Infrared Physics & Technology 69 (2015) 145–154

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Automatic visible and infrared face registration based on silhouette matching and robust transformation estimation

Tian Tian [a], Xiaoguang Mei [b], Yang Yu [c], Can Zhang [b], Xiaoye Zhang [d,*]

[a] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Shanghai Institute of Technical Physics, Chinese Academy of Science, Shanghai 200083, China
[d] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

HIGHLIGHTS

- The silhouette is used to represent a face image for multi-sensor data registration.
- The spatial transformation is found using a robust maximum likelihood framework.
- An automatic visible and IR face registration method is proposed.
- Results on both synthetic and real data show the advantages of our method.

ARTICLE INFO

Article history:
Received 10 November 2014
Available online 7 February 2015

Keywords:
Registration
Face recognition
Infrared
Image fusion
Silhouette

ABSTRACT

Registration of multi-sensor data (particularly visible color sensors and infrared sensors) is a prerequisite for multimodal image analysis such as image fusion. In this paper, we proposed an automatic registration technique for visible and infrared face images based on silhouette matching and robust transformation estimation. The key idea is to represent a (visible or infrared) face image by its silhouette which is extracted from the image's edge map and consists of a set of discrete points, and then align the two silhouette point sets by using their feature similarity and spatial geometrical information. More precisely, our algorithm first matches the silhouette point sets by their local shape features such as shape context, which creates a set of putative correspondences that may contaminated by outliers. Next, we estimate the accurate transformation from the putative correspondence set under a robust maximum likelihood framework combining with the EM algorithm, where the transformation between the image pair is modeled by a parametric model such as the rigid or affine transformation. The qualitative and quantitative comparisons on a publicly available database demonstrate that our method significantly outperforms other state-of-the-art visible/infrared face registration methods. As a result, our method will be beneficial for fusion-based face recognition.

1. Introduction

Multi-sensor data is often available in many fields such as computer vision, remote sensing, medical imaging, as well as military applications, and it is able to provide complementary information about the region surveyed [1–4]. Image fusion which aims to create higher resolution images from such data offering more complex and detailed scene representation has then emerged as a promising research area. Particularly, for the face recognition problem, the use of multi-sensor data such as visible and thermal infrared (IR) images has shown to be able to achieve better recognition performance [5,6]. For example, visible information is better for establishing a discriminative face model, while thermal IR images are not affected by illumination variation or face disguise. However, successful image fusion requires a critical and challenging step that the image pairs to be fused have to be correctly co-registered on a pixel-by-pixel basis [7–9]. In this paper, we focus on registration of visible and thermal IR face images for fusion-based face recognition.

For face recognition systems, registration of multi-sensor images can be implemented accurately by hardware. However,

* Corresponding author.
E-mail addresses: tiantian@hust.edu.cn (T. Tian), meixiaoguang@hust.edu.cn (X. Mei), yuyang@mail.sitp.ac.cn (Y. Yu), zhangcan318@gmail.com (C. Zhang), xiaoyza@gmail.com (X. Zhang).

http://dx.doi.org/10.1016/j.infrared.2014.12.011
1350-4495/

due to elevated cost and low availability, a special-purpose imaging sensor assembly producing co-registered image pairs in the visible and thermal IR domains may not be practical in many situations. In this context, software-based registration utilizing off-the-shelf low cost visible and infrared cameras without requiring any additional hardware may be more appropriate for large-scale deployment. While the accuracy of the pixel-to-pixel alignment can be reduced compared to the hardware-based registration, the salient image structures can still be matched sufficiently well in a software-based registration process [5]. In this paper, we are interested in software-based registration techniques.

To perform the software-based registration, we should first determine what type of information should be extracted to represent the image. In many visible stereo methods, this information could be appearance features such as graylevels/colors, textures and gradient histograms, where the global statistical dependence of the images to be aligned is assumed [10,2]. However, the application of visible stereo methods for visible-infrared camera configurations is not straightforward since infrared and visible images are manifestations of two different phenomena, and hence those typical appearance features will not likely match. Instead, a significant common feature that might be preserved in visible-infrared face images is the edge [11,5,6], as shown in the middle column of Fig. 1. Nevertheless, the edge map still involves a lot of noise, for example, the edges on the eyes, nose and mouth of the image pair do not represent exactly the same physical locations. To address this challenging problem, here we consider matching the object silhouettes in the image pair, which are much more stable and "exactly matchable", as shown in the right column of Fig. 1. The silhouettes are further discretized into point sets, and the registration problem then reduces to determining the correct point correspondence and to finding the underlying spatial transformation between the two point sets.

The next issue in the registration problem is to determine the transformation model between image pairs. Registration problems can be categorized into rigid or nonrigid case depending on the application and the form of the data. Rigid registration (similar to affine and projective) is relatively simple and involves a small number of parameters, while non-rigid registration is more difficult and the transformation is often unknown, complex, and hard to model [12–14]. Since we focus on registration of visible-infrared face image pairs which are typically captured at the same times and from similar viewpoints, a parametric model such as rigid or affine transformation could be accurate enough to match the salient face structures. Therefore, we focus on the rigid case and investigate a robust algorithm for rigid silhouette registration.

There are two unknown variables we have to solve for in the registration problem: the correspondence and the transformation. These two variables share an intimate relationship, e.g., the solution for one variable is actually mostly trivial once the other is known. Given the set of correspondences, finding a good transformation is often a straightforward least-squares problem. On the other hand, given a transformation, we can apply it to one point set and determine the set of correspondences using some proximity criteria. To solve the registration problem, a popular strategy is to first construct a putative set of correspondences between the two point sets based on their local feature descriptor vectors (e.g., SIFT [15] or shape context [16]), and then estimate the transformation model parameters according to the obtained putative correspondence set [17,12]. But for face silhouette registration, the local shape features of silhouette points are often not discriminative enough (e.g., many silhouette points have similar descriptor vectors); this typically results in a number of false correspondences in the putative set. In this situation, the estimate of the transformation will degrade badly unless it is performed robustly. To address this issue, in this paper we proposed a robust algorithm to estimate the transformation from a given correspondence set involving noise and outliers.

More precisely, we first represent visible/infrared face images by the silhouettes extracted from their edge maps, and then match the silhouette point sets based on their local feature descriptors to generate a putative correspondence set. In this process, the edge maps can be extracted by the Canny edge detectors [18], and feature descriptor such as shape context is used to establish correspondence. Subsequently, we focus on estimating the accurate transformation from the putative correspondence set. We formulate this as a maximum likelihood estimation of a Bayesian model with hidden/latent variables indicating whether correspondences in the putative set are outliers or inliers. The transformation between the image pair is modeled by a parametric model such as the rigid or affine transformation. The problem is solved by using the EM algorithm, and a closed-form solution is derived in the M-step. Finally, the alignments between image pairs could be obtained according to the estimated transformation.

It is worth to mention that matching the silhouettes instead of the edge maps will decrease the number of the possible correct point correspondences. This is obvious but we are more concerned whether we are able to recover the right transformation from the estimated correspondences. The answer is yes as we only need 3 correspondences to find the rigid transformation parameters and we may keep only a part of the good correspondences (e.g., correspondences on the silhouettes) to find the best rigid alignment.

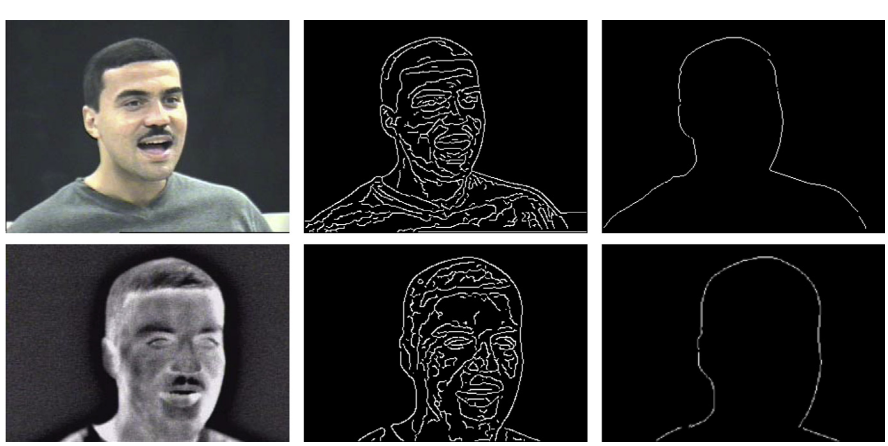

Fig. 1. Schematic illustration of the edge maps (middle column) and silhouettes (right column) of the visible (upper row) and thermal IR (bottom row) images.