MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Automatic Target Recognition by Infrared and Visible Image Matching

Kai-Sheng Cheng and Huei-Yung Lin
Department of Electrical Engineering
Advanced Institute of Manufacturing with High-Tech Innovation
National Chung Cheng University
168 University Rd., Chiayi 621, Taiwan

Abstract

Automatic target recognition based on the long wave infrared (LIR) and visible-light spectrum (VIS) image matching is a very challenging problem. It is due to the fact that the images between LIR and VIS have lots of different textures. This difficulty is inherent in the thermal radiation imaging affected by one of the principal mechanisms so called heat transfer. In this paper, a novel algorithm is presented for object recognition between the LIR and VIS images under various conditions. It is assumed that the visible light images of the target are available a priori, and the newly acquired infrared images are used to perform the target recognition task. The LIR and VIS images are first initialized with edge detection and binary template matching, followed by a local fuzzy threshold to identify the high similarity objects. Our method has has low computational requirements and can be implemented on a real-time system. Several experiments are carried out using the real scene images with various test objects.

1 Introduction

Automatic target recognition (ATR) is a technique to recognize objects or targets by comparing the information stored in the database and acquired from the real-time sensing. It is commonly used to enhance the missile guidance in the type of fire-and-forget applications. In such missile systems, GPS is used to automatically guide the missile to the target in a long-range distance, and ATR then takes over to increase the accuracy in the short-range. The objective is to improve the missile usage efficiency and lower the military expenditure. In general, a missile system adopts an infrared sensor, called forward-looking infrared (FLIR), to avoid low-quality imaging in the battlefield such as in the rainy or foggy weather conditions. In this work, the main target needs to be recognized within several of static targets by their heat emission, along with multiple environmental interferences. However, the thermal radiation follows the equation $\alpha + \rho + \gamma = 1$ with some mechanisms such as heater transfer, convection and conduction, where α, ρ and γ represent components of absorption, reflection and transmission, respectively. Under this image formation process, the appearance of LIR images is fuzzy and represents a totally different concept from the VIS images. Thus, the development of an infrared-visible object recognition system is a very challenging task.

Most approaches for infrared object detection adopt binary template matching to detect targets in the image or video sequence. Furthermore, the real-time performance is generally required for a missile system. In the previous work, the robust template matching techniques are used in [1, 2]. Lewis [1] adopted a well-known cross-correlation method in the transform domain and pre-computed a table which contains the integral of the image to speed up the system execution. Omachi *et al.* [2] and Luigi *et al.* [3] proposed the algorithms that used NCC bounded partial correlation template matching and could perform efficiently with the width and height of the template image different from the partial image. There also exist other methods such as extracting areas of interest (blobs), *PNSF-m*, and orientation map [4, 5, 6] . But most researches in the literature about template matching address the VIS-VIS problems, not the LIR-VIS related issues. Consequently, these methods cannot be applied to the LIR-VIS recognition system.

In this paper, we propose a technique for automatic target recognition by infrared and visible image matching. We use VIS templates to match the objects in the LIR image through the fuzzy cross-correlation coefficient with a proper threshold. In our experiments, it is assumed that some of the VIS images are collected for matching the specific objects in the LIR images. A video sequence is captured when the camera is moving at the distance of 20 meters to 15 meters away from the target objects. Except for the simple test objects, our method is also applied to the real world scenes with vehicles and buildings. To the best of our knowledge, there are very few works attempting to recognize a specific object between NIR and VIS. In this work, the proposed technique for solving the LIR-VIS recognition problem is easy to implement with low computation cost.

2 Our Approach

In the past few years, several adaptive algorithms have been proposed for object recognition, and the methods can be divided into two categories, content-based and feature-based [7]. Some well-known feature-based methods such as SIFT and SURF descriptors in FLIR images are tested in the early work, but it shows poor performance because the feature points in LIR and VIS are different. Since there are very few works that could adapt in the FLIR images, we need to extract stable features which are able to represent the same object in the images obtained from different sensing techniques under various conditions, e.g. the different target size and type of objects, and the complexity of the background scene. Another major problem is that the FLIR image quality can affect the performance of ATR systems. To ensure an ATR tech-

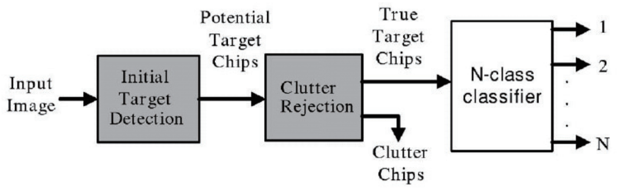


Figure 1. The concept of our ATR system.

nique can adapt in different conditions, the proposed method includes the following steps:

(1) For a pair of input VIS/LIR images, extract the edge-based features.

(2) A Gaussian-like pyramid of the image is constructed with ten spatial scales.

(3) Use local fuzzy threshold for an ROI extraction and binary template matching.

In step 1, a VIS image and an FLIR video sequence are sent to the recognition system, after noise reduction, the edge-based features are then extracted. Since the LIR imaging is formulated through thermal radiation, we found the boundary in each object in the LIR and VIS images. This is an important process in our algorithm for initial object recognition. Another purpose to extract the edge-based features is that we want to perform a center-surround difference operation. The center-surround difference is used to find the salient regions of an input image. A pixel with the intensity value smaller than a threshold will be removed to derive the binary template and LIR binary image.

2.1 Binary template generation

Due to the object templates are VIS image, it's easier to collect all possible outward appearance. We assume that each VIS template already contains the approaching path property For each given input VIS image $f(x, y)$, a region template $T(x, y)$ can be found by binary template generation, and it contains the object in the region. Notice that VIS image f can be taken by any camera. We first remove the image noise and extract the edge features by a convolution process. For each LIR image $I(x, y)$ with the resolution of 320×240, a function $\omega(L_{\lambda_1}, L_{\lambda_2})$ which denotes the infrared intensity energy is then computed. The function $\omega(L_{\lambda_1}, L_{\lambda_2})$ contains an upper bound temperature emissivity L_{λ_1} and a lower bound temperature emissivity L_{λ_2}, and is given by

$$\omega = \frac{L_{\lambda_1}}{L_{\lambda_2}} \tag{1}$$

The emissivity function ratio is directly proportional to the temperature, and matched to a single channel 8-bit image. We then convolve the VIS and LIR images using the edge kernel $g(x)$ to extract edge features, i.e.,

$$E_{T(x,y)} = T(x, y) * g(x) \tag{2}$$

$$B_{E_T} = \begin{cases} 0, & \text{if} \quad E_{T(x,y)} < \delta \\ 255, & \text{if} \quad E_{T(x,y)} > \delta \end{cases} \tag{3}$$

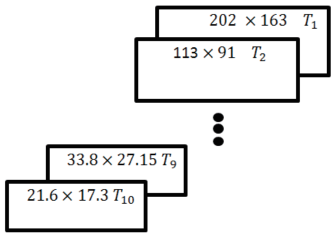


Figure 2. Gaussian pyramid with scale coefficient.

where $E_{T(x,y)}$ is the edge template convolution by the gradient kernel $g(x)$, and the binary edge template is constructed by a filter with a threshold δ.

We then use a Gaussian-like pyramid to reduce the size of each template (see Figure 2). This setting is due to the physical rule of the missile path in our experiments. The image scale change is similar to a Gaussian pyramid. A scale coefficient is added to the equation to let it more adaptable to the real scene conditions.

A formal Gaussian with a scale coefficient is defined as follows:

$$g(x, y; t) = \frac{1}{2\pi t} e^{\frac{-(x^2+y^2)}{2t}} S(d) \tag{4}$$

where $S(d)$ denotes the scale coefficient under different distance d for a VIS binary template with the size of 225×181. The distance (in meter) in the experiments is $d = 2, 4, 6, \cdots, 20$, and the perfect matching results with the most suitable detection region are $S(d) = 90\%, 50\%, 35\%, 25\%, 20\%, 17\%, 15\%, 12\%, 10.5\%, 9.6\%$, respectively.

2.2 Binary template matching

In [1, 8], a template matching strategy is based on cross-correlation, and affinities are computed by M = #(T AND I)/#(T XOR I)+1, where T and I represent the pairs of VIS binary templates (B_{E_T}) and the binary edge LIR images in the current frame of an FLIR video sequence. In the two-dimensional case, given the intensity of the input image $I(i, j)$ with the ith column and the jth row, the NCC equation can be written as

$$NCC(x, y) = \frac{C(x, y)}{||I(x, y)||_2 \cdot ||T||_2} \tag{5}$$

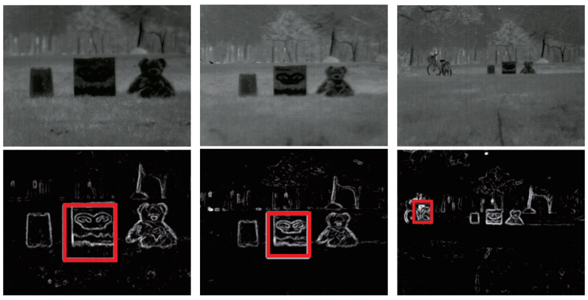
where

$$C(x, y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(x+i, y+j) \cdot T(i, j) \tag{6}$$

$$||I(x, y)||_2 = \sqrt{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(x+i, y+j)^2} \tag{7}$$

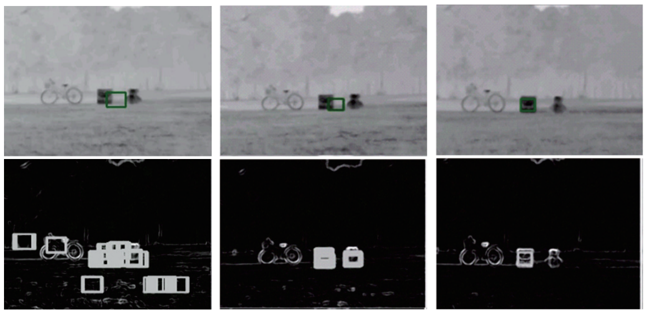
$$||T||_2 = \sqrt{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} T(x+i, y+j)^2} \tag{8}$$

Here the goal is to get the production of the NCC similarity rate $S = \{S_1, S_2, \ldots, S_n\}$, where n is the number of box-filtering count. We can reduce the computation cost by pre-constructing the integral images.



(a) (b) (c)

Figure 3. NCC method performed at difference distances, the target to be recognized is in the middle. The binary template size is 225 × 181. (a) $d = 6$m, $S(d) = 35\%$. (b) $d = 12$m, $S(d) = 17\%$. (c) shows an incorrect recognition result. $d = 18$m, $S(d) = 10.5\%$



(a) (b) (c)

Figure 4. Performance comparison of different algorithms at $d = 18$m, $S(d) = 10.5\%$. (a) NCC fuzzy threshold without coefficient. (b) SAD coefficient with fuzzy threshold. (c) NCC coefficient with fuzzy threshold.

Usually the NCC_{max} will be taken for the recognition results. We also implemented the Sum of Square Difference (SSD) metric to compare with NCC metric in VIS-LIR object recognition. The SSD is given by

$$SSD(x, y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(x + i, y + j) - T(i, j))^2 \quad (9)$$

where the best result in normal SSD is denoted by SSD_{min}. An experiment using the NCC method have been tested, and the results are shown in Figure 3. The target object is found correctly in Figures 3(a) and 3(b), but a mismatch occurs in Figure 3(c). This is mainly due to the large distance between the objects and the infrared sensor, which makes the objects look similar in such a small scale. In this situation, two objects with the same shape but different content inside, the results with the similarity computed by NCC could be incorrect.

2.3 NCC with coefficient and local fuzzy threshold

To deal with the problem in Section 2.2, we consider the region for object recognition by using the coefficient with NCC and a local fuzzy threshold [9]. For an NCC product LIR image $I(x+i, y+j)$ and the binary template $T(i, j)$, we rewrite the equation as

$$NCC_{coe} = \frac{C'(x, y)}{||I'(x, y)||_2||T'||_2} \quad (10)$$

$$C'(x, y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I'(x + i, y + j) \cdot T'(i, j) \quad (11)$$

$$I' = \sum_{i,j=0}^{M-1,N-1} I(x+i, y+j) - \frac{1}{WH} \sum_{i',j'=0}^{M-1,N-1} I(x+i', y+j') \quad (12)$$

$$T' = \sum_{i,j=0}^{M-1,N-1} T(i, j) - \frac{1}{WH} \sum_{i',j'=0}^{M-1,N-1} T(i', j') \quad (13)$$

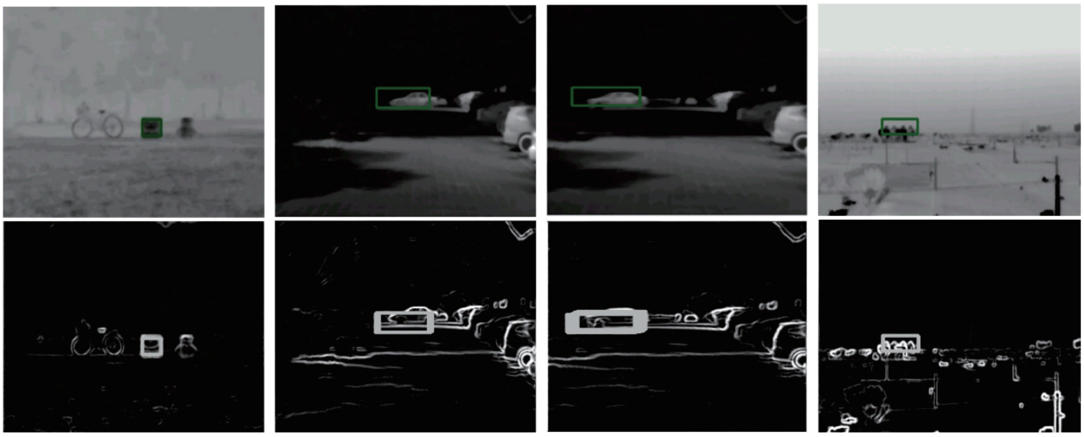
Notice that the proposed modification adds an average term in Eq. (10), and it considers the whole object region for processing. To minimize the heat transfer and improve the previous method, we set a fuzzy threshold to select the NCC similarity rate which is greater than the threshold. Given a fuzzy threshold F_δ, the output ROI is the one with $\gamma(\gamma_1, \gamma_2, \cdots, \gamma_n) > F_\delta$. In final step, we select the result from all ROIs with the highest similarity rate.

As shown in Figure 4(a), the ROIs are identified if the similarity rate is higher than a fuzzy threshold, i.e., $\gamma > F_\delta = NCC(x, y) = 0.96$, for the NCC method without a coefficient term. However, it contains too many false-positive regions. In Figure 4(b), we test the SAD coefficient method with a fuzzy threshold and it shows that the result is better than Figure 4(a). Figure 4(c) shows the best result obtained using NCC coefficient with a fuzzy threshold. The target ROI is identified perfectly and it resolves the false detection problem as shown in Figure 3(c). In addition, to search an object within a $w \times h$ LIR image using an $i \times j$ template, the time complexity is $O(whn)$, where n is the number of all ROIs. To construct the binary template, the time complexity is given by $O(i^2j^2)$.

3 Experimental Results

In our experiments, several scenes containing sample objects and real world objects have been tested. First, three test objects were placed in a field, and the VIS template images were collected as a database for online target recognition using the FLIR images. The emissivity of the infrared camera was set as 0.9ϵ. We simulated three different missile paths in a scale of $1 : 100$, and the size of the main object to be recognized was about 57×40 cm^2. In a real missile system, the target to be recognized is at about 2 km away, which corresponds to 20 meters in our experiment. We recorded three sets of FLIR images from 20 meters to 2 meters away from the target with different approaching angles, and one video sequence was recorded from 20 meters to 15 meters away from the target. Each set of FLIR images contains 10 frames, and the video sequence contains 150 frames.

Figure 6 shows the suitable template sizes for target recognition in the approaching paths of three different angles, 260°, 270°, 280° respectively. The matching range represents the template scale which is suitable for the target recognition task. Each



(a) (b) (c) (d)

Figure 5. The object recognition results. a) A result from the video sequence experiment. b) The recognition of a vehicle from the side view. c) The vehicle is recognized with a different direction. d) The target recognition result of a building.

point on the line is the correct result that we have tested over 300 times. The precision rates for the three different angles are given by $(angle, precision) = (260°, 90\%), (270°, 100\%), (280°, 90\%)$. Another experiment is carried out using an LIR video sequence with two different emissivity. The accuracy is 100% and the result is shown in Figure 5(a). The template size is determined by the Gaussian-like pyramid with a scale coefficient. In this experiment, the scale in the distance between 20 meters to 15 meters is more stable than other distance ranges. That is, we are able to use one template scale to recognize the objects in the range of 20 meters to 15 meters.

In the last experiment, the real world scene is used to test our algorithm. The target recognition results are shown in Figures 5(b), 5(c) and 5(d). In Figures 5(b) and 5(c), the vehicles with different viewing directions are recognized successfully using only one VIS template. Figure 5(d) demonstrates the successful recognition of a building at approximately 1.2 kilometers from the camera.

4 Conclusions

This paper presents an LIR-VIS object recognition method based on template matching and NCC with fuzzy coefficient. We use the collected VIS images to recognize the target object presented in the LIR images. The experimental results demonstrate that we can recognize the real world objects at 1.2 kilometers away from the camera. Furthermore, the proposed technique is able to deal with the approaching angle different from the VIS image acquisition positions. In the future work, we will improve the method to recognize the objects under possible rotation and distortion, and track the target till a close distance.

References

[1] J. P. Lewis: "Fast Template Matching" *Vision Interface*, pp. 120–123, 1995.

[2] S. Omachi and M. Omachi: "Fast Template Matching With Polynomials" *IEEE Trans. on Image Processing*, Vol. 16, No. 8, pp. 2139–2149, 2007.

[3] L. D. Stefano and S. Mattoccia: "Fast template matching using bounded partial correlation" *Machine Vision and Applications*, Vol. 13, pp. 213–221, 2003.

[4] L. Cole and D. Austin: "Visual Object Recognition using Template Matching" *ARAA*, December 6-8, 2004, Canberra, Australia.

[5] J. S, Bae and T. L. Song: "Image Tracking Algorithm using Template Matching and PSNF-m" *IJCAS*, Vol. 6, No. 3, pp. 413–423, 2008.

[6] D Nguyen, W. Li, and P. Ogunbona: "An improved template matching method for object detection" *ACCV*, Sep. 23-27, 2009, Xi'an, China.

[7] H. Y. Lin, C. Y. Hsu, and Y. Y. Chiang: "Detecting Structured Image Region Using Local Features and Clustering Analysis," *MVA*, pp. 53-56, Kyoto, Japan, May 20-23, 2013.

[8] P. Wang, J. Sun, L. Li, and Y. Chen: "A Local Fuzzy Thresholding based Metric for FLIR Image Quality Assessment" *JCIT*, Vol. 7, No. 3, pp. 599-607, 2012.

[9] S. G. Sun: "Target detection using local fuzzy thresholding and binary template matching in forward-looking infrared images" *SPIE*, Vol. 46(3) 03642, 2007.

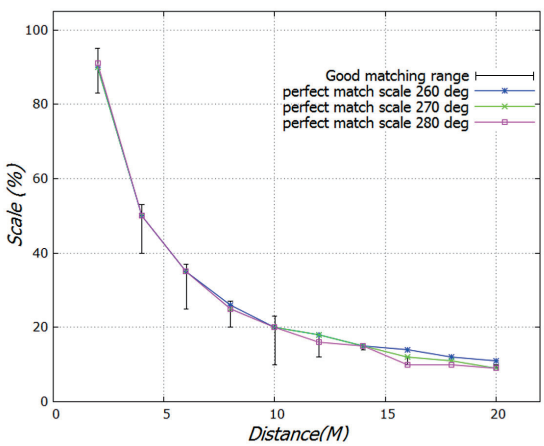


Figure 6. The matching results with different distance and fuzzy scale.