Expert Systems with Applications 41 (2014) 752–764

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Fingerprint classification based on decision tree from singular points and orientation field

Jing-Ming Guo [a], Yun-Fu Liu [a], Jla-Yu Chang [a], Jiann-Der Lee [a,b,*]

[a] Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan
[b] Department of Electrical Engineering, Chang Gung University, Taipei, Taiwan

ARTICLE INFO

Keywords:
Fingerprint
Fingerprint identification
Fingerprint singularity
Decision tree classifier
Fingerprint analysis

ABSTRACT

In this study, a high accuracy fingerprint classification method is proposed to enhance the performance in terms of efficiency for fingerprint recognition system. The recognition system has been considered as a reliable mechanism for criminal identification and forensic for its invariance property, yet the huge database is the key issue to make the system obtuse. In former works, the pre-classifying manner is an effective way to speed up the process, yet the accuracy of the classification dominates the further recognition rate and processing speed. In this paper, a rule-based fingerprint classification method is proposed, wherein the two features, including the types of singular points and the number of each type of point are adopted to distinguish different fingerprints. Moreover, when fingerprints are indistinguishable, the proposed Center-to-Delta Flow (CDF) and Balance Arm Flow (BAF) are catered for further classification. As documented in the experimental results, a good accuracy rate can be achieved, which endorses the effectiveness of the fingerprint classification scheme for the further fingerprint recognition system.

1. Introduction

Fingerprint is widely used in individual identification, largely due to the bio-invariant characteristic of human fingerprints, which also provide more details for distinguishing various persons. Former "fingerprint verification" methods normally demand users to input their personal information through various means, such as a name or an ID card. This kind of system verifies the correspondence between the captured fingerprint and the user's personal information, yet the system is inefficient as the users have to operate the system. For example, keying in their name or directly inserting an ID card is an additional operation for user identification. To avoid this inconvenience, an alternative approach namely "fingerprint identification" (Maltoni, Maio, Jain, & Parbnhankar, 2009) which does not require user's interaction was presented, yet extensive processing complexity caused by its cross reference of fingerprints in the database is required. To cope with this, former works (Ratha, Chen, Karu, & Jain, 1999; Tan, Bhanu, & Lin, 2003) used a strategy to pre-classify the fingerprints in database into different categories. As a result, a fingerprint to be verified simply needs to cross-reference the fingerprints in the category which identical to the verified fingerprint. This manner effectively reduces the number of fingerprints for further matching process.

The researches focusing on fingerprint classification are discussed as below. For instance, in Henry's work (Henry, 1900) the fingerprints were separated into four classes (namely 4C): Arch (A), Whorl (W), Left loop (L), and Right loop (R), some examples are illustrated in Fig. 1(a), (c)–(e). Another classification approach indicates that the category A can be further classified into A and Tented arch (T) Watson & Wilson, 1992 (namely 5C), and Fig. 1(b) shows an example of the additional Tented arch category. Even more numbers of fingerprint categories are also employed (Cappelli, Lumini, Maio, & Maltoni, 1999), but which also raise other issues such as reduced accuracy (ambiguous categories even cannot be classified by experts (Maltoni et al., 2009; Tan & Bhanu, 2005). Thus, in this work, the 4C system is adopted as the standard for classification.

Feature extraction for fingerprint classification is another important issue. Several well-known types of methods have been proposed, including orientation field (OF), singular point (SP), ridge flow (RF), and Gabor filter (GF). Moreover, lots of classification methods based on these features are established, such as rule-based (RB), SYntactic (SY), STRuctural (STR), STAtistical (STA), Neural Network (NN), and Multiple Classifiers (MC). Among these, the RB approach is the most straightforward method than the others. This method relies on the acquired number and positions of the extracted singular points (Karu & Jain, 1996; Kawagoe & Tojo, 1984; Msiza, Leke-Betechuoh, Nelwamondo, & Msimang, 2009; Wang & Dai, 2007; Zhang & Yan, 2004) to classify fingerprints. Since the singular points cannot be extracted from a fingerprint

* Corresponding author. Tel.: +886 928856123.
E-mail addresses: jmguo@seed.net.tw (J.-M. Guo), yunfuliu@gmail.com (Y.-F. Liu), g78974110@hotmail.com (J.-Y. Chang), jdlee@mail.cgu.edu.tw (J.-D. Lee).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.07.099

Fig. 1. Fingerprint categories. (a) Arch (A). (b) Tented arch (T). (c) Whorl (W). (d) Left loop (L). (e) Right loop (R).

image directly, the SY classification approach (Chang & Fan, 2002) adopted the global distribution of 10 basic ridge patterns, the analysis of the ridge shapes, and the sequence of ridge distributions to achieve their work. The STR method (Cappelli et al., 1999) was proposed to partition the orientation field of a fingerprint into different orientation regions, and the related graphs of these regions were employed to classify fingerprint. Meanwhile, the STA approaches were based on a different input features to determine multi-dimensional regression equations. After that, the data features are extracted directly and input into the classifier, and the classified results can be obtained efficiently (Jung & Lee, 2009; Min, Hong, & Cho, 2006; Lia, Yau, & Wanga, 2008; Park & Park, 2005). The STA and NN approaches employed a special strategy to classify fingerprints as they imitated human perception and empirical model. These approaches require lots of training data to yield the classifier, and sufficient data should be obtained to yield a more effective classifier (Bernard, Boujemaa, Vitale, & Bricot, 2001; Kristensen, Borthen, & Fyllingsnes, 2007; Senior & Boll, 2004). Conversely, insufficient training data will be a problem since it significantly degrades the accuracy of a system. In all of the classification methods, the two key issues always affect the accuracy rate: (1) The quality of a fingerprint image, and (2) the ambiguity of a classification scheme.

In this work, clear and explicit rules are established to remove ambiguous classes of fingerprints. The RB technique is then applied since it is easy to implement, and which does not require a training procedure for classifier, while the high accuracy still can be achieved. Moreover, two issues affecting the accuracy rate indicated above are discussed. Finally, a decision tree is designed to realize an automatic fingerprint classification system.

The rest of this paper is organized as follows. Section 2 introduces the preprocessing and feature extraction steps. The results of our analysis and the descriptions of decision trees are provided in Section 3. Section 4 provides the experimental analyses and performance comparisons. Finally, conclusions are drawn in Section 5.

2. Preprocessing and feature extraction

The quality of the captured fingerprint ridge is very important since it dominates the singular point extraction. In an ideal case, a captured fingerprint should include sharp ridges and valleys, yet these are obstructed by other factors (Amengual et al., 1997). Thus, to achieve a better performance, an image enhancement process is highly demanded. In this study, the three public fingerprint databases FVC datasets (Fingerprint database – FVC, 2000, 2002, 2004) are adopted in this study for conducting experiments, in which the fingerprints are captured from different devices, thus perfect and imperfect fingerprints are involved. The critical issues involved can be solved by the proposed fingerprint classification as shown in Fig. 2, and the detail flows of the "preprocessing" is shown in Fig. 3 which is discussed below firstly.

2.1. Preprocessing

2.1.1. Histogram equalization

To obtain an image with stable contrast distribution, due to the foreground and background are simple, the global histogram equalization is utilized. The transformation function is formulated as below,

$$HE(i,j) = 255 \times \frac{\sum_{g=0}^{Image(i,j)} H(g)}{P \times Q}, \tag{1}$$

where $H(g)$ denotes the histogram value at grayscale g and variable $Image(i,j)$ denotes the grayscale value of the captured fingerprint image of size $P \times Q$ at location (i,j). Notably, in this study the full black and white colors are defined at 0 and 255, respectively. Fig. 4 shows a series of results of each sub-function in the "preprocessing" block, and the result corresponding to histogram equalization is shown in Fig. 4(b).

2.1.2. Grad field

The grad field represents the high-frequency energy distribution of the captured fingerprint image, and it can yield a mask to assist the following segmentation's process.

$$G_H(i,j) = \frac{\partial}{\partial j} Image(i,j), \tag{2}$$

$$G_V(i,j) = \frac{\partial}{\partial i} Image(i,j), \tag{3}$$

$$Grad(i,j) = \frac{1}{W^2} \sum_{u=i-\mathrm{round}(\frac{W}{2})}^{i+\mathrm{round}(\frac{W}{2})} \sum_{v=j-\mathrm{round}(\frac{W}{2})}^{j+\mathrm{round}(\frac{W}{2})} \sqrt{G_H^2(u,v)^2 + G_V^2(u,v)}, \tag{4}$$

where the notations H and V denote horizontal and vertical, respectively; the constant $W = 9$ denotes the average filter size; the round$(\cdot)$ denotes the round down operation. Fig. 4(c) shows two different examples of grad results.

2.1.3. Segmentation

The $HE(i,j)$ obtained by Eq. (1) is separated into foreground and background by the variance thresholding method (Mehtre, 1993). The processing steps are described as below:

$$AverageGrayScale(i,j) = 255 - \frac{\sum_{m=-r/2}^{r/2} \sum_{n=-r/2}^{r/2} HE(i+m, j+n)}{(r+1)^2}, \tag{5}$$

$$SegmentMap(i,j) = \frac{AverageGrayScale(i,j) + Grad(i,j)}{2}, \tag{6}$$

$$Mean = \frac{1}{P \times Q} \sum_{i=1}^{P} \sum_{j=1}^{Q} SegmentMap(i,j), \tag{7}$$

$$Var = \frac{1}{P \times Q} \sum_{i=1}^{P} \sum_{j=1}^{Q} (SegmentMap(i,j) - Mean)^2, \tag{8}$$