Forward Looking Infrared Target Matching Algorithm Based on Depth Learning and Matrix Double Transformation

Qiongfei WU[1], Yong ZHU[2], Yi CHEN[1], Zhiqiang ZHANG[1,]

[1]Wuhan Institute of Design and Sciences, Hubei Wu Han, 430205, China

[2]School of Math & Computer, Wu Han Textile University, Hu Bei Wu Han, 430200, China

Abstract. Targeting at the human target detection in infrared sequence images, the extraction method of feature region based on feature points is adopted. The depth-learning algorithm is firstly used to extract the feature points rapidly. Based on the feature points extracted by matrix double transformation, LBP algorithm is used to extract the feature region. After acquiring the feature region (ROI region) interested, feature extraction of wavelet entropy based on discrete wavelet transformation is conducted for ROI region. Then ROI region is classified through compound classification method.

Introduction

Because of the low contrast, edge blur, low signal-to-noise ratio, large noise and non-rigid body of human target, it is difficult to detect the human target in infrared image. Currently, the common detection method for moving target is background subtraction method [1], frame difference method [2] and optical flow method [3]. When the camera is not fixed, the target detection method based on background will become difficult. Therefore, it is more feasible to adopt the method of single-frame image segmentation. The detection method for human target in infrared sequence image adopted in this Paper: (1) The extraction of feature point is conducted in single-frame image by FAST *(Features from Accelerated Segment Test)* [4] algorithm. (2) Based on the feature point detected, the algorithm of kernel function based on Local Binary Pattern (LBP) is used to extract the local feature of image and acquire the region of interest in human. (3) Double-density dual-tree complex wavelet transformation is used to extract the feature information of training sample and the mode recognition method is adopted to extract the region of interest (ROI) for classification detection and acquire the region human target located.

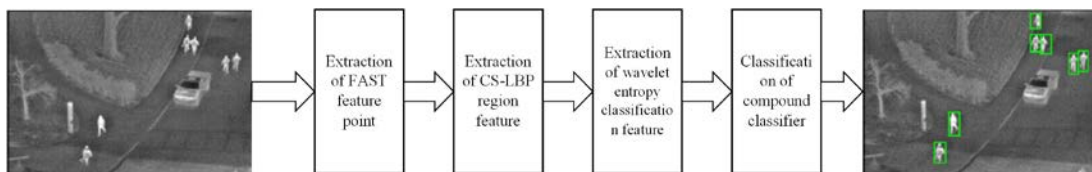## Initial Location of Human Target Region Based on Feature Points

Fig. 1 Detection method of human target and its effect in infrared image adopted in the paper

For the traditional target detection method based on video, when the camera is not fixed, the operand is large and the robustness is poor. Therefore, the detection for human target is conducted based on single-frame image. This Paper adopts the method of extracting the feature points in infrared image firstly, and searching for human target in neighborhood of feature points to extract the human target region.

Feature point extraction based on depth-learning algorithm

FAST algorithm is a detection method of feature point with simple operation. Firstly, the feature point is defined through *Segment-Test* algorithm: in *Bresenham* circle, centering on the pixel pint p with radius of r, if the brightness of more than N continuous pixels exceeds the certain threshold of pixel p or is lower than certain threshold of pixel p, then such pixel p is in a feature angle point. In

application, in order to reduce the quantity of feature points extracted, the neighboring feature points shall be combined, shown as Fig. 2. The left Figure is the original infrared image and the right Figure is the image after combining the feature points in neighboring space.

Fig. 2. Original Infrared Image (Left) and Feature Points Extracted (Right)

Regional feature extraction based on depth-learning and matrix double transformation algorithm

LBP algorithm is put forward as an operator of texture operation, which has been successfully applied in many computer vision fields.

LBP algorithm describes the feature of such pixel according to the relative difference between each pixel and its neighboring pixel. Therefore, it has strong robustness for the changes of pixel value caused by the change of ray condition. If the brightness value of neighboring pixel is larger or equal to the brightness value of center pixel, the value of attribute corresponding to such neighboring pixel is 1, otherwise the value is 0. *CS-LBP algorithm* [7] is to compare the brightness value between each pair of centrosymmetric pixels on circumference. Therefore, in *CS-LBP*, the time of comparison is reduced by half than LBP algorithm. In the Figure, the pixel *CS-LBP* value at (x, y) position is shown as formula (1) according to its neighboring information:

$$CS - LBP_{R,N,T}(x, y) = \sum_{t=0}^{(N/2)-1} s(n_i - n_{i+N/2}) 2^i$$

$$s(t) = \begin{cases} 1, & t \geq 0 \\ 0, others \end{cases} \tag{1}$$

Where, T is a small threshold. According to formula (1), $CS - LBP_{R,N,T}$ has $2^{N/2}$ possible value. Therefore, the feature dimension of statistical histogram of $CS - LBP_{R,N,T}$ value on whole image region is $2^{N/2}$. After the feature angle point is detected, an image region shall be defined around it. The vector shall be calculated according to the information in such region to describe the local vision feature of such region. According to the principle of CS-LBP algorithm, centering on such feature and taking 1 as the radius, the gray value of pixel in 8 neighborhood shall be compared to calculate its LBP histogram. According to the feature of histogram, the feature region around the feature angle point could be extracted according to the features of histogram, namely the candidate region of interest (ROI) extracted. Fig. 3 and Fig. 4 respectively shows the LBP histogram calculated and the ROI region extracted.
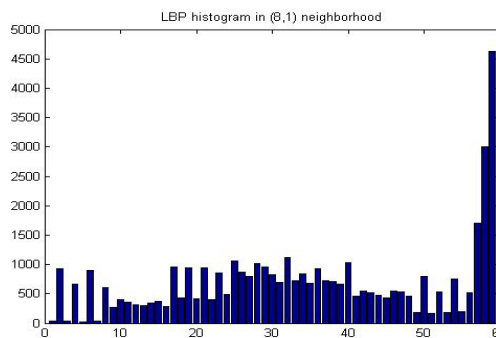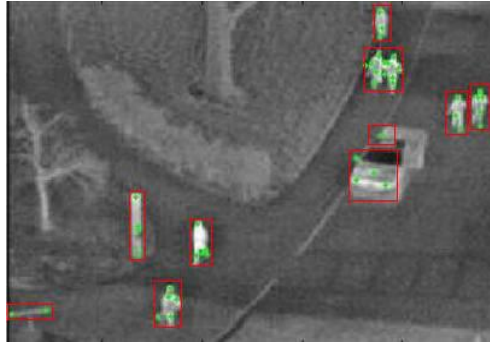

Fig. 3. LBP histogram in (8, 1) neighborhood

Fig. 4. ROI region extracted

Classification Detection of Candidate ROI

This Paper extracts the classification feature of training sample based on discrete wavelet transformation (DWT) and wavelet entropy. Based on this, in order to extract the human object from the selected region of interest more accurately, it also uses the mode recognition method of Boosting [9], support vector machine (SVM) [10] to constitute the compound classifier to classify the candidate ROI region. This Paper mainly trains three samples, such as human, vehicle and others (including house, trees and other non-human and non-vehicle target).

Double density dual-tree DWT [8] combines the advantages of double density and dual-tree wavelet, which has translation invariance, anti-aliasing feature, base-2 redundant and unchanged dimension, approximate continuous wavelet transformation and perfect direction feature. Double density dual-tree DWT (DD-DT DWT), has two different scaling functions $\phi_h(t), \phi_g(t)$ and four different wavelet functions $\psi_{h,i}(t), \psi_{g,i}(t), (i=1,2)$, where, $\psi_{.,1}(t)$ is constituted from 0.5 deviation of $\psi_{.,2}(t)$, namely:

$$\psi_{h,1}(t) = \psi_{h,2}(t-0.5)$$
$$\psi_{g,1}(t) = \psi_{g,2}(t-0.5)$$

(2)

Where, $\psi_{h,1}, \psi_{g,1}$ 和 $\psi_{h,2}, \psi_{g,2}$ is respectively the Hilbert transform pair, namely:

$$\psi_{g,1}(t) \approx H\{\psi_{h,1}(t)\}$$
$$\psi_{g,2}(t) \approx H\{\psi_{h,2}(t)\}$$

(3)

The image generates 4 low-frequency sub-bands and 32 high-frequency sub-bands through double-density dual-tree discrete wavelet transformation. The wavelet coefficient could reflect the detail information on 16 frequency bands and the wavelet entropy of image could reflect the energy distribution information of frequency space of wavelet transformation. The combination of double-density dual-tree discrete wavelet transformation and wavelet entropy has greater representational capacity.

The wavelet energy E_j at single scale j is defined as the quadratic sum of wavelet coefficient $C_j(k)$ at such scale:

$$E_j = \sum_{k=1}^{N} |C_j(k)|^2, (j=1,2,\cdots J)$$

(4)

Total energy of signal as:

$$E_{total} = \sum_{j=1}^{J} E_j = \sum_{j=1}^{J}\sum_{k=1}^{N} |C_j(k)|^2$$

(5)

Relative wavelet energy as:

$$P_j = \frac{E_j}{E_{total}} \quad j=1,2,\cdots J, \sum_{j=1}^{J} p_j = 1$$

(6)

If the wavelet coefficient matrix is regarded as a probability distribution sequence, the entropy

calculated by such sequence reflects the sparsity of such coefficient matrix, namely the order degree of probability distribution, which is the wavelet entropy. According to the principle of Shannon entropy, the wavelet entropy is defined as:

$$S_w = -\sum_{j=1}^{J} p_j \ln p_j \tag{7}$$

The machine-learning sample adopted by this Paper mainly comes from OTCBVS Benchmark Dataset database and the infrared images shot. Fig. 5 shows part of machine learning sample. Due to the demands of programming, the width of all the samples is the multiple of 4 and height has no such limit. Positive sample is expanded for 2 pixels based in human silhouette and the negative sample is generated randomly in existing image sequence and video by self program. Finally, 382 positive samples of human body and 288 positive samples of vehicle are acquired.
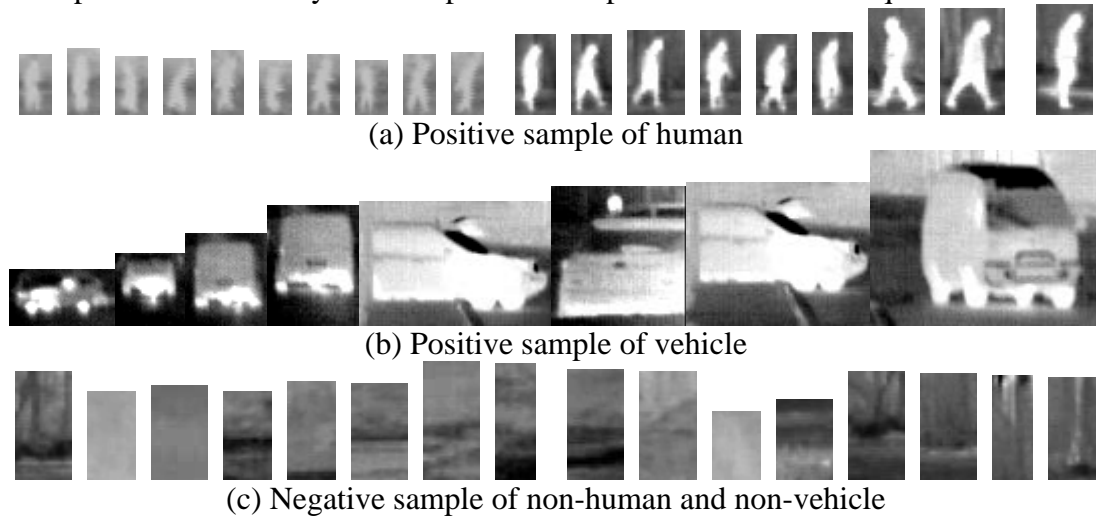

(a) Positive sample of human

(b) Positive sample of vehicle

(c) Negative sample of non-human and non-vehicle
Fig. 5 Parts of sample in training

Experimental Results and Analysis

In order to verify the performance of the method proposed in the Paper, we conduct on-site experiment. The experiments are conducted at the time with higher temperature in the afternoon and lower temperature in the evening. The experimental scenes respectively choose the filed space with relatively sample background and urban road scene with relatively complex background. All the sequence images are shot by fixed camera. The size of image is 384×288. The detection results for moving human target in infrared sequence image are shown as Fig. 6:
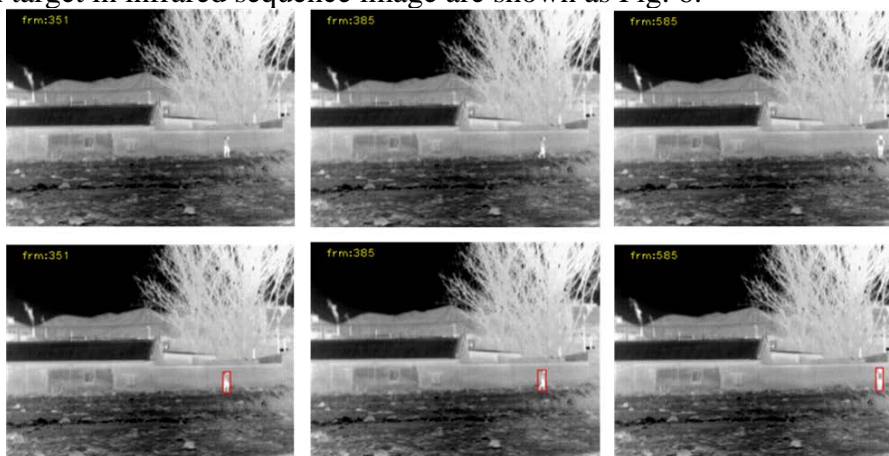

Fig. 6-1. Results of field experiment at daytime (with distance of 300 meters)
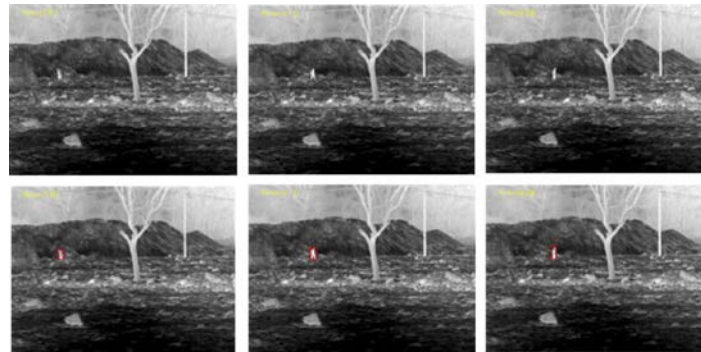
Fig. 6-2. Results of field experiment at night (with distance of 300 meters)
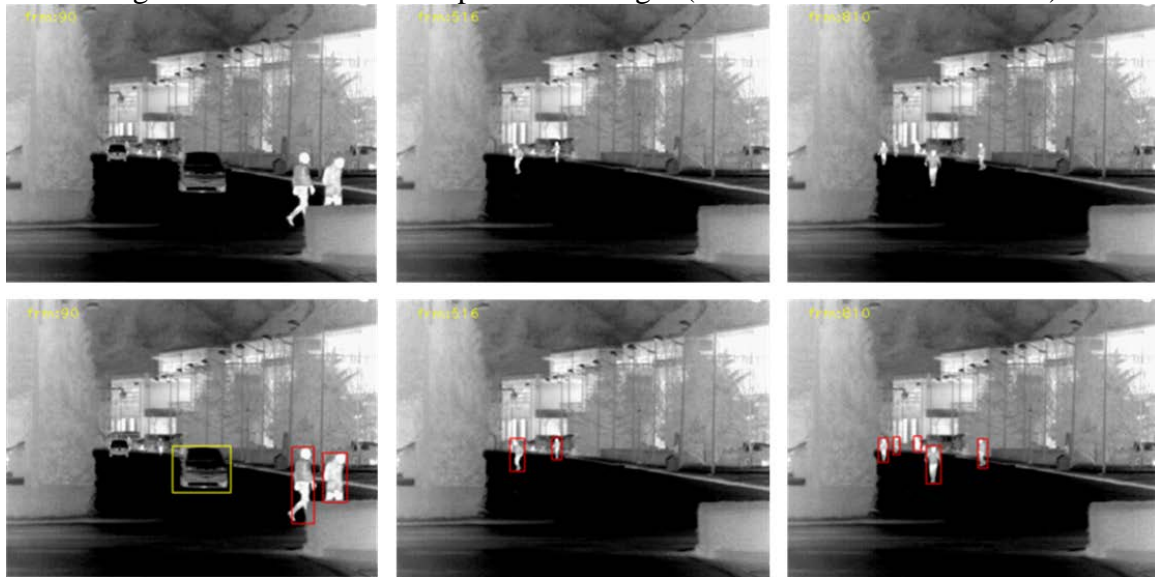

Fig. 6-3. Results of experiment at urban road at night (with distance of 50-300 meters)

According to the experimental results, the algorithm has higher recognition rate and low alarm rate in the environment with relatively simple background, which could accurately recognize the human target in infrared sequence image. However, such algorithm has defects as well. It is shown as Fig. 6-3 that in the environment with relatively complex background, there is detection leakage from 81st frame image to 90th frame image when the temperature is low. While in subsequent image sequence, the detection rate is higher.

Conclusion

Targeting at the features of infrared image, this Paper adopts the detection method for infrared human target based on depth-learning and matrix double transformation and the wavelet entropy based on discrete wavelet transformation for the extraction of classification feature, which makes the feature extraction for candidate region more accurate. The compound classifier composed by Boosting and SVM is used for the classification of target region, ensuring the robustness of target detection. Through the verification of experiment, the method proposed by this Paper has certain reliability. However, when the temperature is low and the background is complex, the detection effect is poorer than that under other cases. The detection rate is lower than that under other cases and the false alarm rate is higher than that under other cases. This problem could be solved through increasing various target learning of classifier. The method adopted by this Paper has strong robustness for the target detection, which will not affect the operation speed of system with the increase of target quantity in image. It has better detection effect when several targets are cross-blocked.

References

[1] Li X, Cao Z, Yan R, et al. Forward-looking infrared 3D target tracking via combination of

particle filter and SIFT[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2013, 8918(48):08.

[2] Li X, Cao Z, Yan R, et al. Forward-looking infrared 3D target tracking via combination of particle filter and SIFT[C]// Eighth International Symposium on Multispectral Image Processing and Pattern Recognition. International Society for Optics and Photonics, 2013.

[3] Li X, Cao Z, Yan R, et al. Forward-looking infrared 3D target tracking via combination of particle filter and SIFT[C]// Eighth International Symposium on Multispectral Image Processing and Pattern Recognition. International Society for Optics and Photonics, 2013.

[4] Borghys D C J, Acheroy M P J. Long-range automatic detection of small targets in sequences of noisy thermal infrared images[J]. Proceedings of SPIE - The International Society for Optical Engineering, 1996, 2235:4--8.

[5] Nasrabadi N M. Improved target detection algorithm using dualband infrared imagery[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2001, 4379:187-198.