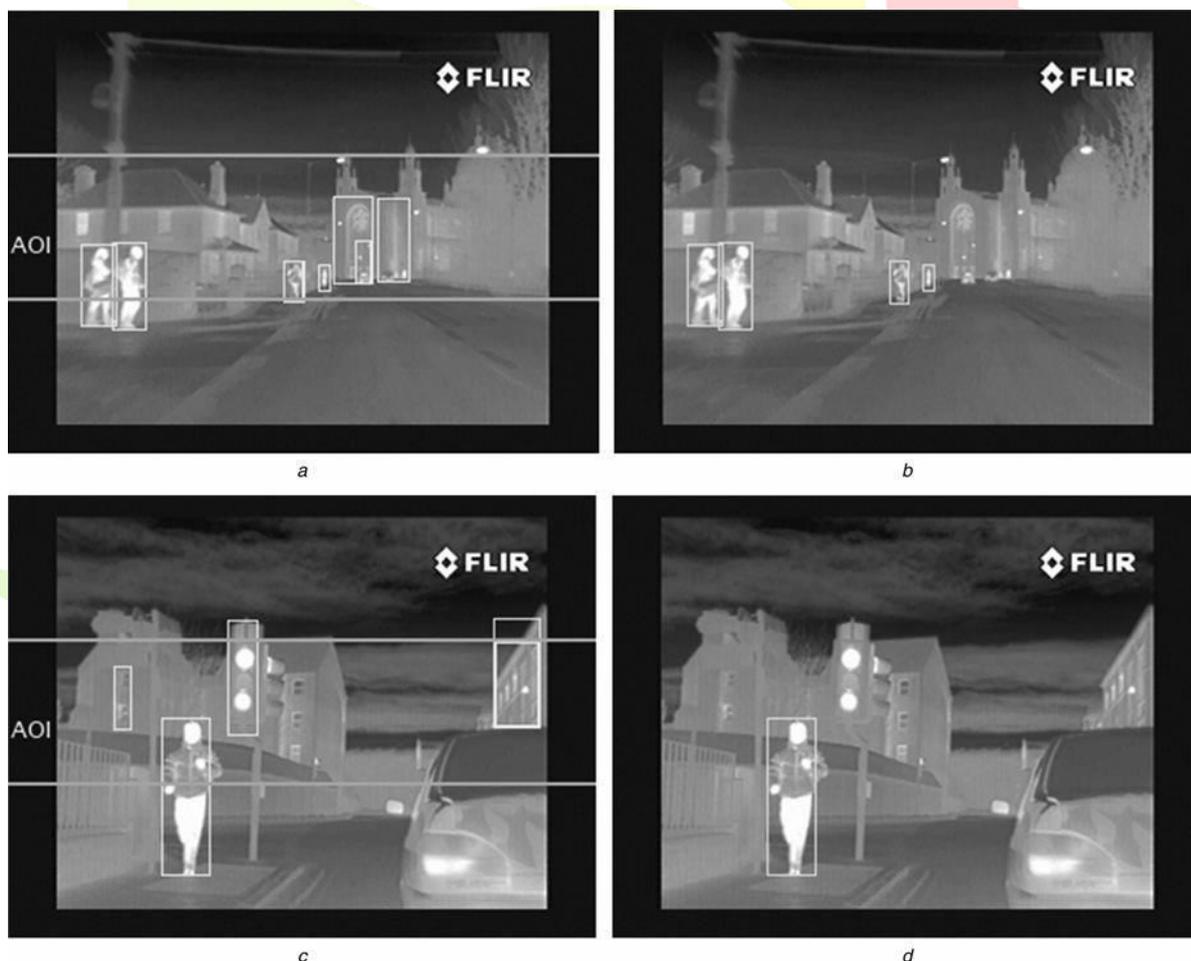
Fig. 3 *Sample images from the outputs of seeded region growing*

a and *b* ROIs that satisfy both extent and aspect ratio filters are encompassed by a bounding box, the AOI is located between the horizontal lines within the image
c and *d* ROIs that satisfy the distance estimation filter are not removed. The large candidates in the background containing non-pedestrians are removed prior to classification

This method removes a large number of common false positives (street lights, wheel arches etc.) while preserving the true positives within the image for classification.

A number of samples of the output from the ROI isolation stages are shown in Fig. 3. Figs. 3*a* and *b* display the output of the seeded region growing stage, a ROI encompassed by a yellow bounding box has satisfied the aspect ratio and extent metrics, the thresholds defining the boundaries of the AOI are overlaid. Figs. 3*c* and *d* display the resulting filtered regions according to their location and size. Note the removal of background object ROIs not containing pedestrians. This method of ROI isolation is more adaptable to different pedestrian poses and variations in ambient temperature than a window-based method. This leads to fewer classification operations by the SVM resulting in fewer opportunities for the SVM to generate false positives, and a lower computational overhead for the system.

3   ROI classification

The second stage of many existing pedestrian detection algorithms is classification of the segmented ROIs. This section discusses the features extracted from the ROI and used by the classifier to determine if the isolated ROI contains a pedestrian.*Training data*

For training and validation purposes, a database of grey-scale images was generated by extracting ROIs from on-road FIR video data, with the region growing algorithm. A total of 2000 ROIs were extracted and manually labelled (1000 pedestrian and 1000 non-pedestrian). A variety of pedestrian poses with differing levels of clothing distortion were used to provide a representative sample of real-world pedestrian targets. Fig. 4 illustrates some examples of pedestrian and non-pedestrian database entries.

Both HOG and LBP feature vectors were calculated for each image in the database. These features were then concatenated and used to train the SVM classifier.

3.1   Histogram of oriented gradients

To generate a HOG feature vector for a ROI, the ROI must first be scaled to a fixed size prior to generating it is HOG feature vector. Each ROI is resized to a fixed size of 20×40 pixels [27] prior to HOG feature vector generation.

The gradient image of the ROI is first calculated by performing a convolution with a horizontal gradient kernel [−1, 0, 1] and with a vertical gradient [−1, 0 1]^T and combining the results. An example of a gradient image is shown in Fig. 5*b*.
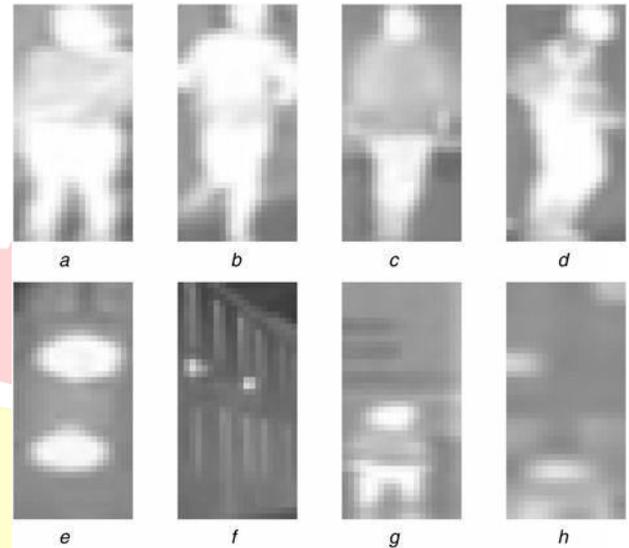
Fig. 4   *Examples of data used for training the SVM classifier*
All images are resized to 20×40 pixels for training purposes
a–d Pedestrians
e–h Non-pedestrians

The resultant gradient image is divided into cells of 5×5 pixels, as shown in Fig. 5*c*. The HOG feature is then computed for each cell by accumulating votes into bins for each orientation. Each bin effectively represents the 'strength' of an edge through 9 orientations from 0° to 180° for a total of 9 histogram bins.

When all of the histograms have been computed, a single feature vector descriptor is formed by concatenating all of the histograms. The resulting feature vector is normalised with L^2-norm [26]. The L^2-norm of an un-normalised vector containing histograms of a block (u) is

Table 2   Parameters used in hog feature vector generation

Parameter	Value
image height	40 pixels
image width	20 pixels
cell width	5 pixels
cell height	5 pixels
no. of orientations	9 (0°−180°)
overlap	0.5
normalisation method	L^2-Hys

division by 0. The value for v is clipped at 0.2 [26, 28]. This is referred to by Dalal and Triggs as L^2-Hys normalisation [17]. This results in a descriptor vector which describes a grey-scale ROI of 20×40 pixels. A HOG feature is generated for each 20×40 FIR image (an example of a HOG descriptor is shown in Fig. 5*d*). A summary of the parameters used for HOG feature generation are shown in Table 2.

3.2   Local binary patterns

LBP were originally proposed as a method of texture classification by Ojala *et al.* [20]. The LBP technique has been applied to areas such as facial detection with visible

spectrum cameras [21] and NIR cameras [22]. LBP features are determined for a ROI by traversing the ROI one pixel at a time with a structuring element, for example, Ojala *et al.* [20] used a structuring element of 3×3 pixels. In essence, the LBP value for a structuring element is calculated by thresholding the outer pixels

$$v = \frac{}{_2 + e^2 \atop \|u\|^2} \tag{3}$$

where *e* is a small constant whose purpose is to prevent

against the centre pixel of the structuring element (Fig. 6*b*) and the resultant 8-bit binary number is then converted to an integer (Fig. 6*c*). Once the image has been traversed and integers have been computed for all pixels, a histogram is generated. The structuring element can be extended to use

differing radii, Ojala *et al.* [20] found that a larger number of sampling points and radii provided higher detection rates in classifying textures in images with a resolution of 256×256 pixels.
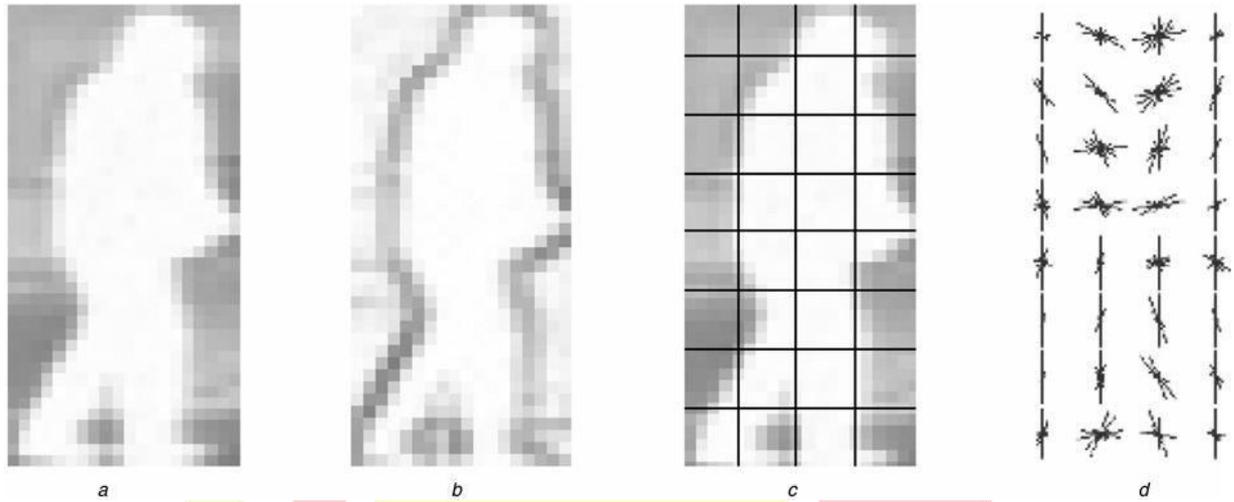
Fig. 5 *Images from the stages of generating a HOG feature vector*
a Pedestrian ROI, scaled to 20×40 pixels
b Gradient image
c ROI divided into cells of 5×5 pixels, resulting in 4 cells $\times$ 8 cells
d HOG descriptor for the ROI showing the gradient orientation histograms in each cell
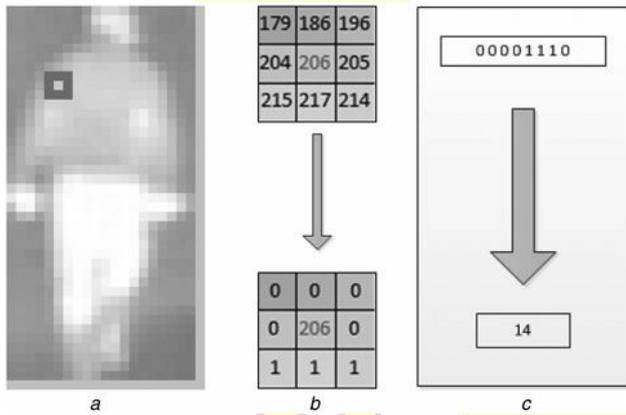
Fig. 6 *Images from the stages of generating a LBP feature vector*
a Pedestrian ROI, scaled to 20×40 pixels with 3×3 structuring element highlighted in red
b Outer values of LBP 3×3 structure thresholded against centre value (206)
c 8-bit binary number (14) generated from structuring element converted to decimal for storage in histogram bin

This paper uses the following notation for LBP structures is $\text{LBP}_{P,R}$, where P is the number of sampling points used and R is the radius of the structuring element. When a sampling point does not fall on integer co-ordinates, the pixel value at that point is bilinearly interpolated. The LBP label for the centre pixel (x, y) of the structuring element is denoted by

$$\text{LBP}_{P,R}(x, y) = \sum_{i=0}^{N=1} s(n_i - n_c)2^i, \quad s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (4)$$

To reduce the total number of bins generated by the LBP stage, uniform patterns are employed. In [20], the authors state that uniform patterns make up approximately 90% of textures when a 3×3 structuring element is used for LBP feature generation. A uniform pattern is achieved when the number of spatial transitions in the output is less than or equal to 2. Examples of uniform patterns are '00110000' and '11110111'. The total number of uniform patterns that can be achieved with a $\text{LBP}_{8,1}$ structuring element is 58.

SVM operates by calculating the optimal separating hyperplane between classes in higher dimensional space. In this paper, a SVM is used to classify ROIs identified during the previous stage as either pedestrian or non-pedestrian based on their HOG–LBP features. A radial basis function (RBF) has been used as a kernel (K) for SVM classification in this paper

$$K(x, y) = e^{-y||x-y||^2} \quad (5)$$

4 Pedestrian tracking

Tracking of pedestrians between frames adds robustness to the system. Information acquired from previous frames can be used to search for previously detected pedestrians in the current frame, and for interpolation between frames. Tracking can also be used to predict the location of a pedestrian in a frame if classification temporarily fails because of partial or full occlusion of the pedestrian.

A Kalman filter [31] has been used in this work to track pedestrians in sequential frames in this system. Kalman filters have been used previously in automotive applications such as pedestrian detection [12] and vehicle detection [32]. The equations that govern the Kalman filter are presented in [31]. The system specific entities for Kalman tracking are presented in this section. Targets are tracked in the system by using four parameters associated with the pedestrian bounding box, namely x-position, y-position, width and height, and these parameters are used to form a measurement vector (z). Predictions of the state vector ($\hat{x}$) and state error covariance matrix (P^-) are generated for a ROI at time k. The speed and direction of a pedestrian does When calculating LBP patterns for an image, 58 histogram bins are used for the various uniform patterns and 1 extra

not change significantly between frames, assuming that the camera is travelling straight with constant speed [33], the state transition matrix A is an identity matrix. The role of A

is to relate the state at the previous time step $(k{-}1)$ to the current time step (k), in the absence of either a driving

bin is used for storing the total number of non-uniform patterns within the image. The number of uniform patterns

generated for larger structuring elements is scaled according to the number of sampling points used.

3.3 HOG–LBP features

When the HOG and LBP features have been calculated for the ROI, they are concatenated to form a single HOG–LBP feature vector. The fusion of HOG features with LBP features as training data for a SVM has been utilised for pedestrian detection in the visible spectrum where increased detection rates have been achieved compared with a SVM

trained purely using HOG vectors [23].

3.4 SVM classifier

A SVM classifier is used in conjunction with the combined HOG–LBP feature vectors for classification of isolated ROIs. SVMs have been used in a range of fields including facial recognition [29] and gesture recognition [30]. An

function or process noise. For the proposed system, A is

$$
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} \tag{6}
$$

The resulting updated state vector $(\hat{x})$ is

$$
\hat{x}_k^- = \begin{pmatrix}
x_{k-1} + \mathrm{D}x_{k-1} \\
y_{k-1} + \mathrm{D}y_{k-1} \\
w_{k-1} + \mathrm{D}w_{k-1} \\
h_{k-1} + \mathrm{D}h_{k-1} \\
\mathrm{D}x_{k-1} \\
\mathrm{D}y_{k-1} \\
\mathrm{D}w_{k-1} \\
\mathrm{D}h_{k-1}
\end{pmatrix} \tag{7}
$$

proposed system is

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (8)$$

The measurement noise covariance matrix R of the Kalman filter determines the sensitivity of the tracker to updates. A high value of R results in smoother movement and less weighting on detections in the current frame, while a small value will result in a more responsive tracker and heavier weighting on the current measurements. However, a value too small can cause the tracker to become unstable. For the video data used in this paper (320×240 pixels, 25 Hz frame rate) an R matrix of $0.1I$ has been found to be suitable for ensuring the tracker is responsive, and also remained stable in the presence of noise caused by variations in the road surface. For the proposed system R is

$$R = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix} \quad (9)$$

The co-ordinates of each classified pedestrian are stored between frames, the previous frame's co-ordinates are then compared with classified ROI in the current frame. If there is a low correlation between the current set of co-ordinates and the co-ordinates found within the previous frame, a new tracker object is defined. If there is a high degree of correlation between the current set of pedestrian position co-ordinates being analysed and co-ordinates found in the previous frame, this is noted and a counter is incremented to keep track of the number of times a pedestrian has been classified in a series of consecutive frames. If a ROI is classified in more than five consecutive frames, the system then creates a tracker object for the ROI and draws a bounding box around the ROI on the in-vehicle display. Tracking has been found to remove a large number of false positives from the system, as false positives tend to occur for only a short duration, (generally 1-2 consecutive frames). If a target goes undetected for more than ten frames then it is deemed to have moved beyond the camera field of view and its associated tracker object is discarded, this allows for regions to go undetected for a short period of time without discarding them. This method of tracking pedestrians allows for more robust system performance and has been shown to generate fewer false positives than a system that does not utilise tracking.

5 Results and discussion

The proposed system has been implemented in Matlab and executed on an Intel Core i7 860 processor with a clock frequency of 2.80 GHz at a rate of 3 frames per second. The performance of the system will be greatly improved through implementation on an embedded system. FIR video has been captured with an automotive grade FIR microbolometer sensor at a rate of 25 fps and a resolution of 320×240 pixels (QVGA). Video segments have been captured in urban, suburban and rural environments at speeds ranging from 0 to 100 km/h. The various classifiers were tested on the database of 2000 FIR images and

approximately 15 000 frames of captured FIR video, the streams of FIR video are independent of the training database.

The performance of the SVM classifier in this paper is presented in two subsections: subsection (A) focuses on detection of pedestrians in FIR images which have been isolated from captured streams of FIR video and subsection (B) focuses on the performance of the classifier on streams of FIR footage. All of the data within the FIR streams is independent of the data used to train the SVM classifier.

5.1 Performance of classifier on FIR database

A range of approaches have been used in the literature to quantify performance in IR pedestrian detection algorithms. A performance evaluation of pedestrian detection systems was presented in [34] which notes that receiver operating characteristics (ROC) curves are an effective tool for quantifying performance. Results are presented in the form of a detection rate with/without tracking and the number of false positives present. The detection rate is the proportion of frames in which a pedestrian is successfully detected (d), out of the total number of frames in which they are present (n)

$$\text{Detection rate} = \frac{d}{n} \quad (10)$$

The detection with tracking rate is the proportion of frames where a pedestrian is successfully detected or tracked (t), out of the total number of frames in which a pedestrian appears (n)

$$\text{Detection with tracking rate} = \frac{t}{n} \quad (11)$$

The recording and measuring of the number of false positives is an important factor in a pedestrian classification system. The false positive rate is a measure of how many regions the classifier falsely determines is a pedestrian (f) out of the total number of frames (n)

$$\text{False positive rate} = \frac{f}{n} \quad (12)$$

K-fold cross validation has been used to test the performance of the SVM when trained with each of the various feature vector sets. Cross-validation has been performed with $K = 10$, which divided the database of images used for training

Table 3 Confusion matrix displaying detection rates for multiple LBP structuring elements

Structuring element		Predicted	Class
$LBP_{8,1}$	pedestrian	0.91 (TP)	0.17 (FP)
	non-ped	0.09 (FN)	0.83 (TN)
$LBP_{16,2}$	pedestrian	0.90 (TP)	0.09 (FP)
	non-ped	0.10 (FN)	0.91 (TN)
$LBP_{24,3}$	pedestrian	0.72 (TP)	0.12 (FP)
	non-ped	0.28 (FN)	0.88 (TN)
$LBP_{8,1 + 16,2}$	pedestrian	0.90 (TP)	0.11 (FP)
	non-ped	0.10 (FN)	0.89 (TN)
$LBP_{8,1 + 24,3}$	pedestrian	0.72 (TP)	0.09 (FP)
	non-ped	0.28 (FN)	0.91 (TN)
$LBP_{16,2 + 24,3}$	pedestrian	0.73 (TP)	0.09 (FP)
	non-ped	0.27 (FN)	0.91 (TN)
$LBP_{8,1 + 16,2 + 24,3}$	pedestrian	0.76 (TP)	0.08 (FP)
	non-ped	0.24 (FN)	0.92 (TN)

Table 4 Confusion matrix for SVM pedestrian classifier trained with various feature vectors on database of FIR imagery

		Predicted Pedestrian	Class Non-Ped
HOG	pedestrian	0.94 (TP)	0.02 (FP)
	non-ped	0.06 (FN)	0.98 (TN)
$LBP_{8,1}$	pedestrian	0.90 (TP)	0.16 (FP)
	non-ped	0.10 (FN)	0.84 (TN)
$HOG–LBP_{8,1}$	pedestrian	0.98 (TP)	0.01 (FP)
	non-ped	0.02 (FN)	0.99 (TN)
$HOG–LBP_{16,2}$	pedestrian	0.87 (TP)	0.01 (FP)
	non-ped	0.14 (FN)	0.99 (TN)

True positive (TP), false positive (FP), true negative (TN) and false negative (FN) detection rates are all displayed

into ten groups. Nine groups have been used to train the SVM and one group for testing, this process is repeated with different groups until every group has been classified. The resulting confusion matrix displaying the metrics of each feature vector are presented in Table 4. Cross-validation increases the time it takes to train the classifer but it ensures that each image is used for both training and testing. This gives a more detailed representation of classifier performance on unseen data. The ROC curve for each type of feature vector is presented in Fig. 7a, the upper left-hand corner of the curve is enlarged for clarity in Fig. 7b. The ROC curves presented are generated based on the performance of the classifier on the database of 2000 images.

The combination of multiple structuring elements has been shown to provide better detection rates for texture classification [20]; therefore a range of structuring elements have been used. The results of tests with detection with multiple structuring elements is presented in Table 3. These results show that combinations of different scales of LBP structuring elements yield no improvement in detection rates in this application. The structuring elements which yield the best results were $LBP_{8,1}$ and $LBP_{16,2}$. These are fused with HOG features to create $HOG–LBP_{8,1}$ and $HOG–LBP_{16,2}$ features to determine which arrangment yields the best detection rates in SVM training. The larger structuring element $LBP_{24,3}$ does not perform as well as the smaller structuring elements since image dimensions of the

structuring element exceed that of the scaled image, that is, 20×40 pixels.

The confusion matrix for the SVM classifier is presented in Table 4. The operating point for the confusion matrix has been chosen at a detection rate of 0.98 and a false positive rate of 0.01 for the $HOG–LBP_{P,R}$ + SVM classifier. This point gave the best trade-off between detection rates and false positive rates. The SVM parameters that yielded this performance were a cost (C) of 2048 and RBF kernel parameter of (γ) 10.

The detection rates presented in this work illustrates that a SVM using a fusion of HOG and LBP feature vectors provides more accurate detection results in FIR imagery than individual HOG or LBP with SVM. The true positive detection rate of 98% is a significant improvement over the HOG only based classifier which had a detection rate of 94% and the LBP classifier which had a detection rate of 90%. The false positive rate of the HOG–LBP classifier is 1%; lower than the HOG only classifier rate of 2% and the LBP only classifier rate of 16%. These figures show that the combination of HOG and LBP feature vectors result in more accurate pedestrian detection in automotive FIR imagery.

5.2 Performance of classifier on FIR video

The results from FIR video are presented in Table 5. The results on real-world FIR footage shows that HOG–LBP feature vectors yield significantly higher detection rates and lower false positive rates than just HOG or LBP trained classifiers alone. A large number of pedestrians are detected at close range and are classified correctly with an SVM trained with HOG–LBP feature vectors. Examples of pedestrians classified with all three feature sets are

Table 5 Results of pedestrian classification on streams of fir footage

Feature	Length	Detection rate	False positive rate
HOG	15 000	94.6	0.017
$LBP_{8,1}$	15 000	91.6	0.14
$HOG–LBP_{8,1}$	15 000	98.41	0.008
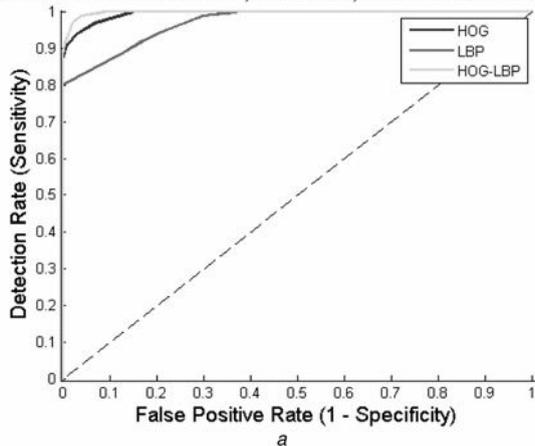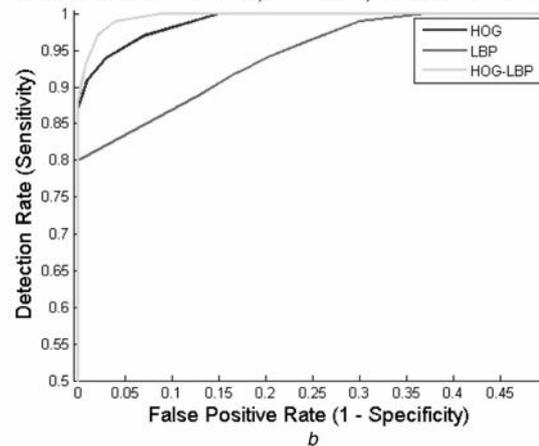
Fig. 7 *ROC curves for various feature vectors*

a ROC curve containing results of cross-validation tests for HOG, LBP and HOG–LBP

b Enlarged view of upper left portion of ROC curve
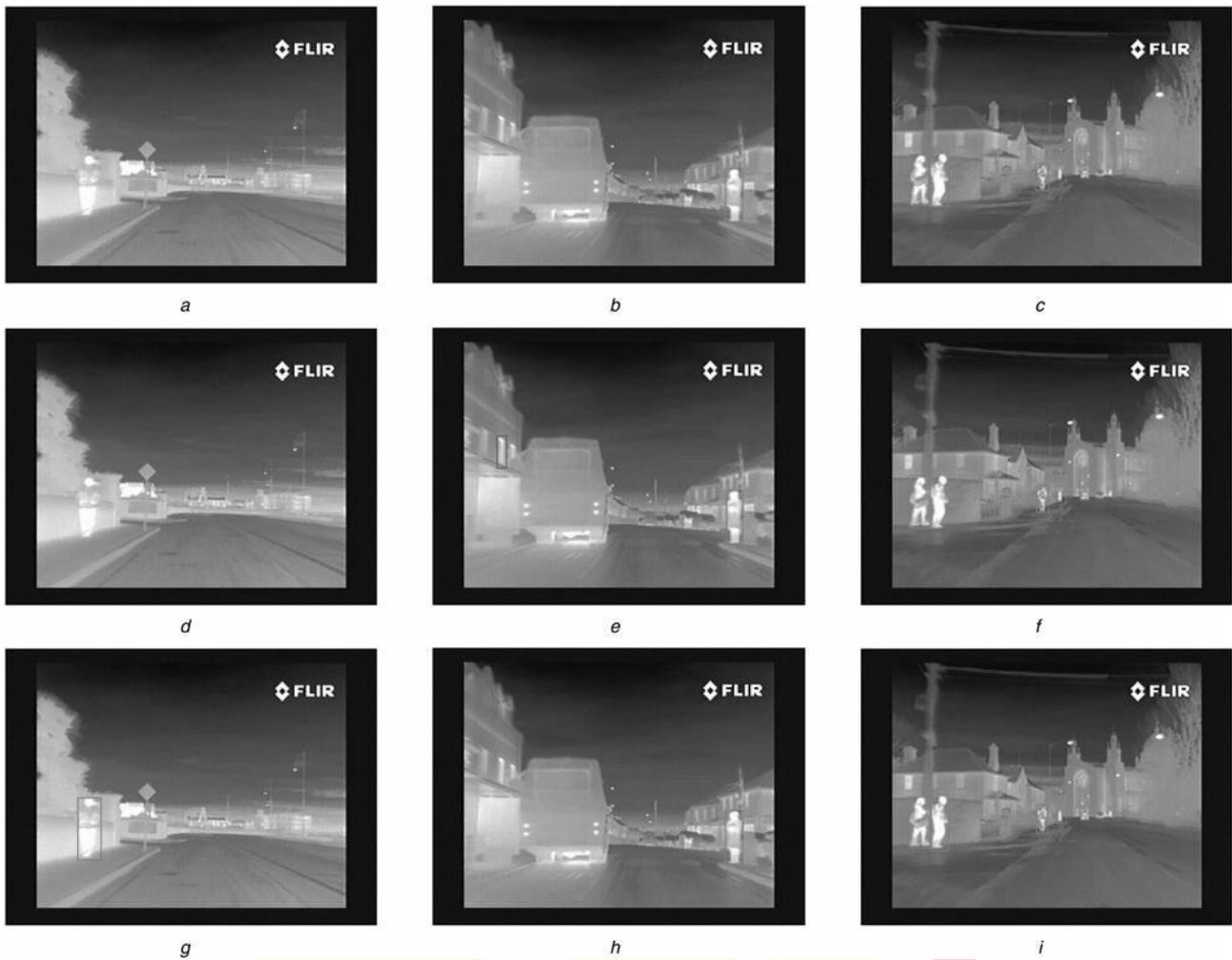
Fig. 8 *Output of SVM classifier trained with different feature vectors*
a–c HOG + SVM
d–f LBP$_{8,1}$ + SVM
g–i HOG–LBP$_{8,1}$ + SVM
True positives are outlined in blue and false positives are outlined in red

presented in Fig. 8. It is difficult to classify pedestrians that are a great distance from the camera [35], mainly because of the relatively low resolution of FIR cameras. Pedestrians also have very little textural information when they are situated far from the camera.

5.3 Discussion of results

A sample of frames processed using each feature vector are shown in Fig. 8: the HOG-SVM output is shown in Figs. 8*a–c*. The LBP$_{8,1}$ + SVM output is shown in Figs. 8*d–f*, the HOG–LBP$_{8,1}$ + SVM output is shown in Figs. 8*g–i*.

All of the features examined within this paper function quite well at close to medium ranges (10–50 m) in environments where there is a small number of other hot objects and there is a significant contrast between the pedestrian and the surrounding environment. However, if there are a large number of bright objects in the centre of the image, such as in a busy urban environment containing a large number of windows on heated dwellings and other vehicles on or near the road, the HOG + SVM and LBP$_{8,1}$ + SVM detection rates can decrease. Fig. 8*e* shows a false

positive caused by heat from the window of an adjacent building.

A number of IR pedestrian detection algorithms are presented and compared in Table 6. The HOG–LBP method of classifying pedestrians in FIR imagery presented in this paper achieves higher detection rates than other previously published methods in this table. The closest detection rate to the HOG–LBP rate of 98% achieved in this work is 95% [26, 38]. HOG–LBP, with a monocular IR camera even out-performs some stereo IR configurations [39]. Using a single camera results in a system that is cheaper to manufacture, easier to install and calibrate and, consumes less power.

The HOG–LBP trained SVM used in this pedestrian detection system demonstrates the ability to detect pedestrians across a wide range of distances from very close to the vehicle, to beyond 75 m (where a pedestrian is little more than a few pixels high). Early detection of pedestrians gives the driver more time to react and adjust the path of the vehicle to avoid collision with an identified pedestrian. Failure to detect pedestrians is most commonly caused by occlusion or close proximity to other bright objects in the surrounding environment. The greatest improvement was observed in the tests performed with video streams using

Table 6 Summary of pedestrian detection methods used in IR spectrum

Authors	Sensor config.	ROI isolation	Object classification	Tracking	Detection rate, %
Bertozzi *et al.* [36]	stereo	dynamic binary threshold + disparity information	Silhouette matching	–	90
Xu *et al.* [12]	mono	dynamic threshold	SVM	Kalman, Mean Shift	95
Bertozzi *et al.* [37]	stereo	–	histogram of oriented gradients + SVM	–	91
Ge *et al.* (2009) [33]	mono	dynamic binary thresholding	histogram of oriented gradients + Haar-Based AdaBoost	Kalman	93
O'Malley *et al.* [26]	mono	seeded region growing	histogram of oriented gradients + SVM	Kalman	95
Sun *et al.* [38]	mono	SUSAN	histogram of oriented gradients + SVM	–	95
Hurney *et al.*	mono	seeded region growing + distance estimation filtering	(histogram of oriented greadients + local binary patterns) + SVM	Kalman	98

the HOG–LBP trained classifier. The number of false positives encountered by the system was much lower in comparison with the HOG only trained classifier and LBP only trained classifier.

6 Summary and conclusion

This paper has presented a novel method of classifying pedestrians in FIR imagery using HOG–LBP feature vectors to train a SVM. Pedestrians are first isolated from the background scenery using a morphological closing operation to remove clothing distortion present in the image. Seeded region growing is then used to grow seed regions within the image. This is performed by using binary thresholds of lowering intensities. When seeds reach a set of geometric criteria their co-ordinates are examined and further filtered based on their location and size within the frame. The regions that satisfy these requirements are saved and used to segment the ROI for classification. HOG and LBP features are then calculated for each segmented region and are concatenated to form a single HOG–LBP feature vector. The feature is then passed to a SVM which has been trained with a database of 2000 pedestrian and non-pedestrian images and is classified as either pedestrian or non-pedestrian. If an ROI has been classified as a pedestrian in multiple consecutive frames, it is tracked with a Kalman filter. The location of the ROI can then be estimated in frames where detection may have failed because of occlusion of the pedestrian or a false classification by the SVM.

The SVM trained with HOG–LBP features achieves higher detection rates than previous literature in the area which utilised exclusively HOG or LBP feature vectors to train a SVM classifier for pedestrian classification in FIR imagery. The HOG–LBP classifier achieves a true positive detection rate of 98% which is an improvement of 4% on the classifier trained with HOG features alone (94% TP) and an improvement of 7% on a classifier trained with LBP features alone (91% TP). The HOG–LBP trained SVM also achieved a lower false positive rate. The detection rate of the HOG–LBP classifier is also an improvement on previous literature which achieved a maximum detection rate of 95% [12, 26, 38]. The combination of this improved pedestrian detection system with more stringent driver testing, road infrastructure and policing could significantly reduce the number of road fatalities that occur during the hours of darkness resulting in a much safer road environment for pedestrians.

Work is currently underway to implement the proposed algorithm on an embedded hardware platform to achieve real-time performance.

7 Acknowledgment

8 References

1 European New Car Assessment Program (Euro NCAP): http://www.euroncap.com/, accessed January 2014
2 Hobbs, A.: 'Euro NCAP/MORI survey on consumer buying interests (speech and presentation)'. Proc. Euro NCAP Conf.: Creating a Market for Safety 10 years of Euro NCAP, Brussels, Belgium, November 2005
3 European Road Safety Observatory: 'Traffic Safety Basic Facts 2012 – pedestrians', March 2012
4 O'Cualain, D., Hughes, C., Glavin, M., Jones, E.: 'Automotive standards-grade lane departure warning system', *IET Intell. Transp. Syst.*, 2012, 6, pp. 44–57
5 Chan, Y.M., Huang, S.S., Fu, L.C., Hsiao, P.Y., Lo, M.F.: 'Vehicle detection and tracking under various lighting conditions using a particle filter', *IET Intell. Transp. Syst.*, 2012, 6, pp. 1–8
6 Kaparias, I., Bell, M.G.H.: 'Testing a reliable in-vehicle navigation algorithm in the field', *IET Intell. Transp. Syst.*, 2012, 36, pp. 314–324
7 Unger, C., Wahl, E., Ilic, S.: 'Parking assistance using dense motion-stereo', *Mach. Vis. Appl.*, 2011, 6, pp. 1–21
8 O'Malley, R., Glavin, M., Jones, E.: 'Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions', *IET Intell. Transp. Syst.*, 2011, 5, pp. 1–10
9 Hanqvist, M.: 'An object detection system', Autoliv Department AB, May 2008, WO 2008/057042
10 Fang, Y., Yamada, K., Ninomiya, Y., Horn, B., Masaki, I.: 'Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection'. IEEE Intelligent Vehicles Symp., June 2003, pp. 505–510
11 Fang, Y., Yamada, K., Ninomiya, Y., Horn, B., Masaki, I.: 'Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection', *IEEE Trans. Veh. Technol.*, 2004, 53, pp. 1679–1697
12 Xu, F., Liu, X., Fujimura, K.: 'Pedestrian detection and tracking with night vision', *IEEE Trans. Intell. Transp. Syst.*, 2005, 6, pp. 63–71
13 O'Malley, R., Glavin, M., Jones, E.: 'An efficient region of interest generation technique for far-infrared pedestrian detection'. ICCE Int. Conf. Consumer Electronics, January 2008, pp. 1–2
14 Sivaraman, S., Trivedi, M.: 'Active learning for on-road vehicle detection: a comparative study', *Mach. Vis. Appl.*, 2011, 6, pp. 1–13
15 Gavrila, D.M., Giebel, J.: 'Shape-based pedestrian detection and tracking', *IEEE Intell. Veh. Symp.*, 2002, 1, pp. 8–14
16 Shashua, A., Gdalyahu, Y., Hayun, G.: 'Pedestrian detection for driving assistance systems: single-frame classification and system level performance'. IEEE Intelligent Vehicles Symp., June 2004, pp. 1–6

17 Dalal, N., Triggs, B.: 'Histograms of oriented gradients for human detection'. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, June 2004, vol. 1, pp. 886–893

18 Chang, F., Yang, X., Wu, W.P., Cho, Y.A., Chen, S.W.: 'Night-time pedestrian detection using thermal imaging based on HOG feature'. Int. Conf. System Science and Engineering, June 2011, pp. 694–698

19 Xia, D., Sun, H., Shen, Z.: 'Real-time infrared pedestrian detection based on multi-block LBP'. 2010 Int. Conf. on Computer Application and System Modeling (ICCASM), October 2010, vol. 12, pp. 139–142

20 Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, 24, pp. 971–987

21 Ahonen, T., Hadid, A., Pietikainen, M.: 'Face description with local binary patterns: application to face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, 28, pp. 2037–2041

22 Li, S.Z., Zhang, L., Liao, S., *et al.*: 'A near-infrared image based face recognition system'. IEEE Seventh Int. Conf. Automatic Face and Gesture Recognition (FGR), April 2006, pp. 455–460

23 Wang, X., Han, X., Yan, S.: 'An HOG-LBP human detector with partial occlusion handling'. IEEE 12th Int. Conf. Computer Vision, October 2009, pp. 32–39

24 Guo, L., Li, L., Zhao, Y., Zhang, M.: 'Study on pedestrian detection and tracking with monocular vision'. 2010 Second Int. Conf. Computer Technology and Development (ICCTD), November 2010, pp. 466–470

25 Cheng, W.C., Jhan, D.M.: 'A cascade classifier using Adaboost algorithm and support vector machine for pedestrian detection'. 2011 IEEE Int. Conf. Systems, Man, and Cybernetics (SMC), October 2011, pp. 1430–1435

26 O'Malley, R., Jones, E., Glavin, M.: 'Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation', *Infrared Phys. Technol.*, 2010, 53, pp. 439–449

27 Tsimhoni, J.B.O., Minoda, T., Flanagan, M.J.: 'Pedestrian detection with near and far infrared night vision enhancement', *IEEE Trans. Intell. Transp. Syst.*, 2004, 6, pp. 63–71

28 Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints'. Int. Journal of Comp. Vision, 2004, pp. 91–110

29 Zhiguo, N., Xuehong, Q.: 'Facial expression recognition based on weighted principal component analysis and support vector machines'. 2010 Third Int. Conf. Advanced Computer Theory and Engineering (ICACTE), 2010, vol. 3, pp. 174–178

30 Molina, J., Escudero-Violo, M., Signoriello, A., *et al.*: 'Real-time user independent hand gesture recognition from time-of-flight camera video using static and dynamic models', *Mach. Vis. Appl.*, 2011, 6, pp. 1–18

31 Welch, G., Bishop, G.: 'An introduction to the Kalman filter', Tech. Rep. TR 95–041, University of North Carolina at Chapel Hill, Department of Computer Science, 2003

32 Teoh, S., Brunl, T.: 'Symmetry-based monocular vehicle detection system', *Mach. Vis. Appl.*, 2011, 6, pp. 1–12

33 Ge, J., Luo, Y., Tei, G.: 'Real-time pedestrian detection and tracking at nighttime for driver-assistance systems'. IEEE Intelligent Vehicles Symp., June 2004, pp. 1–6

34 Bertozzi, M., Broggi, A., Grisleri, P., Tibaldi, A., Rose, M.D.: 'A tool for vision based pedestrian detection performance evaluation'. 2004 IEEE Intelligent Vehicles Symp., June 2004, pp. 784–789

35 Bertozzi, M., Broggi, A., Ghidoni, S., Meinecke, M.M.: 'A night vision module for the detection of distant pedestrians'. 2007 IEEE Intelligent Vehicles Symp., June 2007, pp. 25–30

36 Bertozzi, M., Broggi, A., Lasagni, A., Rose, M.D.: 'Infrared stereo vision-based pedestrian detection'. 2005 IEEE Intelligent Vehicles Symp., June 2005, pp. 24–29

37 Bertozzi, M., Broggi, A., Del Rose, M., Felisa, M., Rakotomamonjy, A., Suard, F.: 'A pedestrian detector using histograms of oriented gradients and a support vector machine classifier'. IEEE Conf. Intell. Trans. Syst., October 2007, pp. 143–148

38 Sun, H., Wang, C., Wang, B.: 'Night vision pedestrian detection using a forward-looking infrared camera'. 2011 Int. Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011, pp. 1–4

39 Suard, F., Rakotomamonjy, A., Bensrhair, A., Broggi, A.: 'Pedestrian detection using infrared images and histograms of oriented gradients'. IEEE Intelligent Vehicles Symp., June 2006, pp. 206–212