Pattern Recognition Letters 36 (2014) 254–260

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Fused intra-bimodal face verification approach based on Scale-Invariant Feature Transform and a vocabulary tree

Carlos M. Travieso *, Marcos del Pozo-Baños, Jesús B. Alonso

Signals and Communications Department, Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria, Las Palmas de Gran, Canaria, Spain

ARTICLE INFO

Article history:
Available online 30 August 2013

Communicated by Luis Gomez Deniz

Keywords:
Bimodal interaction
Information fusion
Visible and thermal face verification
Face detection
SIFT parameters
Vocabulary tree

ABSTRACT

This work studies the intra-bimodal face-based biometric fusion approach composed of the thermal and spatial domains. The distinctive feature of this work is the use of a single camera with two sensors which returns a unique image with thermal and visual images at a time, as opposed to the-state-of-the-art, for example the multibiometric modalities and hyperspectral images. The proposed system represents a practical bimodal approach for real applications. It is composed by a verification architecture based on the Scale-Invariant Feature Transform algorithm (SIFT) with a vocabulary tree, providing a scheme that scales efficiently to a large number of features. The image database consists of front-view thermal and visual image as a single image, which contain facial temperature distributions of 41 different individuals in 2-dimensional format and 18 images per subject, acquired on three different-day sessions. Results showed that visible images gives better accuracy than thermal information, and with independency of range, head images give the most discriminative information. Besides, fusion approaches reached better accuracy, up to 99.45% for score fusion and 100% for decision fusion. This shows the independency of information between visual and thermal images and the robustness of bimodal interaction.

1. Introduction

The usage of different biometric systems on security applications has become more and more common nowadays. The reason is a series of advantages versus other methods such as carrying magnetic cards or reminding passports or PIN numbers, which can be forgotten or used by non authorized persons. Identification systems based on human body measures are well accepted and perceived naturally by both men and women. Therefore, biometric identification methods are achieving outstanding results and truthfulness in the security market.

Human recognition through distinctive facial features supported by an image database is still been studied. Note that this problem presents various difficulties. What will occur if the individual's haircut is changed? Is make-up a determining factor in the process of verification? Would it distort the facial features significantly?

The usage of thermal cameras, originally conceived for military purposes, has expanded to other fields of application such as control process in production lines, detection/monitoring of fire and even security and anti-terrorism applications. Therefore, its use in human identification tasks, in scenarios where the lack of light restricts the operation of conventional cameras, can also be considered. Thermal cameras can also be a great tool against look variations, which in some cases could be quite extreme. Different looks of the main role from the film The Saint are shown in Fig. 1. Val Kilmer modifies his look in this film spectacularly in order to not to be recognized by the enemy.

A correct matching between the test face and that stored in the image database is expected, and this is a hard task to solve even if natural distortion effects such as illumination changes or interference are not considered. The recognition problem should be split in three stages, that is, acquisition of facial images for testing, features extraction from specific facial regions and finally, verification of the individual's identity (Soon-Won et al., 2007).

Currently, computational face analysis is a very lively research field, in which new interesting possibilities are being studied. For example, there are approaches aiming to improve a system's performance when working with low resolution images (LR) and decreasing computational load.

In Huang and He (2011), a facial recognition system was presented, which works with LR images using nonlinear mappings to infer coherent features that favor higher accuracy of the nearest neighbor (NN) classifiers for the recognition of a single LR face image. It is also interesting to cite the approach of Imtiaz and Fattah (2011), in which a multi-resolution feature extraction algorithm for face recognition, based on two-dimensional discrete

* Corresponding author.
E-mail addresses: ctravieso@dsc.ulpgc.es (C.M. Travieso), mpozo@idetic.eu (M.d. Pozo-Baños), jalonso@dsc.ulpgc.es (J.B. Alonso).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.08.016

Fig. 1. Facial changes of the character played by Val Kilmer in the film The Saint.

wavelet transform (2D-DWT), was proposed. Such method exploits local spatial variations in a face image effectively, obtaining outstanding results with 2 different databases.

Images from subjects are often taken in different poses or with different modalities, such as thermographic images, presenting different stages of difficulty in their identification.

In Socolinsky and Selinger (2004), results on the use of thermal infrared and visible imagery for face recognition in operational scenarios were presented. These results showed that thermal face recognition performance is stable over multiple sessions in outdoor scenarios, and that fusion of modalities increases performance.

In the same year 2004, L. Jiang proposed in Jiang et al. (2004) an automated thermal imaging system that is able to discriminate frontal from non-frontal face views with the assumption that at any time, there is only 1 person in the field of view of the camera and no other heat-emitting objects are present. In this approach, the distance from centroid (DFC) shows its suitability for comparing the degree of symmetry of the lower face outline.

The use of correlation filters in Heo et al. (2005) showed its adequacy for face recognition tasks using thermal infrared (IR) face images due to the invariance of this type of images to visible illumination variations. The results with Minimum Average Correlation Energy (MACE) filters and Optimum Trade-off Synthetic Discriminant Function (OTSDF) in LR images (20×20 pixels) prove their efficiency in Human Identification at a Distance (HID).

Scale Invariant Feature Transform (SIFT) algorithm Lowe, 1999 is widely used in object recognition. In Soyel and Demirel (2011), SIFT appeared as a suitable method to enhance the recognition of facial expressions under varying poses over 2D images. The usage of affine transformation consistency between two faces to discard SIFT mismatches has been demonstrated.

Gender recognition is another lively research field working with SIFT algorithm. In Jian-Gang et al. (2010), faces were represented in terms of dense-Scale Invariant Feature Transform (d-SIFT) and shape. Instead of extracting descriptors around interest points only, local feature descriptors were extracted at regular image grid points, allowing dense descriptions of face images.

However, SIFT usually generates a large number of features from an image. This huge computational effort associated with feature matching limits its application to face recognition. To overcome this problem, Majumdar and Ward (2009) proposed the usage of a discriminating method. Computational complexity was reduced more than 4 times and accuracy increased in 1.00% on average by checking irrelevant features.

Another interesting idea is the building method, which is well scaled with the size of a database and allow finding one element of a large number of objects in acceptable time. This work is inspired by Nister and Stewenius (2006), where object recognition by a *k*-means vocabulary tree was presented. Efficiency was proved by a live demonstration that recognized CD-covers from a database of 40,000 images. The vocabulary tree showed good results when a large number of distinctive descriptors form a large vocabulary. Many different approaches to this solution have been developed in the last few years (Ober et al., 2007) and (Slobodan, 2008), showing its competency organizing several objects. Based on this good results, this solution will be tested in this paper with SIFT descriptors in a vocabulary tree.

In addition, references using two different images in different ranges; visible and infrared-thermal, can be found in the state-of-the-art. In Buyssens and Revenu (2010), the authors used a PCA and Sparce analysis before applying a fusion module. This reaches mean recognition rates between 95% and 99% after the fusion process for 63 users. In Bhowmik et al. (2012), the system fuses the thermal and visible images; captured by two individual sensors, in a unique image. Using 70% of the visual image and 30% of the thermal image and classifying with a SVM, the system reaches up to 97.28% on an identification approach.

Other lines of research, multimodal biometric systems have been focused. For example, Almohammad et al. (2012) was based on the fusion of face and gait biometrics, Tong et al. (2010) was based on face and fingerprint biometric fusion, Javadtalab et al. (2011) on the fusion of face and ear recognition, and Raghavendra (2012) on the feature level fusion of face and palmprint biometrics. This kind of multimodal approaches usually need longer times or uncomfortable devices from the user and application points of view.

Finally, different references based on Multispectral Face Recognition and Multimodal Score Fusion can also be found. Zheng and Elmaghraby (2011) and Bourlai et al. (2012) are two examples, where authors used different sensors and cameras; or one camera for a certain range by bandpass filters or for broadband.

In this context, the aim of the present work is to propose, innovate and evaluate on the field of bimodal face biometrics, for visible and thermal ranges. The proposed method could be used in a real application with the convenience of a unique device for fast tracking and the advantages of the fusion of bimodal information. All this gives it an added value versus the state-of-the-art. In addition, a study to search the main source of information is also included here. In particular, the system applies the SIFT algorithm and obtains local distinctive descriptors from each image based on Crespo et al. (2012). The construction of the vocabulary tree enables to have these descriptors hierarchically organized and ready to carry out a search to find a specific object.

For each test image, only its new descriptors are calculated and used to search through the hierarchical tree in order to build a vote matrix, in which the most similar image of the database can be easily identified. This approach mixes the singularity of the SIFT descriptors to perform reliable matching between different views of a visual and thermal face, and the efficiency of the vocabulary tree for building a high discriminative vocabulary. A more detailed description of the system is provided in the next subsections.

This paper is organized as follows. The proposed approach is presented in Section 2. In Section 3, the experimental settings, results and discussions are shown. Finally, conclusions are given in Section 4.

2. Approach proposed

The innovation upon this work focuses in the implementation of fused bimodal face verification approach implemented by a unique device, which gives an image from two sensors for visible and thermal ranges respectively. Finally, a bimodal verification approach is implemented using SIFT descriptors as feature extraction and a vocabulary tree through the use of the *k*-means function as a classification system. The score and decision fusions for both ranges have been applied. Besides, localized discriminative information has been searched between ranges and its fusions, and between regions of interest (head versus face). This work is a novel study and opens a door for an application on real conditions.

In this section, the whole approach and each of its parts will be explained.