Multisensor long distance target detection using support vector machine

Xuan Yang (杨 烜)[1] and Jihong Pei (裴继红)[2]

[1] *College of Information Engineering, Shenzhen University, Shenzhen 518060*

[2] *Intelligent Information Processing Institute of Shenzhen University, Shenzhen 518060*

Received October 19, 2006

Multisensor image fusion could improve system performances such as detection, tracking, and identification greatly. In this paper, a long distance target detection approach is presented based on multisensor image features fusion. This method extracts two different features from visual and infrared (IR) image sequences respectively to detect regions of motion information content. Temporal change feature is extracted from the visual image sequence using temporal decomposition based on wavelet, which reflects the dynamical content variation at a pixel at any time. And correlation features between local regions are extracted from IR image sequence to distinguish regions with potential moving targets. All these features are merged into a multi-dimensional space and the support vector machine is trained to select regions that have the potential target at each pixel location. The method is robust and feasible to detect long distance targets in clutter background scene.

OCIS codes: 100.5010, 100.2960, 100.2000.

Detection and tracking of long distance targets in multisensor imagery is a challenging problem in targets detection and tracking[1−10]. In contrast to traditional targets, long distance targets have low signal-to-noise ratio (SNR), which results in limited information for detection. Fusion techniques are used to improve performance since it reduces system dependent on a single sensor and increases noise tolerance[1].

Visible sensor and infrared (IR) sensor are capable of revealing different information in a scene. It is necessary to fuse multisensor images to generate a decision image. Only a little work has been done in the multisensor imagery. Dawoud *et al.*[2] proposed a decision fusion algorithm for target tracking in forward-looking infrared (FLIR) image sequence, which allows the fusion of complementary ego-motion compensation and tracking algorithms to estimate the position of the target. Lakshminarayanan *et al.*[1] performed temporal and Behavior-Knowledge-Space fusion to classify civilian targets. Kwon *et al.*[3] extracted four different features from visible, near-infrared, and far-infrared bands multisensors to detect low-contrast targets. Fay *et al.*[4,7] simulated biological models of the human retina and visual cortex to combine imagery from low-light charge coupled device (CCD) camera or a short-wave IR camera, with thermal long-wave IR imagery for visualization. Meltzler *et al.*[5] computed the probability of detection of targets in static IR and visually cluttered scenes using the fuzzy logic approach. Borghys *et al.*[8,9] presented approaches to the long range automatic detection of vehicles, using multisensor image sequence. However, traditional feature-based fusion achieved by weighted features combining is difficult to determine weight coefficients, and decision-based fusion depends on decision results of each single sensor greatly. In this paper, a novel multisensor target detection technique based on feature level fusion using support vector machine (SVM) is proposed.

We are concerned with motion detection in the case of a stationary camera. Temporal multiscale decomposition[10] is used to extract motion feature in visual image sequences. The image sequence is considered as a set of one-dimensional (1D) time varying signals. A sequence of L frames of size $M \times N$ pixels is then a set of $M \times N$ 1D signals with L samples[10]. Temporal decomposition of the 1D temporal signal allows us to reflect the dynamical motion content of the sequence at any pixel. For ideal signal, when there is no motion target on any pixel, the 1D temporal signal can be considered as a constant signal. Otherwise, the temporal signal is a time varying signal. Wavelet transform is adopted to characterize the dynamical motion components because of the good ability of the temporal and frequency localization. The temporal change feature (TCF) is defined as

$$\mathrm{TCF} = \sum_i |D_{k^*}(i)|, \tag{1}$$

where D_{k^*} is the difference information at level k^*, k^* is the best level of wavelet decomposition.

For IR image sequence, because of the difference in thermal radiation between the targets and background, the target regions display similar intensity variations and distributions in successive frames. Firstly, motion detection is performed by computing frame difference between the current frame and its neighboring frame. The frame difference is segmented by a threshold to identify motion regions. For each pixel in the motion regions, local correlation coefficient (LCC) feature is extracted as follows.

It is reasonable to suppose that targets move in a line in several successive frames. Suppose the current pixel is (x_i, y_i), $(\mathrm{d}x, \mathrm{d}y)$ is the motion relative distance during t time. In the same time t, $\Delta = (\mathrm{d}x, \mathrm{d}y)$ can be used to describe motion parameters of targets. Suppose the current frame is f_k, let the predication position in the next frame f_{k+l} of a target be $(x_{\mathrm{p}i}(\Delta), y_{\mathrm{p}i}(\Delta))$, where $x_{\mathrm{p}i}(\Delta) = x_i + \mathrm{d}x$, $y_{\mathrm{p}i}(\Delta) = y_i + \mathrm{d}y$, with the motion parameter Δ. $R_{\mathrm{T}i}$ is the local regions centered at the current pixel (x_i, y_i) in frame f_k. $R_{\mathrm{p}i}$ is the local predication regions centered at $(x_{\mathrm{p}i}(\Delta), y_{\mathrm{p}i}(\Delta))$ in frame f_{k+l}. LCC with motion parameter Δ is defined using correlation coefficient,

$$\begin{aligned}\mathrm{LCC}_i^l(\Delta) = [N \sum_{\substack{(u,v)\in R_{\mathrm{T}i}\\(m,n)\in R_{\mathrm{p}i}}} f_k(u,v) f_{k+1}(m,n) \\ - \sum_{(m,n)\in R_{\mathrm{T}i}} f_k(m,n) \sum_{(m,n)\in R_{\mathrm{p}i}} f_{k+l}(m,n)] \\ \times[N \sum_{(m,n)\in R_{\mathrm{T}i}} f_k^2(m,n) - (\sum_{(m,n)\in R_{\mathrm{T}i}} f_k(m,n))^2]^{-1/2} \\ \times[N \sum_{(m,n)\in R_{\mathrm{p}i}} f_{k+l}^2(m,n) - (\sum_{(m,n)\in R_{\mathrm{p}i}} f_{k+l}(m,n))^2]^{-1/2},\end{aligned} \quad (2)$$

where N is the number of pixels in the local region. LCC represents the similarity between $R_{\mathrm{T}i}$ and $R_{\mathrm{p}i}$. For each pixel in the motion regions, LCC is computed under various motion parameters. And then, the LCC with maximum value is adopted as the local correlation coefficient of the pixel

$$\mathrm{LCC} = \max_{\arg(\Delta)} \{\mathrm{LCC}_i(\Delta)\}. \quad (3)$$

All these extracted features are combined and construct a two-dimensional (2D) feature space. Correspondingly, targets' feature would be more prominent in this 2D feature space. What we need to do is to choose a suitable classifier to partition this space. And it is different from traditional feature-based fusion in that no weighted coefficients are needed, no features are emphasized and no features are suppressed. Moreover, it is different from the typical decision-fusion in that no decision is needed to make in each sensor, and the fusion results would not be affected by decision errors of each sensor.

Considering that the separation boundary of targets and background might be a nonlinear curve, SVM is trained to partition the multi-dimensional feature space, and optimal separating hyperplane is obtained. SVMs are based on the concept of constructing an N-dimensional hyperplane that optimally separates the data into two categories. The hyperplane is oriented so that the margin between the support vectors is maximized, which is particularly suited to handle tasks that a full separation requires a curve. The radial basis function is used as the kernel function.

Sample image sequence is used to train a SVM to determine the hyperplane separating the targets and the background. The inputs to the SVM are the temporal change feature and LCCs of each pixels, including pixels on targets and pixels on background. Perform testing over the experimental image using the trained SVM. If the SVM output 1 at position (i, j), this pixel is decided as target. Otherwise, this pixel is decided as background.

We have applied the proposed method to a data set, which consists of thermal IR sequence of gray frames and visual sequence of color frames. The color frames are transformed to gray frames at first.

Figure 1(a) is one frame of the visual sequence and Fig. 1(b) is one frame of the IR sequence. They are registered already. The sequence of 16 frames is used to extract temporal change feature because the target moves slowly, and different information in level 3 of the wavelet decomposition is computed. Figures 1(d) and (e) are temporal change features of visual sequence and LCC of IR sequence, respectively. Figure 1(f) is the detection result based on feature fusion using SVM. Labeled targets are shown in Fig. 1(c). In the fusion result, noise could be suppressed satisfactorily.

Figure 2 shows two sensor sequences with multi-targets,

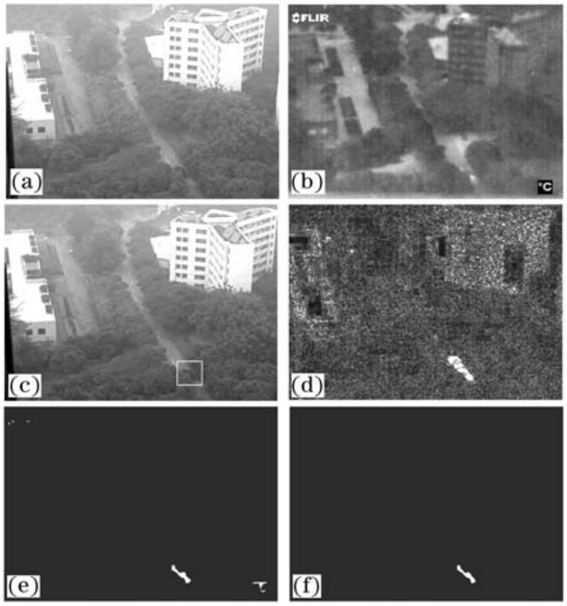

Fig. 1. Detection results of single target. (a) Visual frame; (b) IR frame; (c) labeled target; (d) temporal change feature; (e) local correlation coefficient; (f) detection result.

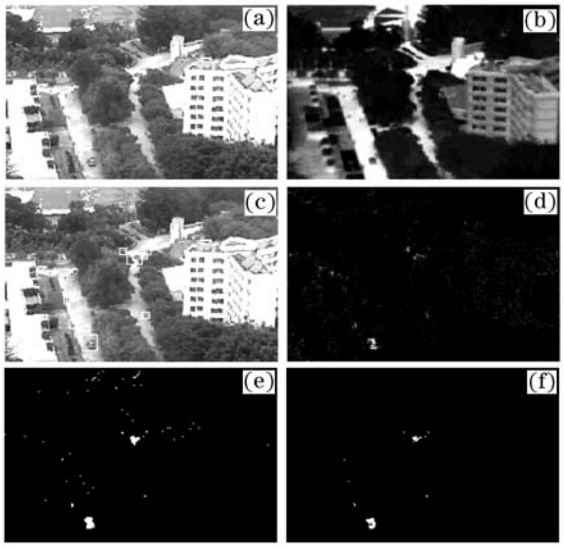

Fig. 2. Detection results of multi-targets. (a) Visual frame; (b) IR frame; (c) labeled targets; (d) temporal change feature; (e) local correlation coefficient; (f) detection result.

such as the passerby and motorcycle on the street besides the moving car. Except the moving car, other targets are very small and dim. The trained SVM of Fig. 1 is used to classify targets and background of Fig. 2. The sequence of 8 frames is used to extract temporal change feature because all these targets are small. Figure 3 shows detection results in another scene. Three targets move across the field of view of the sensors from left to right. In the IR sequences, targets are very dim and small. The sequence of 4 frames is used to extract temporal change feature because all targets are small and move rapidly, and different information in level 2 of the wavelet decomposition is computed. It can be seen that most dim targets are detected in the fused result also.

We have presented a novel feature based fusion method for long distance targets detection in visual and IR sequences. The proposed method extracted two different features from two sensor sequences, respectively. Temporal change feature using wavelet decomposition is computed in visual image sequence, and LCCs of possible areas are computed in IR image sequence. These two features construct a 2D feature space, and SVM is trained in the feature space to obtain the separation boundary of targets and background. Experiments performed on the test sequences show the robustness and feasibility of the proposed moving target detection method.

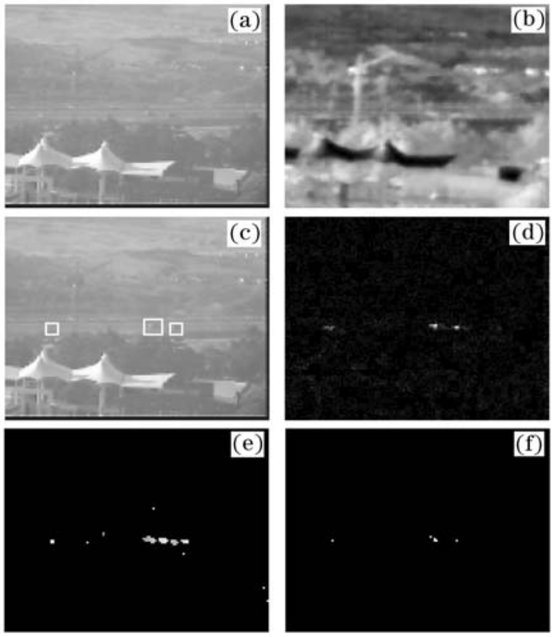

Fig. 3. Detection of multi dim targets. (a) Visual frame; (b) IR frame; (c) labeled targets; (d) temporal change feature; (e) local correlation coefficient; (f) detection result.

The authors wish to thank Dr. A. Schwaighofer for SVM toolbox at http://ida.first.fraunhofer.de/∼anton/software.html. This work was supported by the National Defense Key Laboratory Fundation under Grant No. 51483040105QT5118. X. Yang's e-mail address is pdwxyang@263.net.

References

1. B. Lakshminarayanan and H. Qi, in *Proceedings of AIPR05* 173 (2005).
2. A. Dawoud, M. S. Alam, A. Bal, and C. Loo, Opt. Eng. **44,** 026401 (2005).
3. H. Kwon, S. Z. Der, and N. M. Nasrabadi, Opt. Eng. **41,** 69 (2002).
4. D. A. Fay, A. M. Waxman, M. Aguilar, D. B. Ireland, J. P. Racamato, W. D. Ross, W. W. Streilein, and M. I. Braun, in *Proceedings of the Third International Conference on Information Fusion* TuD3_3 (2000).
5. T. I. Meitzler, H. Singh, L. Arefeh, E. Sohn, and G. R. Gerhart, Opt. Eng. **37,** 10 (1998).
6. X. Chen, Z. Jing, S. Sun, and G. Xiao, Chin. Opt. Lett. **2,** 694 (2004).
7. D. Fay, P. Ilardi, N. Sheldon, D. Grau, R. Biehl, and A. Waxman, in *Proceedings of 7th International Conference on Information Fusion* 499 (2005).
8. D. Borghys, P. Verlinde, C. Perneel, and M. Acheroy, Opt. Eng. **37,** 477 (1998).
9. D. Borghys, P. Verlinde, C. Perneel and M. Acheroy, Proc. SPIE **3068,** 569 (1997).
10. J. M. Letang, V. Rebuffel, and P. Bouthemy, in *Proceedings of 11th IAPR International Conference on Pattern Recognition* **1,** 65 (1992).