Infrared Physics & Technology 68 (2015) 75–83

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Robust object detection based on local similar structure statistical matching

Feiyang Luo, Jing Han, Wei Qi, Yi Zhang, Lianfa Bai *

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China

HIGHLIGHTS

- A composite template set of one object is designed to detect the similar object.
- A simplified method is proposed to reduce the composite template set.
- A matching method of LSSSM is proposed to obtain similarity image.
- Extracting object positions according to the similarity image.

ARTICLE INFO

Article history:
Received 24 April 2014
Available online 11 November 2014

Keywords:
Object detection
Local similar structure statistical matching
Composite template set
Non-maxima suppression

ABSTRACT

We present a robust object detection method to detect generic objects with incompact, complex and changeable shapes without training. First, we build a composite template set, which contains changeful shapes, scales and viewpoints of an interested object class, extract the local structure features from the composite template set and simplify them to construct a non-similar local structure feature set of the object class. Then, we propose a matching method of local similar structure statistical matching (LSSSM) to obtain the similarity image from a test image to the local structure feature set. Finally, we use the method of non-maxima suppression in the similarity image to extract the object position and mark the object in the test image. The experimental results demonstrate that our approach performs effectively on the face and infrared human body detection.

1. Introduction

Object recognition and detection is an important area of computer artificial intelligence. Now, popular object recognition methods, such as probabilistic constellation [1] or parts-and-shape models [2], are based on learning-based classifiers [3], which require an intensive learning/training phase of the classifier parameters and thus are called parametric methods. However, these approaches generally require a lot of training samples in the learning process, and the training process may result in overfitting of training parameters. Besides, they are very slow. In order to avoid the learning and long training process from a large number of samples, no-training method of image recognition is gradually presented, the recognition task with only one sample (or a few) has received increasing attention for many applications [4–7].

The local regression kernel was initially used for image processing and reconstruction [8]. Local regression kernel descriptors can essentially measure the local similarity of a pixel to its neighbors both geometrically and photometrically. It well captures the local structure features of images. To some extent, it can eliminate the influence of noise [9]. So Seo proposed the method of Locally Regression Kernels to detection general object [10], such as face verification [11] and human action [12] detection, by using a single sample image of interested object to find the similar matches in the test image. However, for the object with complex and incompact overall structure, such as human body, it is impossible to predict the posture of the object when it appears in the test image, and the sample picture we choose may be very different from the object in the test image (Fig. 1). As a result, the similarity between sample and the test image is too low to accurately detection the true objects.

In order to solve the problem of above methods, we improve the method in [10] and present a robust object detection method to detect generic objects with incompact, complex and changeable shapes. First, we propose a composite template set, which contains changeful shapes, scales, viewpoints of objects. We build a local structure feature set by use the locally adaptive regression kernel (LARK) descriptor to describe local structure features in the composite template set, and propose a method to simplify the local structure feature set. Then we present a local similar structure

* Corresponding author.
E-mail address: mrblf@163.com (L. Bai).

http://dx.doi.org/10.1016/j.infrared.2014.10.021

statistical matching (LSSSM) method to achieve effective match in object detection without training. The most significant differences between our method and the method in [10] are that our method is robust to the object with complex structures and needs no iteration. So the proposed method can be used to detect objects with complex structure characteristics successfully. And its high performance is demonstrated in infrared human detection.

2. Locally adaptive regression kernel

In order to make the concepts clearer, we first briefly describe the locally adaptive regression kernel (LARK) [11] in this section. LARK captures local geometric structures effectively and efficiently by taking advantage of self-similarity based on gradients. Because LARK vectors can reflect the graphic shapes and gray value changes of local structure features, it can describe the local structure and has good robustness. We define the LARK as a self-similarity between a center and its surroundings as follows:

$$K(C_l, \Delta X_l) = \exp(-ds^2) = \exp\{-\Delta X_l^{\mathrm{T}} C \Delta X_l\}, \tag{1}$$

where $l \in [1, \cdots, P^2]$, P^2 is the total number of pixels in a local analysis window around a center position at the pixel of interested X, $S = \{x_1, x_2, z(x_1, x_2)\}$, in which $z(x_1, x_2)$ is the gray value of coordinate (x_1, x_2). C_l is based on gradients (z_{x_1}, z_{x_2}) in one pixel, and it is an 2×2 covariance matrix computed from the patch $\Omega_l(m \times m)$ centered at position l, as follows:

$$C_l = \sum_{m \in \Omega_l} \begin{bmatrix} z_{x_1}^2(m) & z_{x_1}(m) z_{x_2}(m) \\ z_{x_1}(m) z_{x_1}(m) & z_{x_2}^2(m) \end{bmatrix}. \tag{2}$$

A particular example of computing LARK of size 5×5 is shown in Fig. 2, X_{13} is the center pixel in the gray patch, ΔX_{13} is $[0, 0]^{\mathrm{T}}$, C_{13} is a 2×2 covariance matrix computed from the gray patch Ω_{13} of size 5×5 centered at X_{13}.

LARK in (1) is normalized to a unit vector to be more robust to illumination changes. The LARK of size $P \times P$ patch is normalized as follows:

$$k_i^l = K_i^l \Bigg/ \sum_{l=1}^{p^2} K_i^l \in R^{P^2 \times 1}, \quad i = 1, \ldots, N, \; l = 1, \ldots, P^2, \tag{3}$$

where N represents the total number of pixels in the image. So a column vector of LARK of one patch centered at pixel i as follows:

$$w^i = [k_i^1, k_i^2, \ldots, k_i^{P^2}]^{\mathrm{T}}. \tag{4}$$

Then, we compute the LARK vector of all pixels according to the column order in image to get the LARK matrix of the image as follows:

$$W = [w^1, \ldots, w^i, \ldots, w^N] \in R^{P^2 \times N}. \tag{5}$$

The LARK matrix W is also called local structure feature matrix, each column vector in W describes a local structure in the image.

3. The detail of LSSSM

Different from using one sample image of object to find the similar object in the test image, we study a kind of composite template set to solve the problem of shape and viewpoint variations, and

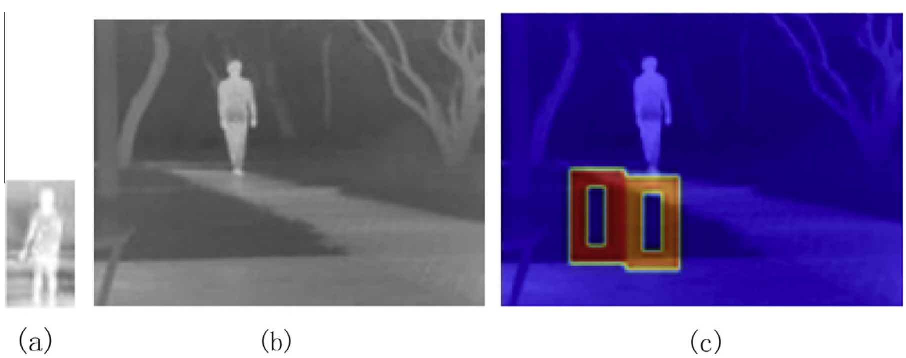

Fig. 1. By using one sample image to detection human body, (a) a sample image, (b) the test image, and (c) detection result.

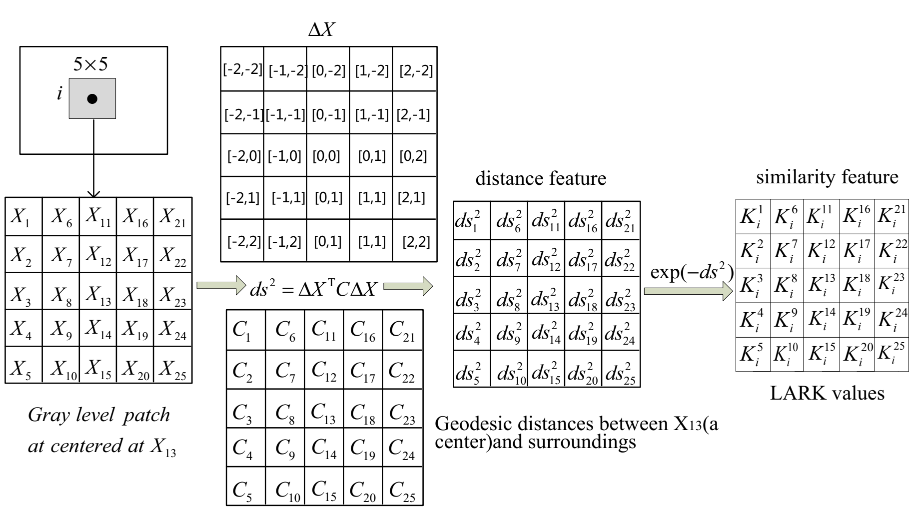

Fig. 2. The way to compute LARK (5×5) values centered at X_{13}.