Information Sciences 268 (2014) 122–132

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Assessing the level of difficulty of fingerprint datasets based on relative quality measures

Shengzhe Li [a], Hakil Kim [a,*], Changlong Jin [b], Stephen Elliott [c], Mingjie Ma [d]

[a] Graduate School of Information and Communication Engineering, Inha University, Incheon, Republic of Korea
[b] Department of Computer Science, Shandong University at Weihai, China
[c] Department of Technology, Leadership and Innovation, Purdue University, West Lafayette, USA
[d] Department of Statistics, Inha University, Incheon, Republic of Korea

ARTICLE INFO

Article history:
Available online 10 June 2013

Keywords:
Level of difficulty
Fingerprint dataset
Biometric similarity score
Sample quality
Relative sample quality
Common area

ABSTRACT

Understanding the difficulty of a dataset is of primary importance when it comes to testing and evaluating fingerprint recognition systems or algorithms because the evaluation result is dependent on the dataset. The difficulty exhibited in this paper represents how difficult it is to achieve better recognition accuracy within the specific dataset. Proposed in this paper is a general framework for assessing the level of difficulty of fingerprint datasets based on quantitative measurements of not only the sample quality of individual fingerprints but also the relative differences between genuine pairs, such as common area and deformation. The experimental results over various datasets demonstrate that the proposed method can predict the level of difficulty of fingerprint datasets which coincide with the equal error rates produced by four comparison algorithms. The proposed method is independent of comparison algorithms and can be performed automatically.

1. Introduction

As the worldwide deployment of fingerprint recognition systems has been increasing, the demand for evaluating their performance is also growing rapidly. Many public and private organizations, including academia, have conducted technology evaluations of fingerprint recognition systems with their own datasets. Because these datasets are collected by different organizations and there is no specific method for measuring the difficulty of a dataset, the evaluation results over various datasets cannot be compared. The difficulty exhibited in this paper represents how difficult it is to achieve better recognition accuracy within the specific dataset.

To test the performance of a fingerprint recognition system, the dataset should be on a standardized corpus, ideally collected by a "universal" sensor (i.e., a sensor that collects samples equally suitable for all algorithms tested) [1]. Nonetheless, performance against this corpus will depend on both the environment and the population in which data is collected. Furthermore, building a standardized corpus with a "universal" sensor is also impractical. Therefore, there are inevitable differences in difficulty between datasets. Because of this, there have been other approaches to quantify performance of the biometric system. One method is that of the biometric zoo or menagerie, which examines the performance of the individual. The most popular zoo examples are that of Doddington et al. [14] and Yager et al. [29].

Doddington et al. [14] originally suggested the biometric zoo in the speech and speaker recognition systems and stated that one major factor affecting performance of these systems is inherent differences in the recognizability of different

* Corresponding author. Tel.: +82 10 6336 7385.
E-mail addresses: szli@vision.inha.ac.kr (S. Li), hikim@inha.ac.kr (H. Kim), cljin@sdu.edu.cn (C. Jin), elliott@purdue.edu (S. Elliott), mmj810611@hotmail.com (M. Ma).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.05.025

speakers. Yager et al. [29] extended Doddington's zoo and defined additional four types of animals in terms of a relationship between a user's genuine and impostor comparison results. The biometric zoo focuses on the results of an algorithm on a dataset. Thus, the results of zoo analysis are algorithm-dependent and difficult to interpret across the datasets. Some studies take account of the user's characteristic and developed many score normalization methods in order to predict the performance of the biometric systems [23,24,26,27]. Poh et al. [26] suggested the biometric menagerie index (BMI) that defined as the ratio of the between-client variance and the expectation of the total variance of user's genuine and imposter comparison. Teli et al. [27] proposed a generalization method for the biometric zoo based on a hierarchical framework. However, these methods still based on the specific algorithms and the difficulty of a dataset cannot directly be represented. O'Connor has proposed a stability score index, which provides insight into particular users who perform poorly or exceptionally well in a particular dataset [25].

Many studies have shown that the sample quality (SQ) of a fingerprint strongly affects the performance of a recognition system [9,11,13,15–17,21]. The SQ of a fingerprint is considered to be the reliability of the features that are extracted from the fingerprint, and it can be adopted as a certain weight to reflect how well it can provide information for comparison algorithms to improve recognition performance. However, the SQ of a single fingerprint cannot represent the relative rotation and deformation (DF) or common area (CA) between the reference and the probe images to be compared. Image quality metrics only can quantify the quality of the reference and probe images separately.

The purpose of this paper is to provide a statistical method for assessing the level of difficulty (LOD) of a given fingerprint dataset by measuring the CA, DF and relative sample quality (RSQ) of mated pairs. These factors are combining into a single LOD score by several methods, such as, linear regression, polynomial regression and neural network. The Kruskal–Wallis test is employed to test the statistical significance of the difference among the resulting LODs for datasets [28]. The proposed method can be applied for characterizing and measuring the level of difficulty of fingerprint datasets used in technology evaluation.

The preliminary version of this paper was previously reported in a conference paper [20]. The present paper provides a complete description of the model selection of the LOD and a comparison with the biometric zoo.

2. Definition and model of LOD

2.1. Definition of LOD

The LOD is defined as a relative measure of a fingerprint dataset that represents how "challenging" or "stressing" the fingerprint dataset is for recognition compared to other datasets [2]. There are several influential factors in the performance of fingerprint recognition, such as sensor type (e.g., total internal reflection, capacitance, thermal, swipe, touchless, ultrasonic, etc.), impression type (e.g., flat, rolled, segmented slap, scanned ink-print, etc.), image resolution, environmental conditions (e.g., temperature, humidity, etc.), demographics (e.g., age, gender, occupation, etc.), finger position (e.g., thumb, index, etc.), cross sensor/cross impression type comparisons, template aging, and subject cooperation [3]. However, some of these factors are impossible to be quantitatively measured from the fingerprint images. In this paper, CA, DF and relative sample quality (RSQ) are considered to be the most influential factors for representing the LOD.

2.2. Common area

Regardless of the comparison algorithms, the CA of a mated pair is one of the major factors determining the similarity score. In general, a larger CA results in a higher similarity score, and conversely, a smaller common area leads to a lower similarity score. Therefore, the similarity score of the mated pair can be considered to be proportional to the CA between the pair. Considering the difference in image size and resolution between datasets, it is more appropriate to define the measure of the CA as the ratio of the intersection between two images to their union rather than the actual common area, as follows:

$$\mathrm{CA} = \frac{F_r \cap F_p}{F_r \cup F_p} \tag{1}$$

where F_r and F_p denote the fingerprint foreground in the reference and the probe fingerprints, respectively. CA ranges from 0 to 1, where a larger CA indicates a larger ratio of the common area between the mated pair. The main steps for computing the CA are summarized as follows:

(1) Segment the input fingerprint pair.
(2) Detect the alignment point of the two fingerprints.
  (a) for non-arch fingerprints, the mated singular points are selected as the alignment points,
  (b) for arch fingerprints, the maximum point pair in the angular difference and the orientation certainty level along the symmetry line are selected as alignment points [19].
(3) Translate, rotate, and align the probe to the reference.
(4) Count the number of pixels in the overlapping region and compute CA between them.