Infrared Physics & Technology 67 (2014) 91–97

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Good match exploration for thermal infrared face recognition based on YWF-SIFT with multi-scale fusion

Junfeng Bai, Yong Ma *, Jing Li, Hao Li, Yu Fang, Rui Wang, Hongyuan Wang

Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

HIGHLIGHTS

- A novel feature matching scheme for infrared face recognition is proposed.
- The method fuses the infrared frame with an auxiliary visual frame to enrich the information of an infrared face.
- The fusion algorithm is based on the multi-scale discrete wavelet transform.
- The matching scheme is based on YWF-SIFT which can efficiently handle mismatches.
- The proposed method can largely enhance the feature matching performance compared to YWF-SIFT.

ARTICLE INFO

Article history:
Received 14 April 2014
Available online 17 July 2014

Keywords:
Thermal infrared image
Feature matching
Image fusion
Multi-scale fusion
Face recognition

ABSTRACT

Stable local feature detection is a critical prerequisite in the problem of infrared (IR) face recognition. Recently, Scale Invariant Feature Transform (SIFT) is introduced for feature detection in an infrared face frame, which is achieved by applying a simple and effective averaging window with SIFT termed as Y-styled Window Filter (YWF). However, the thermal IR face frame has an intrinsic characteristic such as lack of feature points (keypoints); therefore, the performance of the YWF-SIFT method will be inevitably influenced when it was used for IR face recognition. In this paper, we propose a novel method combining multi-scale fusion with YWF-SIFT to explore more good feature matches. The multi-scale fusion is performed on a thermal IR frame and a corresponding auxiliary visual frame generated from an off-the-shelf low-cost visual camera. The fused image is more informative, and typically contains much more stable features. Besides, the use of YWF-SIFT method enables us to establish feature correspondences more accurately. Quantitative experimental results demonstrate that our algorithm is able to significantly improve the quantity of feature points by approximately 38%. As a result, the performance of YWF-SIFT with multi-scale fusion is enhanced about 12% in infrared human face recognition.

1. Introduction

In spite of the thirty years research of machine recognition of human faces in the visible spectrum, two severe problems are still unsolved, i.e., the illumination variation and the face disguise [1]. Recent studies reveal that the human face recognition in the infrared (IR) spectrum may work properly in these two scenarios. This can be attributed to the following two facts. On the one hand, the thermal IR images are generated without taking account of the illumination intensity. On the other hand, the disguise can be distinguished since a thermal pattern of a face is derived primarily from the superficial blood vessels under the skin [2]. The representative infrared methods include elemental shape matching, eigenface, metrics matching, template matching, symmetry waveforms, face codes, as well as SWF-SIFT [3–6]. Among them, the SWF-SIFT approach which is based on the scale invariant feature transform (SIFT) method is more suitable for the infrared human face recognition [7–9]. Therefore, the key step of thermal IR face recognition is to establish more accurate feature matches between the input images, which is the very goal in this paper.

The SWF-SIFT method is able to handle the facial rotation and occlusion problems such as glasses wearing [6]. However, the method usually leads to a number of mismatched feature points and hence degrades the performance. The mismatches could be eliminated by a postprocessing such as mismatch removal [10–14], but this step often relies on a global geometrical constraint. Bai et al. [9] found that the mismatches were typically caused by the feature points with similar textures around them, and hence they proposed a YWF-SIFT method to address this issue. The features extracted by this method are more frequently spread over the images and more stable, which works well in the visible

* Corresponding author.
E-mail address: mayong@hust.edu.cn (Y. Ma).

http://dx.doi.org/10.1016/j.infrared.2014.06.010
1350-4495/

spectrum. However, the performance of YWF-SIFT degrades severely in case of the IR frames. This can be attributed to the decrease of the number of feature points.

With the study of fusion technology, we recognized that one possible solution is to operate on a fused image of the original IR frame and a corresponding auxiliary visual frame generated from an off-the-shelf low-cost visual camera. Clearly, the fused image is more informative than the original thermal IR frame, and hence it is possible to generate more features for the following recognition process [15].

In general, there are three types of techniques for visual and IR frames fusion [16,17], i.e, feature level fusion [18], decision level fusion [19], and pixel level fusion [20]. Among these techniques, fusion at the feature level is expected to provide the best recognition result, since the feature set typically contains the richest information about the raw biometric data. In this paper, we adopt a feature level fusion technique [21], which performs a multi-scale fusion in the discrete wavelet transform domain, to deal with the problem of low number of feature points in IR frames, and hence establish more accurate feature matches.

The remainder of this paper is organized as follows. In Section 2, we describe our method for feature matching in IR face recognition in details, which is composed of four major steps such as registration, multi-scale fusion, feature matching, as well as performance evaluation. In Section 3, we present the experimental comparisons of YWF-SIFT and YWF-SIFT with multi-scale fusion, and also discuss the advantages and disadvantages of our method. In Section 4, we make some concluding remarks.

2. Feature matching based on YWF-SIFT with multi-scale fusion

Fig. 1 illustrates the flow chart of our proposed feature matching scheme in thermal IR face recognition. Generally, the IR frame and the visual frame acquired by two different sensors are not perfectly aligned [15]. If the two images are aligned by hardware, the registration is then optimal. Otherwise, a registration algorithm is necessary to align the two frames ahead. After the registration step, the aligned two images are fused to generate a high-resolution image, followed by the feature matching between the fused images based on YWF-SIFT. Finally, we evaluate the matching performance.

2.1. Registration

Registration of multi-sensor images can be implemented by either hardware or software. For the consideration of reasonable cost, the software-based registration is preferred in this paper, where an off-the-shelf low cost visual camera can be utilized and no additional hardware is needed.

In the scheme of Fig. 1, the original two input frames are from different regions of the electromagnetic spectrum. Most of the registration criterion are area-based, and hence cannot be applied successfully to the multi-sensor image registration. There are in general two methods to align a visual frame with an IR frame. The first approach is based on directional energy maps and is suitable for alignment of man-made structures such as airport runways or buildings [22]. The second one is based on the Canny edge detectors, which is more suitable for the face registration [21]. Here we choose the second scenario.

The alignment of Canny edges is implemented as follows [21]. First, the Canny edges are extracted from the two frames as feature maps. Then the registration is performed based on the extracted edges. Assume that X and Y are the pixel positions of the Canny edges in the long-wave infrared frame and short-wave visual frame, respectively; $S(X)$ and $S(Y)$ are their corresponding 1D attribute vectors which can be intensity as well as color information or local shape descriptors [23], here we use the intensity for efficiency. Thus the binary feature maps of the two frames can be described as point sets $L = \{X, S(X)\}$ and $V = \{Y, S(Y)\}$.

The similarity of the two feature maps can be described by the following Gaussian function $E(T)$ exerted by one point set over the other:

$$E(T) = \sum_{X \in L, Y \in V} \exp\left\{-d^2(X, T[Y])/\sigma^2 - [S(X) - S(T[Y])]^2\right\}, \tag{1}$$

where $d(X, Y)$ denotes the Euclidean distance between pixels X and Y, σ^2 controls the decay with Euclidean distance, and $T[\cdot]$ represents an affine transformation for the registration of the two point sets. In our experiment, we choose the typical value for σ as in [21], i.e., $\sigma = 1$. Let the position Y in the visual frame be (x, y) before the affine transformation and (x', y') after the transformation, we have

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{2}$$

where a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23} are the transformation coefficients. Therefore, the first term $-d^2(X, T[Y])/\sigma^2$ in Eq. (1) defines the spatial correlation of the IR frame and the transformed visual frame. While the second term $-[S(X) - S(T[Y])]^2$ defines the feature map correlation of the two frames. The objective function $E(T)$ maximizes both the overlap and the local attribute similarity of the two frames. Moreover, to avoid spurious results which are mainly caused by local maxima, a regularizing term is added to the objective function and hence it becomes:

$$E'(T) = E(T) + \lambda \sum_{Y \in V} d^2(Y, T[Y]). \tag{3}$$

where λ is a regularization parameter. By minimizing $E'(T)$ with a standard quasi-Newton algorithm, the parameters of the affine transformation in Eq. (2) can be computed [21]. Once we obtain the transformation, we then use it to align the IR frame and the corresponding visual frame.

2.2. Multi-scale fusion

The discrete wavelet transform (DWT) technique is applied to data fusion for visible and thermal IR image pairs, which leads to a multi-scale fusion scheme [21]. First, the wavelet coefficients of the infrared frame ($W_\phi^{\text{thermal}}(m, n)$ and $W_\psi^{\text{thermal}}(m, n)$) and the visual frame ($W_\phi^{\text{visible}}(m, n)$ and $W_\psi^{\text{visible}}(m, n)$) are calculated, where the subscript ϕ denotes the approximation coefficients and the ψ denotes the detail coefficients. Second, the weighted sums of the two sets of coefficients $W_\phi(m, n)$ and $W_\psi(m, n)$ are obtained, as stated in the following two equations [16,21]:

$$W_\phi(m, n) = \alpha_1 W_\phi^{\text{visible}}(m, n) + \beta_1 W_\phi^{\text{thermal}}(m, n), \tag{4}$$

$$W_\psi(m, n) = \alpha_2 W_\psi^{\text{visible}}(m, n) + \beta_2 W_\psi^{\text{thermal}}(m, n), \tag{5}$$

where m and n denote the pixel coordinates; $\alpha_1, \beta_1, \alpha_2$ and β_2 represent the weighted factors of the coefficients. In our experiment, we set $\alpha_1 = \beta_1 = 0.5$ and $\alpha_2 = \beta_2 = 1$ as in [21]. Noted that in this step, the two images are co-registered and of the same size.

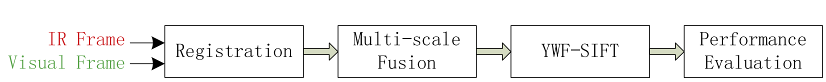

Fig. 1. Flow chart of the face recognition method proposed in this paper.