2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Multi-feature fusion based region of interest generation method for far-infrared pedestrian detection system

Zhiling Wang, Linglong Lin, Yuxin Li

Abstract—**Given the uncontrolled outdoor environments and different physical properties of clothing, the appearance of pedestrians in far-infrared (FIR) images changes dramatically. Finding a robust region of interest (ROI) generation method for pedestrian detection remains challenging. Previous researches can obtain reliable results in some conditions. But they always got inappropriate results in warmer conditions. This study presents a Multi-feature fusion based ROI generation method for FIR pedestrian detection system to solve this problem. We extract two kinds of salient feature regions, namely, highlighting feature areas and vertical feature areas. A reasonable threshold is set to derive the highlighting feature areas and Scharr operator is used to find vertical edges. These areas are not necessarily connected to each other in the image. We think an upright pedestrian is a highly structured target consisting of highlighting feature areas and vertical feature areas. The distribution of feature areas is demonstrated using a skeleton model. So we apply the dilating morphological operation to ensure that these adjacent feature areas within pedestrians' regions will connect together. The size of the structuring element is set adaptively according to the size of feature areas. Finally, the experimental results show the robust performance of our method in different ambient conditions.**

I. INTRODUCTION

Pedestrian detection system (PDS) is a significant component of advanced driver assistance systems[1-5]. The PDS significantly improves driving safety and protects road users. Pedestrians, which are a kind of road users, are the most vulnerable traffic participants because they are often seriously injured when they are involved in traffic accidents, particularly at night[12-19]. The PDS can notify drivers when a pedestrian appears, thus helping to avoid traffic accidents.

Many automobile manufacturers equip their products with the PDS. The PDS usually uses cameras in the visible spectrum during the daytime. When the illumination condition is insufficient (i.e., night, rain, fog, and snow), far-infrared (FIR) cameras are particularly useful. The pixel value of objects with higher temperature is larger than objects with lower temperature in FIR images. Thus, pedestrians tend to be the "hot spot" in the image. FIR cameras also provide less color or texture information.

Based on these characteristics, many studies on FIR pedestrian detection have been conducted. Several techniques used in the visible spectrum can also be applied in FIR cameras. However, crucial techniques vary because their imaging principles are different. If pedestrians' locations can be found in advance, the detection system will have lower computation cost and more accurate results. In the FIR image, pedestrians tend to appear brighter than the surrounding environment. As a result, FIR PDS has region of interest (ROI) generate module to process the FIR image in the beginning. ROI generation is very important because it provides the detecting area for pedestrian classification. The result of the ROI generator directly affects the performance of the PDS.

Zhiling Wang, Linglong Lin are with Hefei Institute of Physical Science，Chinese Academy of Sciences and Anhui Engineering Laboratory for Intelligent Driving Technology and Appllication, Hefei, Anhui, China 230031, (Tel: +86-551-5393190; fax: +86-551-5393190; e-mail: zlwang@hfcas.ac.cn).

Yuxin Li are with Hefei Institute of Physical Science，Chinese Academy of Sciences and Anhui Province Key Laboratory of Intelligent connected technology, Hefei, Anhui, China ,230031

Generation methods based on threshold are widely used and they can get remarkable result already. Junfeng Ge[6] used a dual-threshold segmentation algorithm to generate ROIs. He thinks that the pedestrians appears brighter than environment from the view of horizontal lines. So he calculates a high and a low threshold from the horizontal scan lines. Finally he filtered and generated the candidate areas with the aspect ratio of pedestrians. Some methods generate ROIs using horizontal or vertical projection. Many researchers extract ROIs with more than one kind of method. Q. Liu et al.[10] used a gradient vertical projection curve to subdivide the image. The curve presents the number of pixels with high gradient magnitude. Then a local dual-threshold segmentation algorithm is adopted. These two thresholds are calculated by using the horizontal and vertical neighborhood pixels around the anchor point. Connected regions that meet the requirements (i.e., aspect ratio and distribution location of pedestrians) are finally selected as ROIs. J. Li et al.[7] projected the horizontal and vertical curve roughly at first. Then, an additional intensity-oriented projection method is employed on the basis of the width/height ratio of candidate ROIs to obtain more accurate results. R. O'Malley et al.[11] preprocessed the FIR image using morphological closing operation with a flat vertical rectangle to compensate for the influence of distortion. Then ROIs are generated using feature-based region growing method with high-intensity seeds. The mean shift segmentation algorithm was used by D. Kim [8]and K. Lee[9]. However, pedestrians are usually split into several segmentation parts. Thus, a segment-based ROI generation method was employed. Only segments which are brighter than background are considered. The horizontal overlap ratio between segments is the rule used to determine whether to combine pairs of segments or not.

In this paper, we propose multi-feature fusion based region of interest generation method. The idea of our method is that the pedestrian can be regarded as a combination of different parts. Different kinds of unique feature types can be extracted from these parts. We can locate these parts in the image using two salient features which are also easy to compute, namely, highlighting feature and vertical feature. After these features are extracted, the pedestrians' areas can be located through the fusion of these feature areas. The adjacent feature areas can be connected using the dilating morphological operation. The size of structuring element is adaptive to the size of feature areas, making the ROI result more robust.

The remainder of this paper is organized as follows: We describe our ROI generation method in detail in Section 2. The experimental results are shown in Section 3, and the conclusions are drawn in Section 4.

II. PROPOSED METHOD

ROI generation is important because it provides the detecting area for classification. A good generation method should extract all pedestrians into ROIs. The number of negative ROIs should be suppressed as much as possible. When the only high intensity object is pedestrian or the contrast between pedestrian and environment is high, as shown in Fig1. In these conditions, many algorithms exhibit a good performance. However, due to FIR properties, when the ambient temperature is high, pixel values of pedestrian's may not be larger than background's. The pedestrian is not always brighter than the environment in FIR image. As we can see from Fig 2, there is no obvious contrast between pedestrian and environment. This situation mainly occurred at dawn or dusk. And it is more easily to happen when the climate temperature is higher. These images are taken when the temperature is around 20 centigrade. The intensity of dry road surface is matintained between 60 and 80. Fig 3(b). shows the histogram of this kind of image. As we can see, the intensity of pedestrian i similar to the environment's. Only a small part of points' intensity is higher than 90. There points are caused by street lights, automobile tires or engines and the body parts of pedestrian. The difficulty of ROI generation is raised.

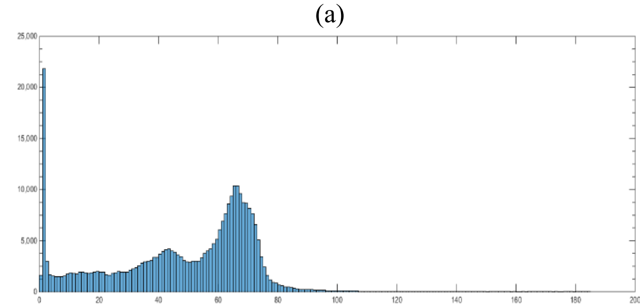
(a) (b)

Fig.1 Examples of far-infrared images. (a) Pedestrian is the only high intensity object. (b) The contrast between pedestrian and environment is high.

Fig 2. FIR images when contrast between pedestrians and environment is low.

The contrast between objects is low in the image, but their own characteristics will not be changed. We can still extract certain features from the object. Although different kinds of objects may be found using one feature, not all objects have the same features. A tall tree or a standing sculpture can have the same salient vertical features, but are not the hot spot in the FIR image. We can categorize objects into different groups using different features. Although different clothing and movement of arms and legs will change the pedestrian' appearance, as long as the pedestrian is upright, the relative position of parts mentioned above is invariant. We think an upright pedestrian is a highly structured target consisting of highlighting feature areas and vertical feature areas.

(a)



(b)

Fig 3. (a) An example of low contrast FIR image. (b) Histogram of the FIR image

A. Highlighting Feature Extraction

A basic characteristic of the FIR image is intensity. Objects above absolute zero emit different amounts of heat depending on its temperature. Objects with different temperature have different intensity in the image. As the temperature of human body is usually higher than the background's, pedestrians should be bright in the FIR image. Some clothes are designed to insulate heat and keep people warm and little infrared radiation is emitted. The warmer weather condition may influence the infrared signature of the pedestrian too. Thus, some parts of human body tend to be dark in the image. Different physical properties of clothing change the appearance of pedestrians greatly. Despite this, the exposed parts of pedestrians always have high intensity values. Pedestrian's head and hands will remain to be bright spot in the image. As a matter of fact, the situation that only the exposed parts of pedestrians have high intensity rarely happens. Several parts of the torso may also appear to be warm objects. Because these bright parts of pedestrians have high intensity, highlighting feature can be extracted from them.

We apply a single threshold to extract highlighting feature from the image. Because different contrast between pedestrian and environment means the different range of image intensity. The threshold T has to be adaptive to the mean intensity I_{mean}, the highest intensity is $I_{highest}$, and I_{lowest} means the lowest intensity. The threshold is set to segment the pedestrian, thus it should also be decided by pedestrian's average intensity I_{pavg}. The threshold T can be expressed as follows:

$$T = w_1 * I_{mean} + w_2 * I_{highest} + w_3 * I_{lowest} + w_4 * I_{pavg} \tag{1}$$

Usually the lowest pixel value can be 0, so the formula can be simplified as:

$$T = w_1 * I_{mean} + w_2 * I_{highest} + w_3 * I_{pavg} \tag{2}$$

where w_1, w_2, w_3 are the weights and they are within the range [0,1]. The optimal values of them have to be acquired through the experiments.

After we get the binary image, there are some noise regions need to be neglected. Each of these noise regions has little pixel points, thus the probability that a pedestrian can exist is low. Even if it belongs to a pedestrian, he should be very far from the vehicle. According to our FIR camera mounting position, when the image size is384×288, the head of pedestrian up to 60 meters away occupies less than 10 points in the image. So we filter the area whose number of points is less than 10. We denote these areas with highlighting features as S_I.

B. Vertical Feature Extraction

Another salient feature we extract from the image is vertical feature. An object with salient vertical features means it has a number of vertical edges. Typically, a standing pedestrian has many vertical edges.

Many edges are calculated by using Sobel operator or Canny operator. The Scharr operator appears to be more rotation invariant compared with Sobel operator. The gradient calculated by Scharr operator is larger than the gradient using Sobel operator. When the objects have obvious contrast against the background, the gradient calculated using Scharr operator will be truncate to 255, while the low contrast area maintains normal gradient value. The gradient value of high contrast areas is truncated, that is equal to extend the gradient value of low contrast area. When the contrast between some parts of pedestrian and the environment is not high, Scharr operator will leads to relatively higher gradient values, ensuring obvious boundaries between them. In short the Scharr operator helps us to extract edges when the contrast is low.

Thus we apply the Scharr operator to get horizontal difference G_x , as follows:

$$G_x = \begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix} * I \tag{3}$$

where I denotes the original image.

There will be a large number of meaningless points in gradient image. So after we get the gradient image, these interfering points need to be eliminated. We set half of the maximum absolute gradient value as the threshold to process the gradient image.

The next step is to find their vertical feature edges. We traverse the binary to find all edges. Not all members of the edges should be considered. These edges need to meet two requirements. Firstly, each edge should have a sufficient number of points. The surrounding of noise has similar appearance in the image, so the edge of it will have few numbers of points. Even if it is not an edge caused by noise, when it doesn't meet the first requirement, the edge cannot be extracted from the pedestrian. Because the vertical feature area extracted from the pedestrian not too far will not be small. As shown in Fig.4, when the image size is384×288, the pedestrian 60 meters away is nearly 35 pixels high. The length of arms, legs and torso of pedestrians there is around 10 pixels. Thus we set the minimum number of points of each edge to 20. This requirement helps us to filter out needless areas. Secondly, the edge should be more vertical than horizontal. We think the height of edge's area has to be larger than 1.5 times the width.

When the edge meets both two requirements, it will be preserved for the next step. These preserved areas with vertical features are denoted as S_v.

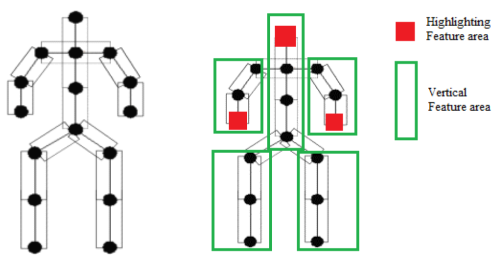
Fig 4. The pedestrian 60 meters away shown in the FIR image

C. Feature Region Fusion

The previously presented process acquires two kinds of salient feature regions. Because they are extracted by using different features, these areas may not be connected to each other. If these areas are regarded as ROIs separately, the pedestrian may be split into many parts. We consider pedestrians as a fusion of multi-feature regions. If we can fuse these feature areas correctly, ROIs containing pedestrians can be generated.

As shown in Fig.5 (a), an upright pedestrian is represented using a skeleton model. Such a body structure is unrelated to pedestrian's movement and posture. Highlighting feature and vertical feature can be extracted from different parts of the pedestrian. As we can see from Fig.5 (b), the top of the human is head, which is always the bright object in the FIR image. Areas below the head is human torso with obvious vertical feature and arms which are approximated as vertical object. Although when pedestrian is walking, there are some changes in the angle of the legs, the legs still have sufficient vertical features. Under ideal conditions, these feature areas are connected together in FIR images. Due to the influence of ambient temperature and light, sometimes there are gaps between two feature areas. In order to fill the gap between these regions, we apply morphological dilation method.

Morphological dilation operation will expand the boundaries of respective feature regions. It's obvious that the structuring element should be a vertical rectangle according to the minimum bounding box of pedestrian. Suppose there are two pedestrians standing at different distances and feature areas extracted are not intersecting. If we dilate such two different scales of pedestrians using the same structuring element, it may lead to unexpected results. If the scale of pedestrian is small while the size of structuring element is very large, although feature areas can be connected together, there will be many superfluous areas in ROIs. If the size of structuring element is not big enough compared to the scale of pedestrian, feature areas may not be connected at all. Thus the most appropriate way is to use structuring element of different sizes for each person. We have to calculate the suitable size of the structuring element according to different feature areas.



(a) (b)

Fig 5. Skeleton Model of the pedestrian and schematic diagram of feature areas.

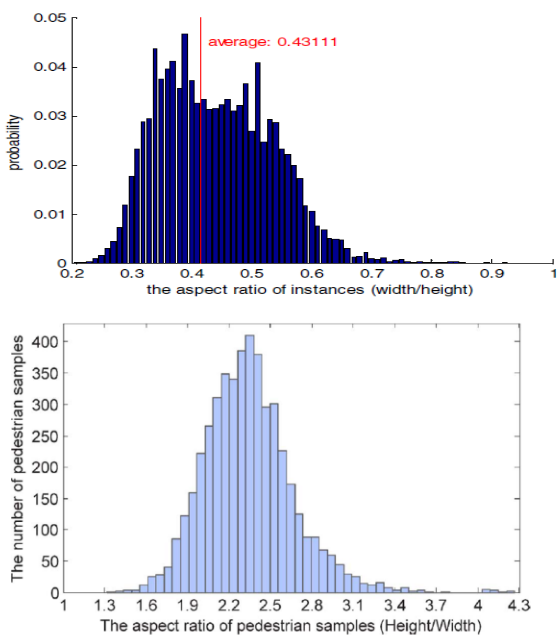


Fig 6. Pedestrians' aspect ratio statistics in others studies

Most aspect ratio of pedestrians (Width/Height) usually follows a certain statistical distribution[6,10], ranging from 0.3 to 0.7 in Fig 6. We select a moderate aspect ratio (i.e. 0.5) to set the size of structuring element and we think that if the width or height of the element is half that of the feature areas, the adjacent areas will certainly be dilated together. However, the width or height of the element should not be too small or the process will be meaningless. The size of the element should not be too large and that will avoid the risk to merge too many feature areas together.

For a given region in the image whose size is 384×288, we let w and h indicated the width and the height of the element E, as follows:

$$w = \begin{cases} 5 & W_R \leq 5 \\ W_R/2 & 5 < W_R \leq 10 \\ 10 & W_R > 10 \end{cases} \quad h = \begin{cases} 10 & H_R \leq 10 \\ H_R/2 & 10 < H_R \leq 20 \\ 20 & H_R > 20 \end{cases} \tag{4}$$

where W_R and H_R are the width and height of the corresponding areas respectively. As we explained above, every single feature area of the pedestrian 60 meters away is about 5 pixels width and 10 pixels height. Although the size of the pedestrian near car can be up to 170×420 in Fig 7, the gap between feature areas won't be too large. We think the maximum width and height of the element is adequate to connect areas together.

Fig 7. The pedestrian near the car shown in the FIR image

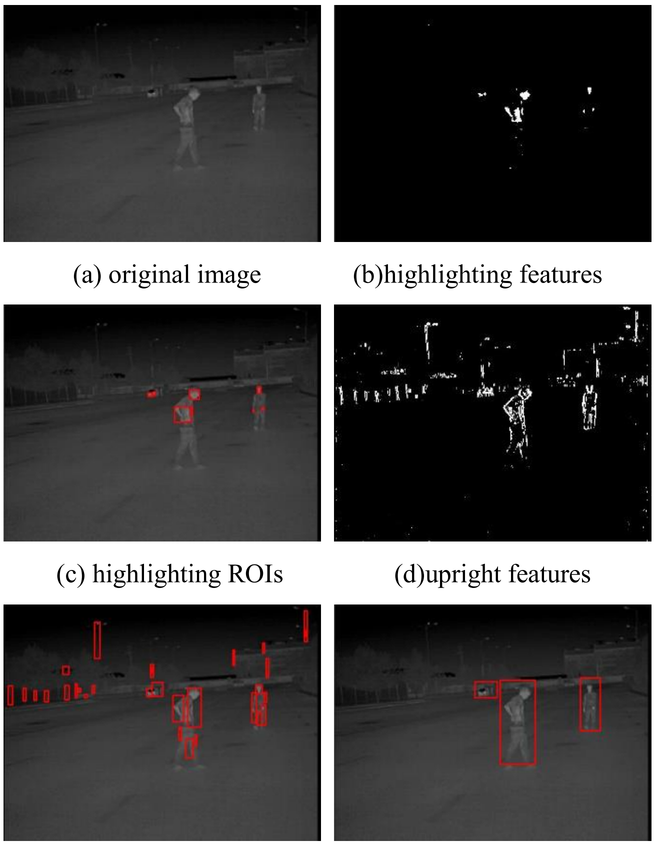
When each structuring element is decided, the dilating morphological operation can be used to generate the final ROIs. The dilation will expand the candidate region, ensuring adjacent highlighting feature area and vertical feature areas connected. As long as the minimum bounding boxes overlap, we consider these areas are connected.

Heads, hands and some body parts are bright objects and highlighting feature areas are always contained in vertical feature areas, as we mentioned in Fig 5. So if we just dilate vertical feature areas, the low computation cost will be ensured. The ROIs is generated as follows:

$$S_{ROI} = S_I \cup (S_V \oplus E) \tag{5}$$

where S_{ROI} is the final result of proposed generation method and E is the structuring element with width w and height h corresponding to each feature-based areas.

Our multi-feature fusion method aims to ensure all pedestrians are contained in the ROIs and improve the performance when the contrast between pedestrians and environment is low. An example of generating ROIs using our method in that condition is shown in Fig 8. With the help of the morphological operations, separate neighboring parts are combined. Remaining areas after all these steps are considered as the ROIs. Small areas are removed when extracting feature. Several ROI generation approaches limit the aspect ratio of the bounding box while our method does not need to. Existing areas are connected based on the structure of pedestrians. The dilating operation will expand the area of the target, but only a few extra background environments are included in the ROIs. Our method provides a raw but fairly accurate result, and other processes (e.g., filter ROI with size requirements) can be applied to detect specific areas.



(a) original image (b)highlighting features

(c) highlighting ROIs (d)upright features

(e) upright ROIs (f)ROIs by feature fusion

Fig 8. An example of generating ROIs using our method when there is low contrast between pedestrians and the environment.

III. EXPERIMENTAL RESULT

In order to evaluate the performance of our proposed method, experiments are conducted on the platform of our autonomous vehicles. The mounting position of FIR camera is shown in Fig 9.

Fig 9. Experiment platform and FIR camera mounting position

Our experimental environment is mainly clear weather at night. We collect the test data and evaluate the algorithm at the different Seasons. We not only test the performance in simple, complex background when the vehicle is stopped, but also dynamically evaluate our method in a residential environment and an urban traffic environment. Before we provide the statistical evaluation, we need to know how to evaluate the result first.

A. Performance Evaluation Criteria

The most intuitive evaluation method is to calculate how many pedestrians are not included in the ROIs. Because the large range of occlusion can lead to a serious lack of pedestrians' infrared signature and if the exposed parts of pedestrians cannot be seen, it's difficult for the algorithm to generate ROI areas based on missing information. Thus we assume that if more than 80% of the pedestrian is occluded, they are not included in the statistics. The miss rate MR is denoted as follows:

$$MR = \frac{N_{p \notin ROI}}{N_p} \tag{6}$$

where $N_{p \notin ROI}$ represents the number of pedestrians not included in ROIs. If less than 20% areas of a pedestrian is contained in a ROI, we consider the pedestrian as a missed target. N_p means the number of pedestrians in the image. The miss rate is an easy way to evaluate the algorithm, while this approach does not reflect the specific performance of the algorithm. Thus we need an evaluation method which is more comprehensive.

The ROI generation method inputs an image and returns an image win ROI bounding boxes (BB_{ROI}) on it. Another kind of bounding boxes we use is ground truth bounding boxes (BB_{GT}) which represent the minimum rectangle areas of pedestrians. The ideal situation is $BB_{ROI} = BB_{GT}$. In fact, current algorithm cannot achieve such a result, usually BB_{ROI} and BB_{GT} overlaps. If the overlap rate is larger than the threshold, we think the area of pedestrian is successfully extracted and BB_{ROI} is a positive match to BB_{GT}. Each BB_{ROI} and BB_{GT} can only be matched once at most. If one pedestrian is split into several ROIs, we only consider the ROI with the highest score. The rate of overlap is denoted as follows:

$$R_O = \frac{area(BB_{ROI} \cap BB_{GT})}{area(BB_{ROI} \cup BB_{GT})} \tag{7}$$

where $area(x)$ means the areas of x in the image. As we can see, a higher value of R_O indicates a better performance of the method. We think the overlap rate must exceed 50% just like the parameters in D.S. Kim's study [8].

B. The Result of Experiment

In order to get the optimal parameters for highlighting feature areas, we tried different values of w_1, w_2, w_3 to get the suitable threshold. These parameters are tested ranging from 0.05 to 1 and the step is 0.05. We calculated the overlap rate of the highlighting feature areas and pedestrian's bright areas. When the maximum overlap rate is achieved, the corresponding w_1, w_2, w_3 will be selected. The average intensity I_{pavg} of 150 positive pedestrian samples is 65. No matter what the contrast between pedestrians and background is, the average intensity of FIR image is within the range from 40 to 60. It is found that we can get the maximum overlap rate when $w_1 = 0.35, w_2 = 0.4, w_3 = 0.4$.

TABLE I. STATISTICAL RESULT IN THE FOUR SEASONS

Experiment environment	Frame numbers	Annotated pedestrians	MR of our method	R_O of our method
Winter night	1179	375	0%	94.0%
Autumn night	1041	792	0%	83.6%
Summer night	2011	388	18%	66.9%
Spring night	1388	1502	1.6%	78.3%

After we get the suitable parameters, we conducted our experiments at night in the different seasons. The statistical results are shown in the table I.

Fig 10. ROI generation result in winter.

Fig 10 shows the result when the car is stopped in winter. There is at least one pedestrian in each image. Pedestrians in different distances away represent three states (i.e., far, mid and near). As it is winter evening, the ambient temperature is very low and the far-infrared radiation emitted is not too much. As a result, the intensity of environment in image is low. The pedestrians are moving heat source. The intensity of exposed parts will be much higher than the background. Although the physical properties of clothing will change pedestrian's appearance, it's sure that the overall intensity of pedestrian is higher than the environment. As long as the pedestrian is not too far away or seriously blocked, the ROI generation method can easily extract pedestrians' regions. As we can see from the result, our method does not miss any one. Pedestrians nearby occupied a large part of the image, but the contrast is very obvious. So good segmentation results are achieved. Although pedestrians in distance tend to blend with the background, the fusion of multi-feature is also able to extract pedestrians' area. These feature areas are expanded using dilating operation. BB_{ROI} is slightly larger than BB_{GT}. Thus the overlap rate is down to 94%

Different clothing and physical condition will influence FIR radiation greatly in autumn. Pedestrians' appearance is a distribution of light and dark areas. As we can see in Fig 11, there is a clear change in brightness between the man nearby and the woman in distance. The only bright areas of woman are hands, legs and the head. Our generation method extracts all pedestrians into ROIs. It can be seen that the algorithm has a good adaptability to pedestrians at different distances. In the first row of the figure, vertical feature areas extracted from the vehicle far away is close to pedestrians' position. Feature areas of them are enlarged during the fusion. As a result, pedestrian's areas and vehicle' areas are connected to each other. The size of ROI generated is larger than normal. Moreover, vertical feature areas extracted from the feet of pedestrian lower left corner is very small while the scale of pedestrian is large, so the feet is not included in BB_{ROI}. Thus the MR of test data in autumn is 0% and R_O is just 83.6%

Fig 11. ROI generation result in autumn.

The temperature in summer is so high that the contrast between pedestrians and environment is low. There is not too much difference in infrared emission between them. Only heads of pedestrians and high temperature areas of vehicles are hot spots in the image. Although the contrast is low and the environment is complex, the multi-feature fusion method still has a good performance. The result can be seen in Fig 12. Since highlighting feature can be extracted from some parts of pedestrian's torso, highlighting feature areas and vertical feature areas can be connected together. These two features can be extracted from a moving vehicle and it's extracted as ROI. Because the pedestrian is close to the car in the image, two bounding boxes are considered as one ROI. Thus BB_{ROI} is larger than expected.

Fig 12. ROI generation result in summer.

Fig 13. ROI generation failures in summer.

However, when the contrast is low, if the pedestrian is very close to the car, the extracted feature area, especially the vertical feature area, cannot reflect the actual pedestrians' scale, which will lead to the failure of the algorithm. As we can see from Fig 13, the leg of pedestrian nearby is almost bended with the road. Areas that can be extracted vertical feature are relatively small. Only the lower legs can be located. The pedestrian is divided into two parts at his waist region because of large area of low intensity. As a result, the fusion between upper and lower parts is failed. The overlap rate between BB_{ROI} and BB_{GT} is less than 50%, so he is defined as a missing target. Similarly, the pedestrians' legs are not included in ROIs. The overall rate of overlap is decreased. The overall MR of the experiment in summer is 18% and R_O is around 67%. The algorithm still shows the performance of ROI generation in low contrast conditions.

We carried out an experiment in a residential environment in spring. Since the temperature at night is low, most interfering object in that district is the vehicles. Others are the walls of houses and street lamps. Fig 14 shows the example of them.

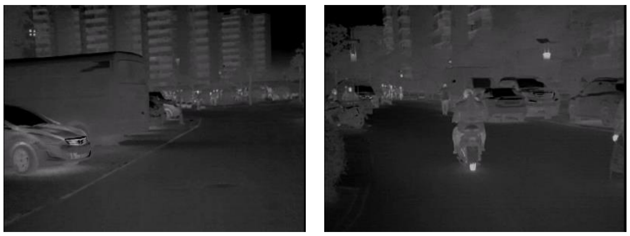


Fig 14. Interfering objects in a community

The residential environment is a real, complex road traffic environment. Due to large number of people within the community, there will be a lot of people showing up at the same time in the image. Many pedestrians walk among the traffic and it's easy for them to be occluded. So we ignored some pedestrians who are seriously blocked. We annotated 1502 pedestrians in 1388 images. As we can see from Fig 15, although the background is complex, our algorithm can still generate BB_{ROI} correctly. Our method includes almost all people in the ROIs. Only 1.6% people are missed because they are too far away. Some cars have high intensity and salient vertical feature, so they are contained in ROIs. That's the mainly reason why the overlap rate is 78.3%.

Fig 15. ROI generation result in spring.

C. Comparison with other algorithm

Apart from conducting our own experiments, comparison with others should be made to show the performance of proposed method. Liu Qiong's algorithm used a curve projection method and dual-threshold method. The experimental results in his study seem to be effective and promising. So we compare our method with his. In his study, the optimal settings of parameters and determined empirically. So we followed the conservatively settings as his instruction. Parameters used in projection curve is $w = 1, T_g = 20$, while $\alpha,\ \beta$ and ω used in dual-threshold is set to 2, 8 and 12.

Table II show the statistical result in the winter and summer. Fig 16 shows the results of our method and Liu Qiong's in winter. First row is ours while the second row is his result.

Fig 16. Comparison of ROI generation results in winter between proposed method in this study and Liu Qiong's method. Images in the first row are the result of our method and images in the second tow are the result of Liu Qiong's algorithm.

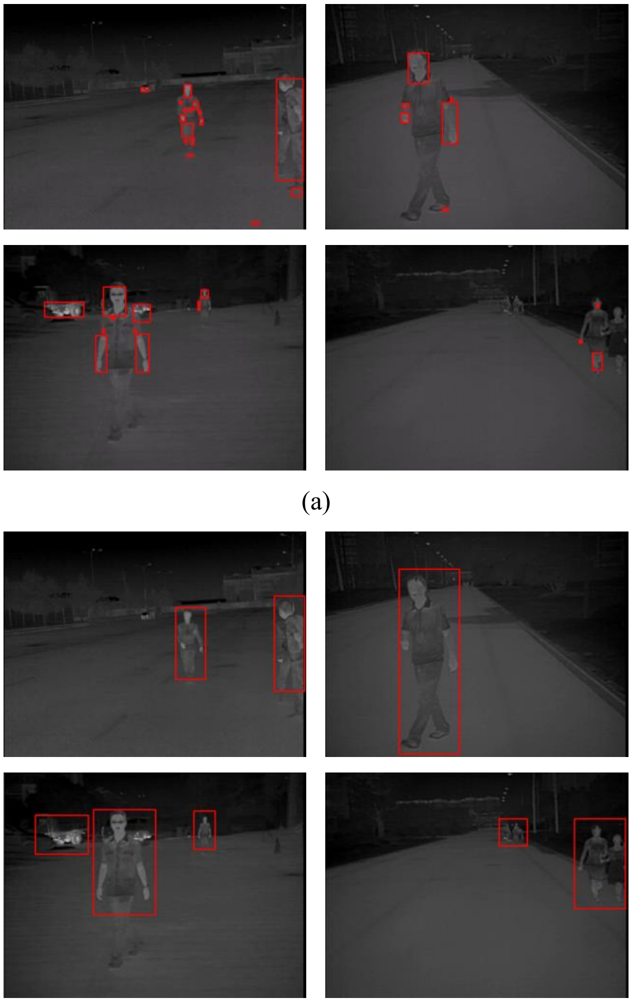
TABLE II. STATISTICAL RESULT OF LIU QIONG'S METHOD

Experiment environment	Frame numbers	Annotated pedestrians	MR of Liu Qiong's	R_O of Liu Qiong's
Winter night	1179	375	0%	97.3%
Summer night	2011	388	3.6%	6.1%

From the table and the image, we can clearly see that both algorithms have low miss rate in winter. The difference is that his method have more accurate ROI result in winter. As the pedestrians are clearly brighter than the background, the threshold can easily distinguish them from the environment. His ROI bounding boxes is almost the same as the ground truth areas. Our multi-feature is connected using dilating operation and the boundaries are expanded. So the overlap rate of ours is slightly smaller than Liu Qiong's. It seems there are not too many advantages of our method when the contrast is high.

But when the contrast between pedestrians and the environment is low, the proposed method shows the robust performance. As we can see from Fig 17(a), Liu Qiong's algorithm can hardly obtain accurate results. The ROIs they generated are small and distributed among the human body's area. When the contrast between pedestrians and background is low, the intensity between them is similar. There is a hidden premise that the contrast between them is high enough, or it's difficult to acquire satisfying results. The only feature he uses is intensity, as a result, his method doesn't have a good performance in summer.

We extract highlighting feature and vertical feature from the image. Although the highlighting feature areas we get may be the same as Liu Qiong's in that circumstance, the vertical feature areas can help us to locate the torso areas and some other areas. Thus even if the intensity of pedestrian is close to the environment's, we can still get robust ROI results in summer. More results can be seen in Fig 17(b).



(a)

(b)

Fig 17. Comparison of ROI generation results in summer between proposed method in this study and Liu Qiong's method. (a) ROI generation result of Liu Qiong's method. (b) ROI generation result of proposed method in this study.

IV. CONCLUSION

This study presents a multi-feature fusion based ROI generation method which aims to improve performance when the contrast between pedestrians and environment is low. As the appearance of pedestrians can be greatly affected by the ambient condition and physical properties of clothing, we extract two kinds of feature areas from the image. These features can be easily found from the pedestrians. We apply a single threshold to extract highlighting feature from the image and the threshold is determined by the mean intensity of sample pedestrians, the maximum intensity and the mean intensity of image. Vertical feature areas are extracted using Scharr operator. We show the distribution of feature areas using a skeleton model. As the human body is a highly structured object, we use the dilating morphological operation to ensure adjacent feature areas will connect together. The size of the structuring element is set adaptively according to the size of feature areas. This kind of process ensures that our proposed method can generate better ROIs results when the contrast is low. Although the experimental results of our approach are encouraging, further improvements regarding this work need to be done. If the pedestrian is too close in the situation when the contrast is not high, the vertical feature areas cannot reflect the actual pedestrians' scale. If there is a method can tell the pedestrians' scale, the ROI generated will be more accurate.

ACKNOWLEDGMENT

This work is supported by Key Program of 13th five-year plan, CASHIPS, with granted No. KP-2017-13; Anhui Province Science Foundation for Young Scientists with granted No. 1808085QF213; Anhui Major Science and Technology Programs with granted No. 15CZZ02037.

REFERENCES

[1] Brehar R, Vancea C, Nedevschi S. Pedestrian detection in infrared images using Aggregated Channel Features[C]. International Conference on Intelligent Computer Communication and Processing. IEEE, 2014:127-132.

[2] Chang S L, Yang F T, Wu W P, et al. Nighttime pedestrian detection using thermal imaging based on HOG feature[C]. International Conference on System Science and Engineering. IEEE, 2011:694-698.

[3] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Computer Society Conference on. IEEE, 2005:886-893.

[4] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: A benchmark[J]. Proc.conf.on Computer Vision & Pattern Recognition, 2009:304-311.

[5] Xu F, Liu X, Fujimura K. Pedestrian detection and tracking with night vision[J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(1):63-71.

[6] Ge J, Luo Y, Tei G. Real-Time Pedestrian Detection and Tracking at Nighttime for Driver-Assistance Systems.[J]. IEEE Transactions on Intelligent Transportation Systems, 2009, 10(2):283-298.

[7] Kai J, Arens M. Local Feature Based Person Reidentification in Infrared Image Sequences[C]. International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2010:448-455.

[8] Kim D S, Kim M, Kim B S, et al. Histograms of local intensity differences for pedestrian classification in far-infrared images[J]. Electronics Letters, 2013, 49(4):258-260.

[9] Kim D S, Lee K H. 2013. Segment-based region of interest generation for pedestrian detection in far-infrared images. Infrared Physics & Technology [J], 61: 120-128.

[10] Liu Q, Zhuang J, Ma J. 2013. Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems. Infrared Physics & Technology [J], 60: 288-299.

[11] O'malley R, Jones E, Glavin M. 2010. Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation. Infrared Physics & Technology [J], 53: 439-449.

[12] Qi B, John V, Liu Z, et al. Use of Sparse Representation for Pedestrian Detection in Thermal Images[C]. Computer Vision and Pattern Recognition Workshops. IEEE, 2014:274-280.

[13] Tsimhoni O, Bärgman J, Flannagan M J. 2007. Pedestrian Detection with near and far Infrared Night Vision Enhancement. LEUKOS [J], 4: 113-128.

[14] Vondrick C, Khosla A, Malisiewicz T, et al. HOGgles: Visualizing Object Detection Features[C]. International Conference on Computer Vision. IEEE, 2014:1-8.

[15] Yajun F, Yamada K, Ninomiya Y, et al. 2004. A shape-independent method for pedestrian detection with far-infrared images. IEEE Transactions on Vehicular Technology [J], 53: 1679-1697.

[16] Yang M Y, Yong X, Rosenhahn B. Feature Regression for Multimodal Image Analysis[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society, 2014:770-777.

[17] Yuan Y, Lu X, Chen X. 2015. Multi-spectral pedestrian detection. Signal Processing [J], 110: 94-100.

[18] Zhang K, Zhang L, Yang M H. Real-Time Compressive Tracking[C]. European Conference on Computer Vision. Springer-Verlag, 2012:864-877.