Infrared Physics & Technology 67 (2014) 111–115

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Good match exploration for infrared face recognition

Changcai Yang [a,b], Huabing Zhou [c], Sheng Sun [d], Renfeng Liu [a,b,*], Ji Zhao [b], Jiayi Ma [e]

[a] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[b] School of Automation, Science and Technology on Multi-Spectral Information Processing Laboratory, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430073, China
[d] School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China
[e] Electronic Information School, Wuhan University, Wuhan 430072, China

HIGHLIGHTS

- SIFT-GSI is proposed to establish accurate SIFT feature correspondences for IR face recognition.
- A smooth spatial mapping function for the underlying correct matches is estimated by using GSI.
- The proposed method can establish accurate correspondences without hurting the correct matches.

ARTICLE INFO

Article history:
Received 10 April 2014
Available online 17 July 2014

Keywords:
Infrared (IR) face image
Feature matching
Spatial mapping function
Thin-plate spline (TPS)
Global structure information (GSI)

ABSTRACT

Establishing good feature correspondence is a critical prerequisite and a challenging task for infrared (IR) face recognition. Recent studies revealed that the scale invariant feature transform (SIFT) descriptor outperforms other local descriptors for feature matching. However, it only uses local appearance information for matching, and hence inevitably leads to a number of false matches. To address this issue, this paper explores global structure information (GSI) among SIFT correspondences, and proposes a new method SIFT-GSI for good match exploration. This is achieved by fitting a smooth mapping function for the underlying correct matches, which involves softassign and deterministic annealing. Quantitative comparisons with state-of-the-art methods on a publicly available IR human face database demonstrate that SIFT-GSI significantly outperforms other methods for feature matching, and hence it is able to improve the reliability of IR face recognition systems.

1. Introduction

Human face recognition is of particular interest in the past 30 years due to increasing demands for security in law enforcement and commercial applications, such as mugshot identification, gateways to limited access areas and surveillance of crowd behavior [1,2]. Current face recognition systems typically work well in case of visual modality and in a controlled environment. However, in uncontrolled operating conditions such as non-uniform illumination, low lighting, makeup, as well as disguise, they often perform poorly because of great errors during extracting reliable features.

Recently, face recognition using thermal infrared (IR) imaging sensors has become an area of growing interest [1–4]. These studies reveal that face recognition in IR spectrum may work much better than in visible spectrum in case of uncontrolled environment. This can be attributed to that the IR images are generated without taking account of the illumination intensity and a thermal pattern of a face is derived primarily from the superficial blood vessels under the skin. The general procedure for face recognition contains three major steps: face detection, feature extraction, and recognition. The first step aims to segment out face-like objects from cluttered scenes. And then in the second step face images are usually represented in terms of feature vectors in lower dimensional feature space for recognition. Finally, face recognition tasks include both identification and verification. In this paper, we focus on the second step, exploring good feature correspondences for IR face recognition.

Due to the ability of extracting distinctive feature points that are invariant to location, rotation, scale, and robustness to affine transformations and illumination changes as well, scale invariant feature transform (SIFT) [5] introduced by Lowe is widely used for object detection, recognition, and tracking. Recent performance

* Corresponding author.
E-mail addresses: changcaiyang@gmail.com (C. Yang), zhouhuabing@gmail.com (H. Zhou), mycalos@gmail.com (S. Sun), claye.liu@gmail.com (R. Liu), zhaoji84@gmail.com (J. Zhao), jyma2010@gmail.com (J. Ma).

http://dx.doi.org/10.1016/j.infrared.2014.07.011
1350-4495/

evaluation [6] has shown that the SIFT descriptor outperforms other local descriptors. However, the IR face images often involve low texture; as a result, very few feature points could be extracted. Moreover, the matching process of SIFT uses only local appearance features, which inevitably leads to a number of false matches, or mismatches, causing difficulties for the following recognition process. This situation will be even worse if the image suffers from occlusion problems, for example, there exist eyeglasses in the images [3]. Therefore, establishing good feature correspondence is a critical prerequisite and a challenging task for IR face recognition.

To improve the matching accuracy of SIFT features, many methods have been introduced in the literatures [2,4,7–10]. Pele and Werman [7] proposed to use the Earth Mover's Distance (EMD) [11] instead of the L_2 metric [5] for comparing SIFT descriptors. Tan et al. [4] introduced a SWF-SIFT method which adds an additional step based on SIFT to filter the mismatches. It was later extended in [2] to two other similar filter patterns, such as CWF-SIFT and YWF-SIFT. These methods [2,4] are evaluated on IR face images and have shown good matching performance. However, similar to SIFT, they still only use the local appearance information to match feature points; therefore, mismatches are still inevitable. Furthermore, the filters used in the methods [2,4] also eliminate a lot of correct matches, which could be fatal for IR face recognition since the IR images typically contain very few feature points. There also exist some methods utilizing global constraints to reject false SIFT correspondence [8,9], but mainly focus on visible images.

In this paper, we consider the global spatial relationship between the SIFT correspondences, and propose an efficient feature matching method which is able to establish accurate correspondences without hurting the correct matches. We illustrate our main idea in Fig. 1. The left IR image pair presents a set of initial SIFT correspondences, where blue lines indicate correct matches and red lines indicate mismatches. The right figure is the corresponding motion field, in which each arrow denotes a correspondence: the head and tail correspond to the positions of feature points in two images. The upper five blue arrows are the five matches on the head under rotation, and the rest blue arrows are correspondences on the clothes under non-rigid motion. We see that the blue arrows are regular and the global spatial motion is slow-and-smooth [12], while the distribution of the red arrows seems random. Therefore, if we can fit a smooth spatial mapping function for the underlying correct matches/blue arrows, then the mismatches will be eliminated accordingly. And this is the very goal in our work. More specifically, we associate each correspondence with a matching confidence variable, and fit a non-rigid mapping function by using the thin-plate spline (TPS) model [13,14]. The procedure involves two techniques such as softassign and deterministic annealing [15].

The rest of this paper is organized as follows. Section 2 presents our method for good feature matching based on the estimation of spatial mapping function. Section 3 illustrates our robust algorithm for estimating spatial mapping function from feature correspondences by using global smoothness constraint. The experiments are provided in Section 4 and the conclusions are drawn in Section 5.

2. Establishing good feature correspondence between IR face image pairs

Suppose we are given a pair of IR face images. The goal is to establish accurate feature correspondence between them, and hence improve the performance of IR human face recognition. To this end, we first use SIFT to extract a set of feature correspondences from the images, i.e., $S = \{(\mathbf{u}_n, \mathbf{v}_n)\}_{n=1}^{N}$. Throughout this paper, we use homogeneous coordinates for the image point, i.e. $\mathbf{u} = (\mathbf{u}^x, \mathbf{u}^y, 1)^{\mathrm{T}}$. The correspondence set S typically contains a lot of false matches; therefore, we need a second step to eliminate the mismatches by using some global geometric constraints.

As mentioned in Fig. 1, the global spatial motion is often slow-and-smooth for the correct matches, and random for the false matches. This indicates that the correct matches satisfy a certain smooth mapping function $\mathbf{f}$, i.e., $\mathbf{v} = \mathbf{f}(\mathbf{u})$ if $\mathbf{u}$ corresponds to $\mathbf{v}$. Clearly, if $\mathbf{f}$ is recovered successfully, the mismatches can be eliminated accordingly since they typically do not satisfy the smooth mapping $\mathbf{f}$.

3. Estimating spatial mapping function via SIFT-GSI

In the following, we will first model the mapping function $\mathbf{f}$ and then demonstrate how to estimate it robustly.

3.1. Mapping function

Generally, the motion of the human face is non-rigid, and hence parametrical model such as affine transformation or homography will not be suitable for modeling the mapping function $\mathbf{f}$. In this paper, we model $\mathbf{f}$ by a non-parametrical model, i.e., the thin-plate spline (TPS) [13]. The TPS is a general purpose spline tool which produces a smooth functional mapping for supervised learning. It has no free parameters that need manual tuning, and also has a close-form solution which can be decomposed into a global linear affine motion and a local non-affine warping component controlled by coefficients $\mathbf{A}$ and $\mathbf{W}$ respectively:

$$\mathbf{f}(\mathbf{u}) = \mathbf{u} \cdot \mathbf{A} + \widetilde{K}(\mathbf{u}) \cdot \mathbf{W}, \tag{1}$$

where $\widetilde{K}(\mathbf{u})$ is an $1 \times N$ vector defined by the TPS kernel, i.e. $K(r) = r^2 \log r$, and each entry $\widetilde{K}_n(\mathbf{u}) = K(|\mathbf{u} - \mathbf{u}_n|)$. Here to reduce the computational complexity, we adopt a sparse approximation and choose an arbitrary subset $\{\tilde{\mathbf{u}}_m\}_{m=1}^{M} (M \ll N)$ from the original input points $\{\mathbf{u}_n\}_{n=1}^{N}$ as control points [16], and hence $\widetilde{K}(\mathbf{u}) \in \mathbb{R}^{1\times M}$ with $\widetilde{K}_m(\mathbf{u}) = K(|\mathbf{u} - \tilde{\mathbf{u}}_m|), A \in \mathbb{R}^{3\times 3}$, and $\mathbf{W} \in \mathbb{R}^{M\times 3}$. Loosely speaking, the TPS kernel contains the information about the internal structural relationships between the SIFT correspondences and is

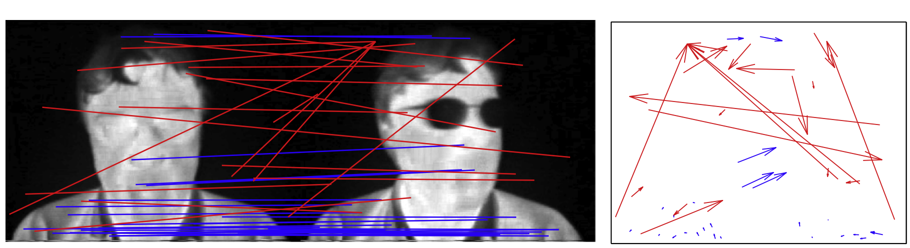

Fig. 1. SIFT correspondence (left) and motion field (right) between a pair of IR images.