USC–SIPI REPORT #437

ADVANCED TECHNIQUES FOR LATENT FINGERPRINT ENHANCEMENT AND RECOGNITION

By

Jian Li

May 2017

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
USC Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

ADVANCED TECHNIQUES FOR LATENT FINGERPRINT ENHANCEMENT AND RECOGNITION

by

Jian Li

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

May 2017

I dedicate this dissertation the ones love me and I love, especially my parents and my sister, who always support me without any limitation and condition.

Acknowledgments

First, I would like to thank my PhD advisor, Professor C.-C. Jay Kuo deeply in my heart. He is forever my role model for both aspects of research and life. He is the right one to lead me, teach me, and help me to be better during my PhD study. It is my great honor to have chance working with and learning from him during the last 5 years. I will treasure all the memorable moments in the Media Communications Lab family at USC.

Second, I wish to acknowledge Professor Feng from Tsinghua University who is always patient and helpful enough to provide very useful advise to me, especially about my research. Besides, I would like to thank Dr. Songtao Li, Dr. Xian Tang and other researchers at 3M Cogent for helpful discussion. Computation for the work described was also supported by the University of Southern California's Center for High-Performance Computing (hpc.usc.edu).

Finally, I would also like to thank all my other committee members Prof. Ram Nevatia, Prof. Aiichiro Nakano, Prof. Antonio Ortega, Prof. Richard Leahy, and Prof. Justin Haldar, for their valuable feedback on this dissertation.

Contents

List of Tables

List of Figures

Abstract

Fingerprints provide one of the most popular biometric data, and have been widely used in individual person identification and verification. The Automated Fingerprint Identification System (AFIS) offers important evidences for criminal investigation, and serves as an important tool for law enforcement. As compared with conventional exemplar fingerprints, latent fingerprints are typically collected in a crime scene. They are often degraded and corrupted, leading to very low identification rates. In a practical system, a latent fingerprint has to be enhanced prior to feature extraction to ensure a reliable fingerprint matching performance. In this research, we study techniques for latent fingerprint enhancement and orientation field estimation to achieve a higher matching rate. Our studies include traditional image processing techniques as well as a new method based on the emerging convolutional neural network (CNN). Several major contributions are detailed below.

In Chapter 3, starting from traditional image processing basis, we propose a new method using the Markov random field (MRF) model and the sparse representation (SR) of ridges to enhance latent fingerprint. The proposed MRF-SR method is inspired by the recent success of dictionary-based methodologies (including both the orientation field dictionary and the ridge dictionary). The idea is detailed below. First, given a set of training local fingerprint patches, we obtain the over-complete sparse dictionary to form a pool of ridge patch candidates. Second, the texture component of latent fingerprints

is extracted using two total variation (TV) models: namely, the adaptive-directional TV (ADTV) model and the TV with the L1 fidelity regularization (TV-L1) model. Third, we assess local image quality using the structure similarity (SSIM) index to determine the proper total variation model to use for a local fingerprint patch in the next optimization step. Fourth, we define an MRF model with an unary potential and a neighbor interaction potential and use it to select the optimal patch candidates for the enhancement of a given patch. As compared with existing fingerprint enhancement techniques based on the ridge dictionary for denoising, the proposed MRF-SR method offers a better scheme for latent fingerprint enhancement with sparse optimization.

In Chapter 4, we show that the MRF-SR method can also be used to extract the orientation field. To begin with, we generalize several well-known traditional orientation estimation algorithms to the context of latent fingerprints. Then, we include the orientation field estimation technique by adding the orientation unary potential based on fingerprint pose to the cost function of the MRF formulation, which is MRF-SR-Modified. It provides a valuable supplementary tool to other orientation estimation algorithms in the literature. Finally, a fusion technique is adopted to boost the overall performance of latent fingerprint enhancement. As an essential feature of fingerprints, the orientation field can enhance a fingerprint image with directional and contextual filtering. Experimental results on orientation field estimation as well as latent fingerprint enhancement are given to demonstrate the effectiveness and robustness of the proposed fusion methodology.

In Chapter 5, we explore the feasibility and study the performance of applying the CNN to latent fingerprint enhancement. Being motivated by recent developments of CNN in image enhancement and restoration applications, we propose a novel encoding-decoding neural network, called the FingerNet. The FingerNet is trained in a pixelwise end-to-end manner for direct fingerprint enhancement. In particular, we develop a novel

data augmentation method to add structured noise (lines, characters, etc.) to good quality fingerprint patches so as to form meaningful training data. Then, we design a multi-task encoder-decoder network that has a leading convolutional module, which is followed by two deconvolutional branches. They are the enhancement branch and the orientation branch. After that, the cost functions of the two branches are added to form one final cost function. We successfully train the proposed FingerNet. The FingerNet outperforms benchmarking methods with a fast computational speed.

Chapter 1

Introduction

1.1 Significance of the Research

Being different from conventional exemplar fingerprints scanned from sensors or rolled with inking, latent fingerprints can be left on any surface of objects such as guns, glass bottles, and even newspapers unintentionally. They may not be visible directly, and some physical/chemical enhancement techniques are required to obtain images of latent fingerprints. Automated latent fingerprint identification can provide important evidence in supporting law enforcement such as for criminal identifying [24]. After a tremendous amount of efforts in last three decades, today's automated fingerprint identification system (AFIS) works well for plain and rolled full fingerprints with a very high rank-1 identification rate [54, 69, 70, 108]. On the other hand, it is common to have blurred ridge structure, uneven ridge/valley contrast, overlapping fingerprints, and structured noise (lines, stains, letters, etc.) in latent fingerprint images [19, 52, 118, 120]. Some exemplary latent fingerprints and their corresponding rolled fingerprint images are shown in Fig. 1.1 for visual comparison.

However, poor quality and special characteristics of latent fingerprints make feature extraction such as the region of interest (ROI), minutiae and singular points difficult to be detected or extracted automatically by machines. In practice, features of latent fingerprints are marked manually by expert examiners so that they can be identified against a large fingerprint dataset using AFIS. However, manual markup of features has three major issues: time consuming, compatibility and repeatability [30, 31, 52, 102, 101].

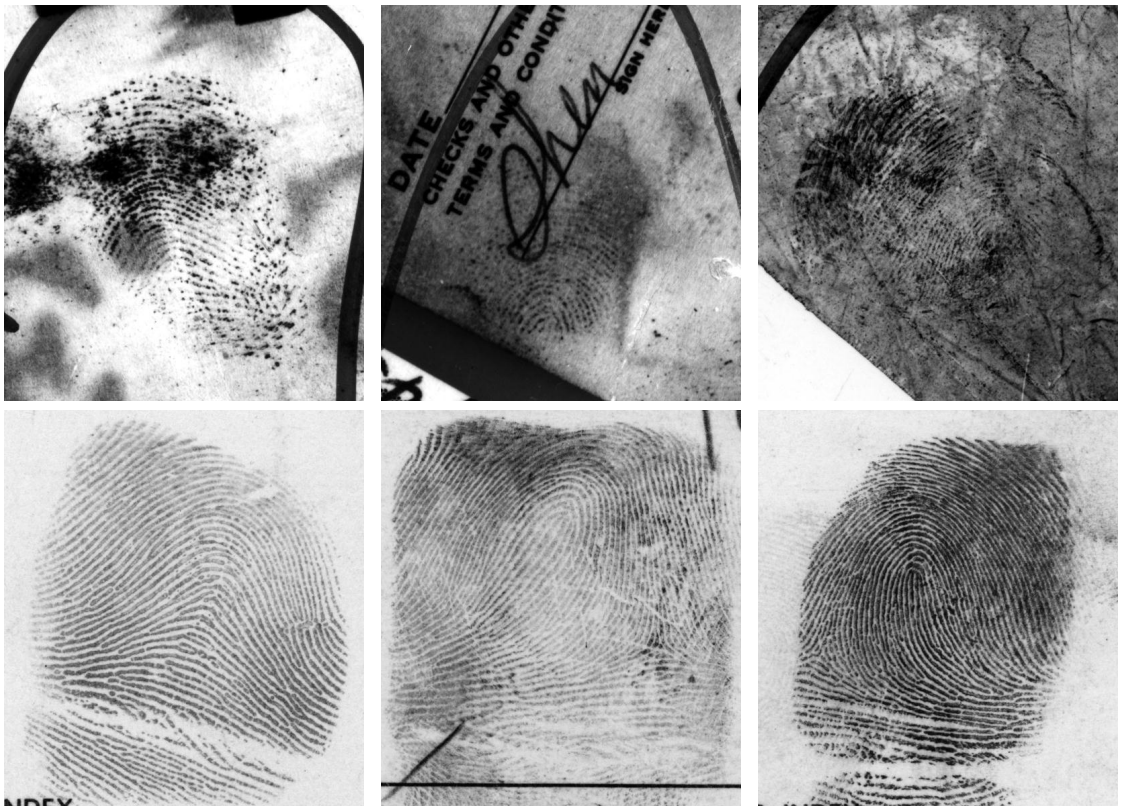


Figure 1.1: Three exemplary latent fingerprint images (the top row) and their corresponding rolled images (the bottom row) from NIST SD27.

The time cost is obvious since minutiae marking for a single latent image takes more than 20 minutes while the number of expert examiners is limited. By compatibility, we refer to the gap of feature extraction accuracy between manual markup in latent fingerprints and automatic extraction in full fingerprints. Finally, repeatability or reproducibility is about the stability of human markup. For the same latent fingerprint, different examiners may have different minutiae marking, and even the same examiner can have different markups at different times. Because of these issues, manual feature marking is not a good solution in the long term [42]. It is very much desired to develop an automatic high-performance feature extraction technique for latent fingerprints [6, 21, 23, 35, 96].

To perform automatic minutiae extraction directly on latent fingerprints tends to result in high missing and false alarm rates, which in turn decrease the identification rate of AFIS significantly [48, 49]. To ensure the success of automatic feature extraction, effective latent fingerprint enhancement as a pre-processing step is the key. By fingerprint enhancement, we improve the contrast between ridges and valleys, connect breaking ridges, separate joined ridges, and remove varieties of structured noise, etc. These efforts can enhance efficiency and robustness of automatic feature extraction such as minutiae detection, which plays significant role in fingerprint matching [36, 14].

1.2 Review of Previous Work

Quite a few fingerprint enhancement algorithms were proposed in literature based on directional or contextual filtering [2, 22, 38, 46, 58, 86, 95, 94]. O'Gorman *et al.* [86] used anisotropic smoothing filters with their directions aligned with the ridge orientation. A fingerprint image can be contrast-enhanced in a direction perpendicular to ridges and smoothed along ridge orientation by these filters. Sherlock *et al.* [94] enhanced fingerprint images by contextual filtering in the Fourier domain. That is, they convolved the fingerprint image with a directional filter whose orientation is aligned with the local ridge orientation everywhere. Almansa *et al.* [2] integrated a shape adaptation process and a scale selection process for fingerprint enhancement. The shape adaptation process achieves smoothing according to the local ridge structure while the scale selection process estimates the local ridge width and adjusts the smoothing level according to the noise level. Greenberg *et al.* [38] proposed a structure-adaptive anisotropic filter whose kernel can be shaped and scaled according to local features yet without local frequency estimation. Chikkerur *et al.* [22] estimated the local ridge orientation and frequency simultaneously using the 2-D Short Time Fourier Transform (STFT) in a probabilistic

manner and, then, applied contextual filtering in the Fourier domain to enhance fingerprint images. Since ridge flows can be viewed as sinusoidal waves with a specified local orientation and frequency, Hong *et al.* [46] enhanced the fingerprint image by applying Gabor filter banks that were tuned with the estimated local orientation and frequency field. Gabor filters are powerful in selecting parameters of a local pattern (*e.g.*, orientation and frequency) so as to achieve joint optimization in scale, frequency and orientation. Although the above-mentioned techniques work well for exemplar fingerprints, they do not perform well for latent fingerprints since the weak fingerprint information is buried by a variety of structured noise in the latter case. Generally speaking, it is difficult to estimate reliable contextual information from latent fingerprint images.

Various contextual filtering methods share one thing in common, *i.e.* orientation field (OF) estimation. Generally speaking, more accurate OF estimation leads to better fingerprint enhancement, which in turn results in a higher identification rate of AFIS. Thus, OF estimation plays a central role and attracts much attention. Several methods were proposed for the OF estimation of a small neighborhood. Examples include the gradient-based, the STFT-based and the Slit-based methods. The gradient-based method [46] estimates the domain orientation of a neighborhood of size 16×16 by computing the average squared-gradients. The STFT-based method defines a probability distribution using the polar form of the Fourier spectrum in a local region, and estimates the domain orientation as the most possible orientation or the mean [22]. The Slit-based method tests the contrast of a parameter (*e.g.*, the standard deviation of a neighborhood) along 16 pairs of perpendicular directions. and selects the highest contrast pair to indicate the ridge orientation [87].

Apparently, these local orientation estimation methods are sensitive to noise. In particular, they are not robust with respect to structured noise (*e.g.*, lines, printed letters and handwriting) that overlaps with the underlying fingerprint. To handle this challenge,

several smoothing methods were proposed as a post-processing step, where regularization is applied by adding more constraints on orientation estimation. Examples include low-pass filtering, multi-resolution smoothing with both small and large windows, and energy minimization using the Markov Random Field (MRF) [7, 68, 87]. Although local smoothing provides improved performance, it lacks the global constraints. To model global OF, global parametric models such as the polynomial and the FOMFE models [40, 105] were proposed. FOMFE treats the whole OF as a 2D matrix, and conducts the 2D Fourier expansion on this matrix. Fourier coefficients of basis functions are estimated and used to update the new OF. In general, global parametric models suffer from over-fitting or under-fitting and their performance highly depends on the accuracy of the initial OF.

Several OF estimation and fingerprint enhancement algorithms were proposed for latent fingerprints with the above-mentioned techniques as their building bricks in recent years [36, 113, 116, 117]. For example, Yoon *et al.* [117] proposed a hypothesize-and-test paradigm using the randomized-RANSAC to obtain an OF model with multiple orientation elements from the STFT method as the initial input. However, manually labeled ROI's and singular points are required for decomposing the OF into singular and residual components. Recently, the dictionary-based approach has achieved good results in OF estimation. Feng *et al.* [36] used an orientation patch dictionary to smooth the initial OF, where clean reference orientation patches are used to create a global dictionary as the prior of the fingerprint orientation. For every initial OF patch, they looked up this dictionary for potential candidates and, then, minimized an energy function to estimate the final OF using loopy belief propagation, where the energy function is defined by orientation similarity and neighboring compatibility. However, since this global dictionary does not consider the location-dependent information, the estimated OF of a patch may not be consistent with its spatial location in the fingerprint. To improve this, Yang *et*

al. [113] constructed different dictionaries at different fingerprint locations. They registered a latent fingerprint with a pose estimation algorithm based on the Hough transform and, then, estimated the OF of a patch by looking up the location-dependent dictionary. Although being more robust, these methods still rely on the accuracy of the initial OF. Besides, they do not incorporate the fingerprint ridge structure information in the OF update.

Instead of working on the OF estimation problem alone, some researchers attempted to exploit the ridge structure to help OF estimation or direct enhancement of latent fingerprints. Examples include the ridge dictionary method [12, 67] and the convolutional neural network (CNN) method [11]. The decomposition of a latent fingerprint image into texture and cartoon components using the total variation (TV) model offers an effective way to extract the ridge structure. That is, the ridge structure is kept as the oscillatory and small-scale pattern in the texture component while the piecewise-smooth content such as structured noise remains in the cartoon component. To improve the conventional TV model, Zhang and Kuo *et al.* proposed a scale-adaptive TV model in [118] and an adaptive directional TV (ADTV) model in [119]. The latter is adaptive to both the direction and the scale of a local region. Though the texture component is treated as the enhanced latent fingerprint in these methods, their performance is still limited in the strong noise region since their scale and orientation parameters cannot be estimated reliably.

With the texture component of latent fingerprints as the input, Cao *et al.* [12] proposed a coarse-to-fine ridge structure dictionary for OF estimation, enhancement and segmentation. A coarse ridge dictionary for larger patches and a set of fine ridge dictionaries (with each fine dictionary designed for one specified direction) for smaller patches are learned from high quality exemplar fingerprint patches using the sparse representation. Both orientation and frequency fields are first estimated using the coarse-level

dictionary and, then, refined by the fine-level dictionary. The final enhancement result is achieved using Gabor filter banks with the estimated frequency and orientation field. Being different from [12], Liu *et al.* [67] proposed an iterative dictionary look-up algorithm with multi-scale dictionaries. The multi-scale dictionaries are constructed directly from different sizes of Gabor element functions without learning, and they are used to reconstruct latent fingerprints. Coherence is used to measure fingerprint quality. Low quality reconstructed regions are improved by checking the dictionary of a larger patch size iteratively.

Cao *et al.* [11] treated the OF estimation problem as a classification problem, and proposed a convolutional neural network (ConvNet) method for its solution. A total of 128 orientation patterns were learned via ConvNet from 128,000 exemplar fingerprint patches with added texture noise. A latent fingerprint was decomposed into overlapping patches after pre-processing, and each patch was classified into one of the 128 orientation patterns by the trained ConvNet. All estimated orientation patches were stitched to form a global orientation field. Then, the Gabor filter with a fixed frequency was applied to obtain the final enhancement result. However, this method has several shortcomings. First, the noise types added to exemplar patches in the ConvNet training contained only lines and zero blocks, they were not sufficient in simulating rich structured noise in practical latent fingerprints. Second, simple quilting was adopted to stitch estimated orientation patches without considering compatibility among neighbors. Finally, the Gabor filter of a fixed ridge frequency has its limitation in the enhancement performance.

1.3 Contributions of the Research

Specific contributions of this research include the following.

- In the current literature, the localized OF dictionary method proposed in [113] offers one of the state-of-the-art enhancement results in latent fingerprint matching. Being inspired by this work as well as the success of sparse representation in image denoising [1, 32] and its application to latent fingerprints such as the ridge structure based methods [12, 67], we propose a new fingerprint enhancement method in this paper. Since it is based on the Markov Random Field (MRF) model and the sparse representation (SR) of ridges, it is called the MRF-SR method.

- The MRF-SR method consists of both local enhancement and global regularization. For local enhancement, the texture component of the latent fingerprint is first extracted by the adaptive directional total variation (ADTV) and total variation (TV) method [118, 119] to serve as the input to the proposed MRF-SR system. A local quality assessment is conducted to determine which method to use based on the local quality. Then, a robust local enhancement process is conducted based on Markov Random Field (MRF) modeling and loopy belief propagation, where overlapping fingerprint patches are treated as random variables and possible enhancement results of each random variable are obtained as structured outputs from a ridge-structured dictionary. The results are further regularized by the global information such as the estimated pose and the localized orientation dictionary for better performance.

- The proposed MRF-SR provides a robust solution to latent fingerprint enhancement by formulating the sparse representation in the MRF framework. To achieve this, we define the unary (vertex) potential based on structural similarity (SSIM) [106] and the interaction (or edge) potential based on combination of orientation and frequency for the MRF modeling. Global manner such as pose of fingerprint is also formulated in the unary potential definition. Due to this new mathematical

treatment, one can obtain enhanced fingerprints directly without the necessity to estimate the OF using directional filtering. It is demonstrated by extensive experimental results that we can achieve significant performance improvement by fusing various outputs of the proposed MRF-SR method and using it as the input of the AFIS.

- The proposed MRF-SR-Modified method can estimate the OF. The resulting OF estimation performs much better than other sparse representation methods with the ridge dictionary. Besides, the OF estimation result from MRF-SR-Modified can server as a diversity supplement to fingerprint orientation estimation research community to further fuse. Then, a machine learning based fusion methodology combining different OF estimation approaches demonstrates this by an overall performance boosting in the OF estimation.

- We propose a CNN-based solution, called the FingerNet, to latent fingerprint enhancement. This CNN design is tailored to the fingerprint application. The network has an encoder-decoder architecture. Its encoding part uses convolutional layers to extract fingerprint features for enhancement. To exploit two different feature types, its decoding part consists of two branches: an enhancement branch and an orientation branch. We train the network successfully and demonstrate the power of CNN in the latent fingerprint enhancement problem. Several deep learning concepts such as residual learning, multi-task learning and network variation are tested. Finally, we show the FingerNet can be executed very efficiently.

1.4 Organization of the Dissertation

This dissertation is organized as follows. The background knowledge about fingerprints is reviewed in Chapter 2. Then, the proposed MRF-SR latent enhancement method is

presented in details in Chapter 3. Chapter 4 introduces an orientation field estimation technique using the MRF-SR method and a fusion scheme. Chapter 5 presents the proposed FingerNet, which is a CNN based method for latent fingerprint enhancement. Finally, concluding remarks and future research directions are provided in Chapter 6.

Chapter 2

Background Review

Comparing with other physical or behavioral biometrics, fingerprint provides a better balance between theoretical and practical requirements for biometric information. For example, fingerprint outperforms face and hand geometry in term of both recognition performance and permanence. Although iris and retinal scan also have high performance, their collectability and users' acceptability are much lower than fingerprint. Therefore, fingerprint as the most popular human biometric plays an important role in automated verification and identification systems in practical use during last two decades. As the terrorism increasing during recent years, fingerprint is highly supporting law enforcement agencies for identifying and convicting criminals as critical evidence.

In this Chapter, background review about fingerprint is introduced including concepts and basics of fingerprint, processing techniques such as total variation models, and fingerprint enhancement from exemplar fingerprints to latent fingerprints.

2.1 Fingerprint Basics

2.1.1 Fingerprint Features

A fingerprint is the pattern of interleaved ridges and valleys that appears on a smooth surface when the fingertip is pressed [4, 16]. A ridge is defined as a single curved segment with dark grayscale, whereas a valley is bright which is the region between two adjacent ridges. Fig. 2.1 shows an visual example of ridges and valleys in a fingerprint.

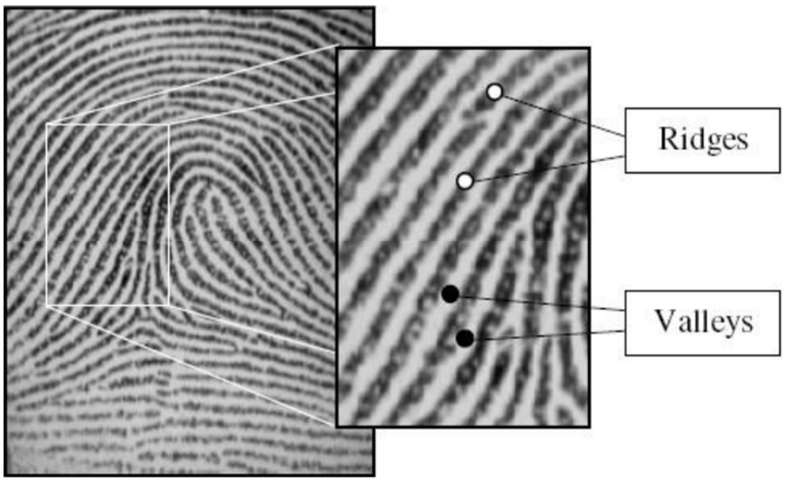


Figure 2.1: An example of fingerprint pattern showing ridges and valleys. Image Credits: [55, 79].

With ridges and valleys, fingerprint features can be extracted and they classified in to three different levels in general [79].

The global ridge details are described by Level-1 features, which include the followings. Fig. 2.2 is an illustration of some fingerprint features mentioned here.

- **Pattern type:** the overall ridge flow formation can be classified into the following 5 categories: Arch, Tented Arch, Left Loop, Right Loop and Whorl [13, 82].

- **Singular points:** they are control points named as core and delta where the ridges are "wrapped" that results in discontinuities in ridge orientation. A core point is the uppermost of a curved ridge, whereas a delta point is the location where three ridge flows meet.

- **Orientation field:** it is a map of local ridge directions of the ridge-valley structure. The direction is defined as angle between the ridge inclination and the horizontal line.

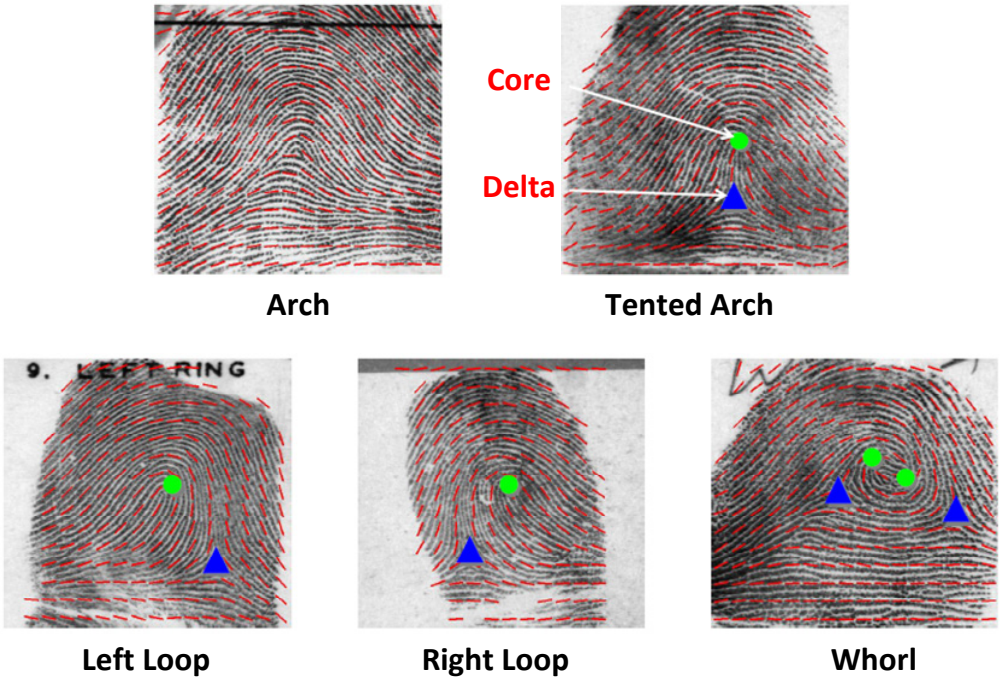


Figure 2.2: A visual illustration about fingerprint features such as pattern type, singular points and the orientation field (red lines indicate the local directions). Image Credit: online resource.

- **Frequency field:** it is a map of local ridge frequency, that is the reciprocal of the number of ridges per unit length along a window which is orthogonal to the local ridge orientation [53].

Level-2 features are about local ridge characteristics, among which two most prominent ones are called minutiae including ridge endings and ridge bifurcations. Ridge ending happens where a ridges ends abruptly. Ridge bifurcation is defined as the location where a ridge diverges into branch ridges. General speaking, minutiae in fingerprints are robust and stable to environment of fingerprint collecting compared with other representations. Minutiae is the key feature for fingerprint matching. More accurate minutiae detection offers higher matching rate for AFIS in general. Fig. 2.3 is a visual illustration about the two types of minutiae and the minutiae detected in a fingerprint.

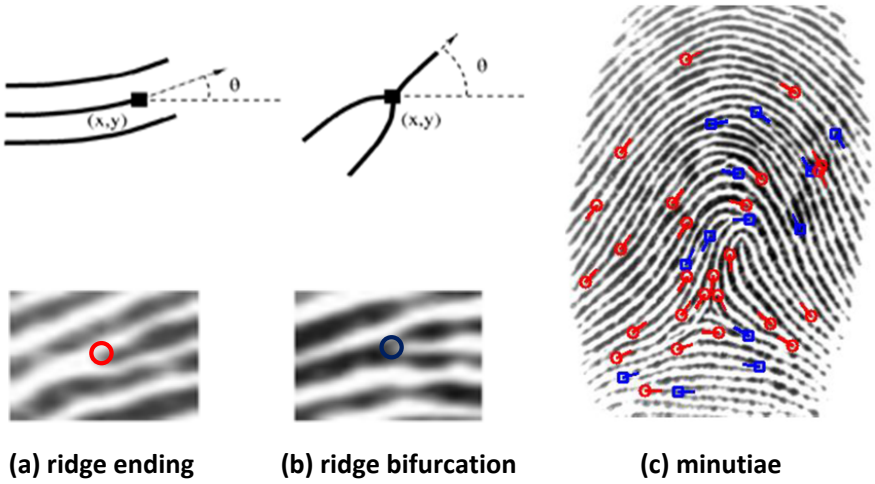


Figure 2.3: Minutiae illustration: (a) ridge ending. (b) ridge bifurcation. (c) a fingerprint example with minutiae marked (red color means ridge ending and blue color means ridge bifurcation). Image Credit: online resource.

Level-3 features of fingerprint focus on very-fine level, that is, intra-ridge details to be detected. Intra-ridge details include width, curvature, shape, ridge contours, and some permanent details such as tiny sweat pores and incipient ridges. However, level-3 features are not practically used in current AFIS since the requirement of good quality fingerprint images with high-resolution in order to extract these fine-level features.

2.1.2 Fingerprint Types

There are three types of fingerprint images as shown in Fig. 2.4: rolled, plain and latent [52]. In order to capture the whole ridge details of a fingertip, rolled fingerprint images are obtained by rolling a finger from one side to the other; plain fingerprins images are impressions pressed down on flat surfaces without rolling; different from rolled and plain fingerprints collected from either impression with inking or scanned using live-scan devices, latent fingerprints are left unintentionally on any surface of objects. Due to invisibility, latent fingerprint images need to be obtained through some physical or chemical processing techniques.

(a) Rolled (b) Plain (c) Latent

Figure 2.4: Three types of fingerprint images: rolled, plain and latent fingerprint images. Image Credit: online resource.

2.1.3 Challenges in Latent Fingerprint Processing

Latent fingerprints are usually captured from crime scenes and they are used as crucial evidence in supporting forensic identification for decades. Due to complexity of crime scenes when latent fingerprints are obtained, quality of these images are very poor in general. Several challenges exist in latent fingerprints including unclear ridge structure, partial fingerprint, overlapping fingerprints, and structured noises (lines, speckles, stains, letters, and etc.). Fig. 2.5 shows several examples to illustrate those challenges mentioned above.

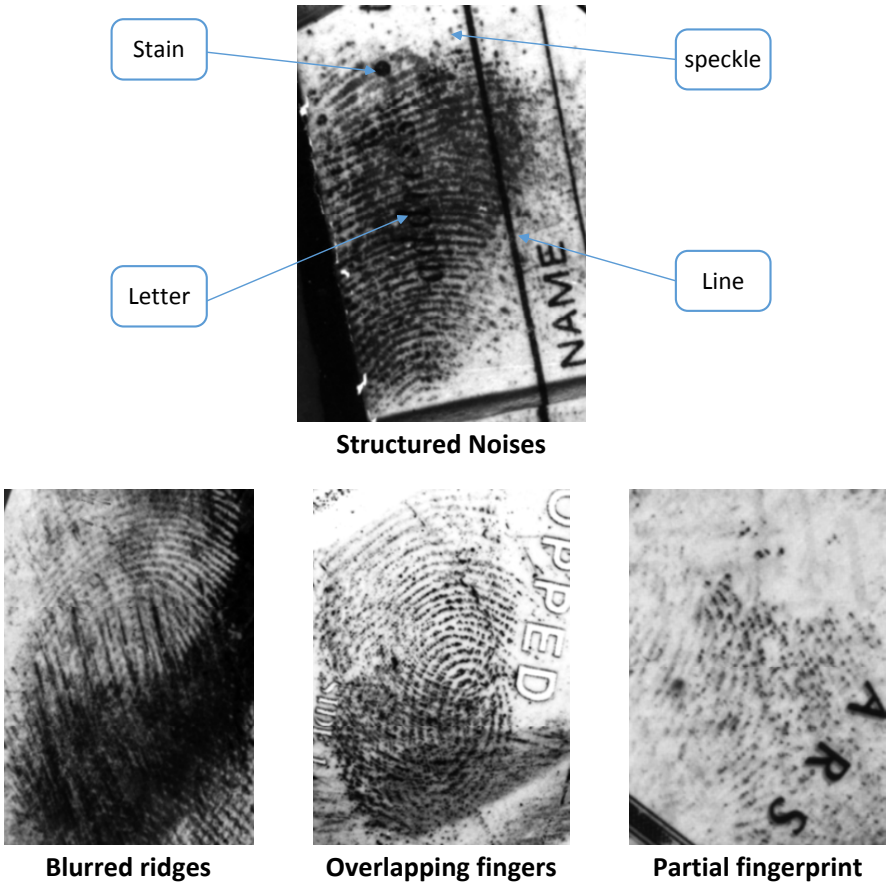


Figure 2.5: Examples of several challenges existed in latent fingerprint.

Unclear ridge structure includes blurred ridges when smearing or ridge breaks because of dry fingers; partial fingerprints and overlapping fingerprints exist a lot since those fingerprints are inadvertently left on complicated object surfaces; structured noises often occur especially in latent fingerprint. Some examples of structured noise can be seen in Fig. 2.5. *"Line"* noise can appear in the format of a single line or even multiple lines (both paralleling and crossing pattern). *"Stain"* noise happens if the fingertip is not pressed properly, especially on a wet or dirty surface. Usually, stains appear with spongy shapes. *"Speckle"* noise has random or regular tiny structures, in which similar scales are shared with fingerprints. *"Letter"* noise is another common structured noise in latent fingerprints. Letters can be handwritten, printed, and have varieties of font

types. Structure noises are very easy to be confused with fingerprint signal because of they share resemblance usually.

These challenges mentioned above make latent fingerprint quality very poor so that feature such as minutiae for latent fingerprints can not be extracted stably and accurately. This further results in a very low identification rate for latent fingerprint in AFIS. Therefore, an effective latent fingerprint enhancement algorithm is very necessary and important to improve feature extraction.

2.2 Total Variation Models for Fingerprint Enhancement

Prior to latent fingerprint enhancement, several pre-processing techniques may help. One of the most useful pre-processing currently is the total variation (TV) decomposition. TV decomposition has been used for image applications such as denoising, deblurring and inpainting [9, 15, 114]. TV decomposition has the potential to reduce structured noise while preserving fingerprint itself. By decomposing fingerprint image into texture and cartoon components, most of the irrelevant contents and structured noises will be kept into the cartoon component, that is, excluded from the texture component. Moreover, TV decomposition has been proved to offer a successful pre-processing tool for latent fingerprint enhancement [12, 67, 11]. Among variations of TV models, TV with L1 fidelity regularization (TV-L1) and adaptive directional TV (ADTV) models are the most used versions for latent fingerprint pre-processing.

2.2.1 TV-L1 Model

TV-L1 model is suitable for multi-scale image decompostion and feature selection, such as facial recognition with illumination variance [20]. Similarly, TV-L1 model decomposes an input latent fingerprint image $f(x)$, into two components:

$$f(x) = u(x) + v(x), \tag{2.1}$$

where $u(x)$ and $v(x)$ are its cartoon and texture components, respectively. The cartoon component consists of piecewise-smooth background noise while the texture component keeps oscillatory and textured patterns such as fingerprint ridges.

This decomposed cartoon component u^* can be achieved by solving the optimization problem in Eq. 2.2:

$$u^* = \arg\min_u \int |\bigtriangledown u| \, dx + \lambda \int |u - f| \, dx, \tag{2.2}$$

where, f, u, v are image intensity about location x. Then, $v^*(x) = f(x) - u^*(x)$ is the decomposed texture component. In Eq. 2.2, the first and the second integrations represent the total variation term and the fidelity term, respectively. The total variation term and the fidelity term are nonlinear and nondifferentiable. Therefore, a steady approximated solution of Eq. 2.2 can be achieved by solving the Euler-Lagrange equation about the optimization problem [17]:

$$\bigtriangledown \cdot (\frac{\bigtriangledown u}{|\bigtriangledown u|}) + \lambda \frac{f - u}{|f - u|} = 0. \tag{2.3}$$

Certainly, several numerical methods are proposed to improve the speed of gradient descent such as the split Bregman iteration, alternation direction method of multipliers (ADMM), the Douglas-Rachford spliting, and so on [110, 34, 92].

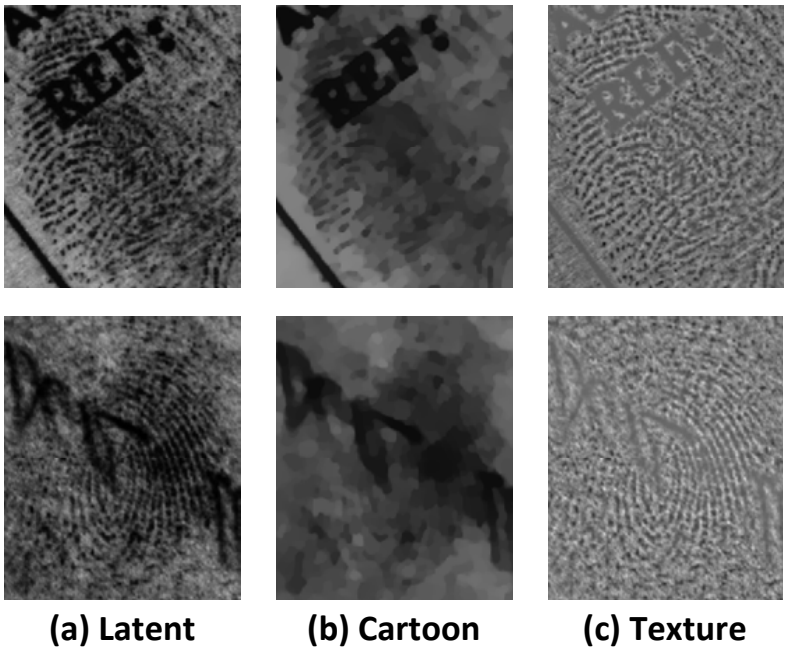


Figure 2.6: TV-L1 decomposition examples: (a) Latent fingerprint images, (b) Cartoon component, (c) Texture component.

By applying TV-L1 model with proper λ value, the decomposed texture layer can extract fingerprint in a certain degree while leaving some unwanted structured noise in the cartoon component. Fig. 2.6 shows examples of the TV-L1 decomposition, where the printed and handwritten letters are mainly kept in the cartoon components and fingerprints are remained into the texture components.

However, due to the fixing λ value, there are two major problems for using TV-L1 model directly. Firstly, some structured noises with small scales (comparable with fingerprint) will still be kept in the texture component. Secondly, using finite differentiating on the non-smoothness boundary results in boundary signals near non-smooth edges.

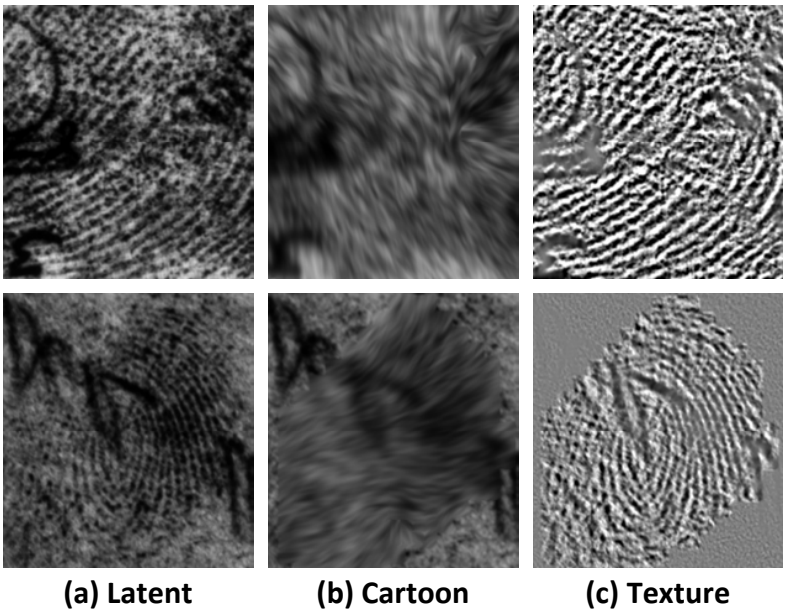


Figure 2.7: ADTV decomposition examples: (a) Latent fingerprint images, (b) Cartoon component, (c) Texture component.

2.2.2 Adaptive Directional TV Model

In order to overcome those limitations mentioned above, Zhang *et al.* [119] proposed an Adaptive Directional Total-Variation (ADTV) method for latent fingerprint decomposition, which provides a better decomposition result over other TV models in general.

In [119], the cartoon component, $u^*(x)$, is obtained by optimizing the following variation problem via modifying Eq. 2.2 to Eq. 2.4:

$$u^* = \arg\min_u \int |\bigtriangledown u \cdot \vec{a}(x)|\, dx + \frac{1}{2} \int \lambda(x)\, |u - f|\, dx, \tag{2.4}$$

and $v^*(x) = f(x) - u^*(x)$, where $\int |\bigtriangledown u \cdot \vec{a}(x)|\, dx$ and $\frac{1}{2} \int \lambda(x)\, |u - f|\, dx$ represent the new total variation term and the fidelity terms. $\vec{a}(x)$ and $\lambda(x)$ denote spatial-dependent orientation and scale parameters, respectively. Parameter $\vec{a}(x)$ gives the direction for texture suppression; namely, the variation is suppressed along the direction of $\vec{a}(x)$ but allowed along its perpendicular direction. Parameter $\lambda(x)$ controls the textured amount

at location x. It is set to a large (or small) value in the structured-noise-dominant region (or the fingerprint-ridge-dominant region) to separate the cartoon component from the texture component. Note that, when $\vec{a}(x) = (1,1)^T$ and $\lambda(x)$ is set to a spatially-independent constant λ, the ADTV model degenerates to the traditional TV-L1 model with fidelity regularization in the $L1$-norm, which is denoted by TV-L1. For more details on the proper choice of parameters $\vec{a}(x)$ and $\lambda(x)$, please refer to [119].

Fig. 2.7 shows some examples from the ADTV decomposition. It can be seen that more fingerprint signal can be kept in the texture component and more structured noise will be in the cartoon layer, which is consistent with its advantage that ADTV model is anisotropic and spatial adaptive in scale.

2.3 Other Fingerprint Enhancement Techniques

Performance of AFIS relies a lot on fingerprint feature extraction, especially minutiae extraction. For an ideal fingerprint image, ridges and valleys alternate and flow in a locally constant direction. With this well-defined case, fingerprint ridges are able to be detected easily and the minutiae can also be extracted precisely. However, fingerprint images in practical are corrupted by noise because of the variations in environment conditions such as wet or dry fingertip skin, senor noise, inappropriate finger pressure and so on. For latent fingerprint, the image quality is particular poor as we introduced before.

To ensure a feasible automatic algorithm of feature extraction such as minutiae, latent fingerprint enhancement is the key step before extracting features. Directly using conventional automatic minutiae extraction techniques for latent fingerprint will result in a large amount of false positives and false negatives, which furthermore decrease identification rate of AFIS significantly. However, fingerprint enhancement as pre-processing,

the goal of which is to enhance contrast between ridges and valleys, connect breaking ridges, separate joined ridges, and remove varieties of structured noise, can make sure efficiency and robustness of automatic minutiae extraction algorithm [36].

2.3.1 Contextual Filtering

General image enhancement techniques often are not proper for latent fingerprint enhancement. A direct applying of general-purpose image enhancement is hard to provide satisfying results for latent fingerprint. The reason is obvious, that is, the goal of fingerprint enhancement is different with general image enhancement. Reducing structured noise and enhance the definition of ridges and valleys are the extreme goal of fingerprint enhancement. In the other hand, "noise" signal in fingerprint sometimes is as strong as the fingerprint signal. This problem is much more serious about latent fingerprints, in which structured noises are even much stronger than real fingerprint.

The most widely used fingerprint enhancement technique is based on contextual filtering. In contextual filtering, the filter characteristics change according to the local context, unlike the traditional image filtering with a single filter. Generally, a set of filters is pre-computed and one of them is selected for each image patch. The context for fingerprint is usually defined by the local ridge characteristics(local orientation and frequency). Actually, ridges and valleys form a sinusoidal-shaped wave that is mainly defined by a local orientation and frequency. Therefore, it is natural to design filter banks that are tuned to the local ridge frequency and orientation. This can efficiently remove the undesired noise and preserve the true ridge and valley structure.

Several techniques of contextual filtering have been proposed in literate such as [2, 22, 38, 46, 86, 94, 112]. Among these techniques, Gabor filtering proposed by Hong *et al.* is the most popular one used nowadays [46]. Gabor filters have optimal joint resolution in both spatial and frequency domains since their frequency-selective and

orientation-selective properties. Gabor filters in the spatial domain can be defined as Eq. 2.5:

$$g(m, n, \theta, w) = exp\{-(\frac{m_\theta^2}{2\delta_m^2} + \frac{n_\theta^2}{2\delta_n^2})\} \cos(2\pi w m_\theta), \tag{2.5}$$

$$m_\theta = m\cos(\theta) + n\sin(\theta), \tag{2.6}$$

$$n_\theta = -m\sin(\theta) + n\cos(\theta), \tag{2.7}$$

where m, n are 2D coordinates in conventional image coordinate system; θ is the perpendicular orientation to the parallel stripes of a Gabor function; w represents the wavelength of a sinusoidal plane wave; δ_m and δ_n are spatial constants of the Gaussian envelope along the corresponding axes; and m_θ and n_θ denote the axes of the filter coordinate frame, respectively.

With Gabor filters, a fingerprint image is filtered using the corresponding filter for each image pixel (m, n) to get the enhanced image via

$$I_E(m, n) = \sum_{i=-\frac{s_x}{2}}^{\frac{s_x}{2}} \sum_{j=-\frac{s_y}{2}}^{\frac{s_y}{2}} g(i, j, O(m, n), w(m, n)) I(m - i, n - j), \tag{2.8}$$

where I denotes the fingerprint image, O is the estimated orientation field, w is the ridge wavelength map obtained from frequency field, and s_x and s_y are the size of the Gabor filter mask. An illustration about the Gabor filters and fingerprint enhancement are shown in Fig. 2.8.

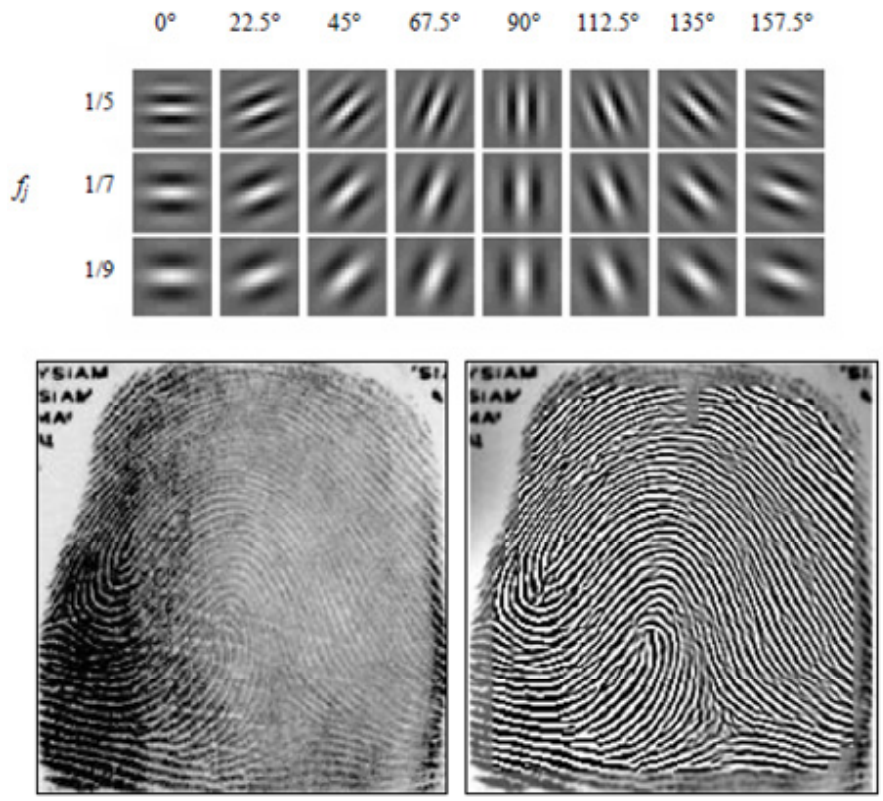


Figure 2.8: An illustration from [78] about Gabor filters and the enhanced fingerprint by Hong method.

2.3.2 Orientation Field Estimation

Since the success of contextual filtering, orientation field estimation as the key factor of fingerprint enhancement tends to attract researchers' great interest so much for the last two decades [22, 46, 47, 71, 59, 99, 100, 105, 121]. Due to the space limit, it is not possible to review too many approaches in this proposal. Only several classical and recent algorithms are reviewed in this subsection.

• Orientation Field Local Estimation

Orientation at a pixel of fingerprint is defined as the angle between the ridge inclination and the horizontal line. Usually, the orientation field consists of local orientations of on-overlapping blocks (such as 16×16 block size). Fig. 2.9 is an illustration about

Figure 2.9: Definition illustration about orientation at fingerprint pixel and an orientation field estimated for an exemplary fingerprint. Blue lines means the estimated orientation in the local block. Image Credit: online resource.

local orientation of fingerprint and an orientation field map estimated for an exemplary fingerprint.

Gradient based Method

The gradient based method was proposed by Hong *et al.* in [46]. A Fingerprint I is firslty normalized to I_N with the fixed mean μ_0 and variance σ_o^2 as follows:

$$I_N(m,n) = \begin{cases} \mu_0 + \sqrt{\frac{\sigma_0^2(I(m,n)-\mu)^2}{\sigma^2}} & \text{,if } I(m,n) > \mu \\ \mu_0 - \sqrt{\frac{\sigma_0^2(I(m,n)-\mu)^2}{\sigma^2}} & \text{,otherwise} \end{cases} \tag{2.9}$$

where μ and σ^2 are the estimated mean and variance. Although this pixel operation does not clear the ridge structure, the variations along ridges and valleys are reduced.

Then, the normalized fingerprint I_N is divided into blocks of $W \times W(16 \times 16)$ and the gradients are calculated as follows:

$$G_x(m,n) = \frac{\partial I_N(m,n)}{\partial m}, \tag{2.10}$$

$$G_y(m,n) = \frac{\partial I_N(m,n)}{\partial n}. \tag{2.11}$$

Finally, the orientation field O is obtained by the following equations:

$$G_{xx}(m,n) = \sum_{i=m-\frac{W}{2}}^{m+\frac{W}{2}} \sum_{j=m-\frac{W}{2}}^{m+\frac{W}{2}} \{G_x^2(i,j) - G_y^2(i,j)\}, \tag{2.12}$$

$$G_{yy}(m,n) = \sum_{i=m-\frac{W}{2}}^{m+\frac{W}{2}} \sum_{j=m-\frac{W}{2}}^{m+\frac{W}{2}} 2G_x(i,j)G_y(i,j)\}, \tag{2.13}$$

$$\theta(m,n) = \frac{1}{2}\tan^{-1}\{\frac{G_{yy}(m,n)}{G_{xx}(m,n)}\}, \tag{2.14}$$

$$O(m,n) = \frac{1}{2}\tan^{-1}\{\frac{Gau(i,j) * \sin(2\theta(m,n))}{Gau(i,j) * \cos(2\theta(m,n))}\}, \tag{2.15}$$

where G_{xx} and G_{yy} are the average squared gradients in a $W \times W$ window, and Gau represents a smoothing Gaussian kernel to smooth the estimated orientation field.

Short-time Fourier Transformation (STFT) based method

Another popular method to estimate the orientation filed is called STFT, which views the magnitude spectrum of the Fourier transform of a local fingerprint in polar coordinate system as a probability distribution. The best local orientation can be estimated as

the most probable orientation or the mean in the local block. This method was proposed by Chikkerur *et al.* in 2007 [22].

In this method, a local region of the fingerprint image is modeled as a surface wave according to Eq. 2.16 and then the 2D STFT is applied to the region as Eq. 2.17,

$$I(x,y) = A\{\cos(2\pi f(x\cos\theta + y\sin\theta))\}, \tag{2.16}$$

$$X(\tau_1,\tau_2,\omega_1,\omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)W_{2d}^{*}(x-\tau_1,y-\tau_2)e^{-j(\omega_1 x+\omega_2 y)}dxdy, \tag{2.17}$$

where τ_1, τ_2 are the spatial position of a 2-dimension window $W_{2d}(x,y)$, and ω_1, ω_2 represent the spatial frequency parameters.

Then, the 2D Fourier spectrum which is only about frequency and orientation parameters can be expressed in the polar form as $F(r,\theta)$, where r, θ represent the frequency and orientation, respectively. A probability density function $p(r,\theta)$ and its marginal density function $p(\theta)$ can be defined as follows:

$$p(r,\theta) = \frac{|F(r,\theta)|^2}{\int_r\int_\theta |F(r,\theta)|^2}, \tag{2.18}$$

$$p(\theta) = \int_r p(r,\theta)dr. \tag{2.19}$$

Finally, the expected value of the orientation can be obtained as Eq. 2.20 to express the local orientation O:

$$O = E\{\theta\} = \frac{1}{2}\tan^{-1}\{\frac{\int_\theta p(\theta)\sin(2\theta)d\theta}{\int_\theta p(\theta)\cos(2\theta)d\theta}\}. \tag{2.20}$$

• Orientation Field Smoothing

Obviously, the local orientation estimation approaches mentioned above are sensitive to noise, especially unreliable when structured noise such as lines, printed letters and handwriting appear. To handle this, several smoothing methods were proposed as post-processing using regularization to add more constraints on estimated orientation. Local orientation smoothing approaches include low-pass filtering, multi-resolution smoothing, and energy minimization using Markov Random Field (MRF) [7, 10, 18, 27, 57, 68, 73, 87, 90].

Although local smoothing provides improved performance, it lacks the global constraints. For example, a MRF model with small contextual information can only have limited prior knowledge about fingerprint ridge structure [8, 65]. To model entire orientation field globally, researches proposed global parametric models such as polynomial model and the Fingerprint Orientation Model Based on 2D Fourier Expansion (FOMFE) method [41, 40, 105, 122, 123]. FOMEF method is used here for an example to show the orientation smoothing.

FOMFE [105] treats the whole orientation field as 2D matrix. Each element θ in the orientation field can be expressed as $\cos(2\theta)$ and $\sin(2\theta)$, that means the original OF matrix $\mathbf{O}$ can be decomposed into one "Cos" matrix $\mathbf{V_c} = \cos(2\mathbf{O})$ and one "Sin" matrix $\mathbf{V_s} = \sin(2\mathbf{O})$. 2D Fourier Expansion is conducted on $\mathbf{V_c}$ and $\mathbf{V_s}$ separately and the Fourier coefficients of the basis functions are estimated using initial orientation field. Then, those estimated coefficients are used to reconstruct a new orientation filed.

2D function $f(x, y)$ can be expressed using a set of basis functions in the following general form as Eq. 2.21 showing:

$$f(x, y) = \sum_{i=0}^{k} \beta_i \Psi_i(x, y) + \varepsilon(x, y), \tag{2.21}$$

where $k \in N$ is the order, $\varepsilon(x, y)$ is the residual, and $\{\Psi_i\}$ are the basis functions.

Since orientations behave in a periodic manner, it is reasonable to use a series of cosine and sine function to take place the basis functions $\{\Psi_i\}$ in Eq. 2.21. Eq. 2.21 will become the 2D Fourier expansion as shown in Eq. 2.22:

$$f(x, y) = \sum_{i=0}^{k} \sum_{j=0}^{k} \Psi(i\nu x, j\omega y; \beta_{ij}) + \varepsilon(x, y), \tag{2.22}$$

$$\begin{aligned} \Psi(i\nu x, j\omega w; \beta_{ij}) = \lambda_{ij}[a_{ij} \cos(i\nu x) \cos(j\omega y) + b_{ij} \sin(i\nu x) \cos(j\omega y) \\ + c_{ij} \cos(i\nu x) \sin(j\omega y) + d_{ij} \sin(i\nu x) \sin(j\omega y)], \end{aligned} \tag{2.23}$$

where $-l \leq x \leq l, -h \leq y \leq h$, $i, j \in N$, $\nu = \frac{\pi}{l}$ and $\omega = \frac{\pi}{h}$ are the fundamental frequencies, λ_{ij} is a constant scalar, and $\{a_{ij}, b_{ij}, c_{ij}, d_{ij}\}$ are the Fouier coefficients to be estimated.

Then, the two decomposed orientation matrix $\mathbf{V_c}$ and $\mathbf{V_s}$ can be expressed based on the 2D Fourier expansions using Eq. 2.22, respectively. A more compact matrix expression can be written as:

$$\mathbf{V_c} = \mathbf{P}(\mathbf{X}) \cdot \mathbf{B_c} \cdot \mathbf{Q^T}(\mathbf{Y}), \tag{2.24}$$

$$\mathbf{V_s} = \mathbf{P}(\mathbf{X}) \cdot \mathbf{B_s} \cdot \mathbf{Q^T}(\mathbf{Y}), \tag{2.25}$$

where $\mathbf{P}, \mathbf{Q}$ are used to form compact matrix expression, $\mathbf{B_c}$ and $\mathbf{B_c}$ are the parameter matrices that consist of the Fourier coefficients to be estimated. The parameter matrices can be obtained by the various optimization techniques to solve this classical linear least square problem.

The gradient based OF can be estimated initially to be $\mathbf{V_c}$ and $\mathbf{V_s}$. With the optimized solution of $\mathbf{B_c}$ and $\mathbf{B_c}$, the smoothed OF can be reconstructed using these two parameter matrices.

Chapter 3

Latent Fingerprint Enhancement Using Markov Random Field and Sparse Representation (MRF-SR) Method

3.1 Introduction

Although dictionary-based methods (*e.g.* the orientation dictionary [36, 113] and the ridge dictionary [12, 67]) has improved the performance of latent fingerprint enhancement, they are still far from satisfaction in challenging cases.

The methods based on the orientation dictionary only [36, 113] do not exploit the ridge information fully. That is, the ridge structure is only used for initial OF estimation. Afterwards, these methods only involve the manipulation and/or smoothing of the initial OF. Without the guidance of the fingerprint ridge knowledge, the estimated OF is not accurate. Furthermore, Gabor filtering with a fixed ridge frequency cannot offer good enhancement results in heavily corrupted ridge regions. The methods based on the ridge dictionary only [12, 67] attempted to resolve the above-mentioned issue by constructing a dictionary with atoms formed by ridge patterns. However, they have the following two drawbacks.

First, the traditional sparse representation is not suitable for fingerprint ridge images especially in regions with heavy noise. One example is illustrated in Fig. 3.2. The results of applying the traditional sparse representation to an exemplary fingerprint ridge image

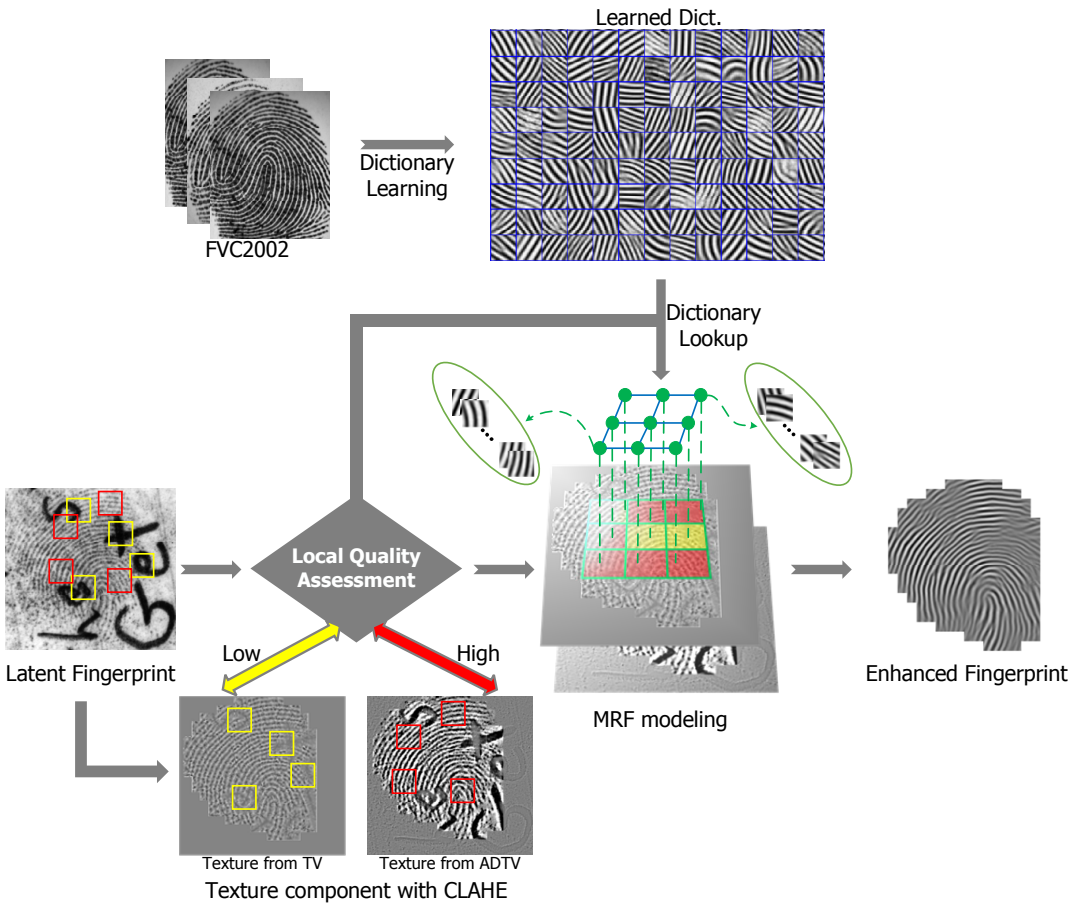


Figure 3.1: The flowchart of the proposed MRF-SR latent fingerprint enhancement method.

are shown in Fig. 3.2, where we focus on a texture patch of size 32×32 within a red square, which is corrupted by a handwritten letter "X". This patch is first enhanced by the total variation model. Then, it is coded by the orthogonal matching pursuit (OMP) sparse coding with trained and Gabor dictionaries [12, 67]. The top 10 corresponding atoms whose linear combination give its closest approximation are shown from left to right in order. The expected enhancement output is a patch consisting of four horizontal ridges, yet a patch of this type ranks #7 and #9 in the trained and Gabor dictionaries, respectively. The desired patch is unlikely to be selected in a practical setting.

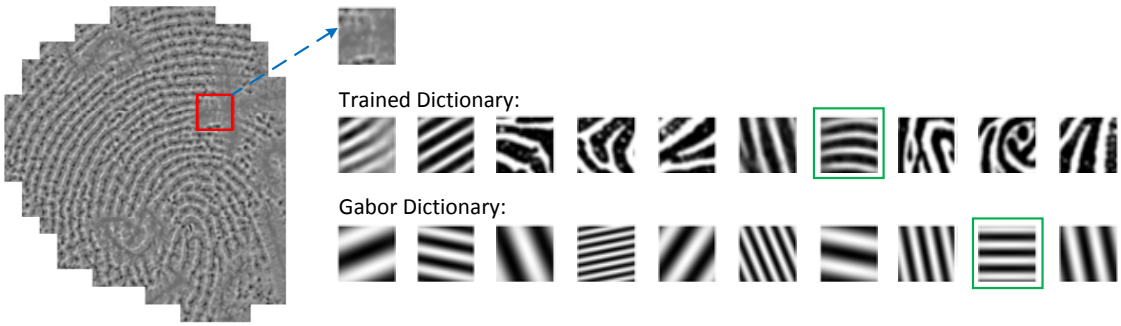


Figure 3.2: A texture patch of size 32×32 corrupted by structured noise is shown within a red square and the top 10 atoms, whose linear combination give its closest approximation using the traditional sparse coding (OMP) with both trained and Gabor dictionaries, are shown in order from left to right.

Second, their result depends on the initial reconstructed result based on sparse coding. As discussed in [12], when the coarse-level reconstruction result is wrong, it cannot be further corrected since the fine-level dictionary is consistent with its coarse-level dictionary. Similarly, as presented in [67], wrongly reconstructed regions of high coherence cannot be fixed in further iterations since only regions of low coherence are rectified. A better scheme that can correct inaccurate initial estimation is much in need.

3.2 System Overview

To solve the above-mentioned problems, we propose a local enhancement method by formulating the sparse representation in the MRF framework. For MRF modeling, we define the unary (or vertex) potential based on the SSIM measure [106] and the interaction (or edge) potential based on the combination of orientation and frequency. Due to this new mathematical treatment, the loopy belief propagation process can select a proper candidate and correct the wrong selection dynamically in the following iteration.

The flowchart of the proposed MRF-SR method is given in Fig. 3.1. With a given ridge dictionary, it consists of the following three major modules.

1. Decompose a latent fingerprint image into its cartoon and texture components using the TV and ADTV decomposition [119] and, then, apply the contrast limited adaptive histogram equalization (CLAHE) to the texture component.

2. Local quality of the latent fingerprints is assessed by measuring the SSIM between the fingerprint and its reconstruction via sparse representation. Then, the texture component from ADTV and TV is extracted for further processing in good and poor quality regions, respectively.

3. Extract overlapping patches from corresponding texture image based on local quality and use the ridge dictionary to generate a candidates pool for each patch. A ridge dictionary can be constructed in two ways – (1) learned from high quality fingerprint images and (2) created by the Gabor function. Both two dictionaries are studied thoroughly in this work.

4. Set up the MRF model for latent fingerprints by treating each patch as a random variable. The objective is to minimize the energy cost function that consists of a data fidelity term and regularization terms so as to determine the most suitable candidate for each patch. Afterwards, we perform simple quilting of all enhanced patches to yield the final enhanced fingerprint image.

The details of each of the above steps will be given in the next section.

3.3 Proposed MRF-SR Method

3.3.1 Pre-Processing of Latent Fingerprints

TV Model

The total variation (TV) decomposition has been used for image applications such as denoising, deblurring and inpainting. Moreover, it offers a successful pre-processing tool for latent fingerprint enhancement [12, 67, 11].

Mathematically, a latent fingerprint, denoted by $f(x)$, can be written as

$$f(x) = u(x) + v(x), \tag{3.1}$$

where $u(x)$ and $v(x)$ are its cartoon and texture components, respectively. The cartoon component consists of piecewise-smooth background noise while the texture component keeps oscillatory and textured patterns such as fingerprint ridges. The decomposed cartoon component u^* can be achieved by solving the following optimization problem:

$$u^*(x) = \arg\min_u \{ \int |\bigtriangledown u(x)|\, dx + \lambda \int |f(x) - u(x)|\, dx \}, \tag{3.2}$$

where, $f(x), u(x), v(x)$ are image intensity at location x, $\bigtriangledown$ represents the gradient operation, the first and the second integrations in Eq. (3.2) represent the total variation term and the fidelity term, and λ is the weight for fidelity term.

Generally speaking, the above optimization procedure maintains u a smooth function using the total variation term and ensures u to be close to f using the data fidelity term. Then, $v^*(x) = f(x) - u^*(x)$ is the decomposed texture component, which is most likely composed of the fingerprint. The parameter, λ, is highly correlated with the scale of the decomposed texture features [17, 115]. By applying the TV model with a

proper λ value, one can separate the fingerprint in the texture component, $v^*(x)$, from unwanted structured noise in the cartoon component, $u^*(x)$.

ADTV Model

Zhang *et al.* [119] proposed an Adaptive Directional Total-Variation (ADTV) method for latent fingerprint decomposition, which is anisotropic and spatial-adaptive in scale. In [119], the cartoon component, $u^*(x)$, is obtained by optimizing the following new variation problem:

$$\begin{aligned} u^*(x) = \arg\min_u \{ & \int |\bigtriangledown u(x) \cdot \vec{a}(x)| \, dx \\ & + \frac{1}{2} \int \lambda(x) \, |u(x) - f(x)| \, dx \}, \end{aligned} \tag{3.3}$$

where $\int |\bigtriangledown u(x) \cdot \vec{a}(x)| \, dx$ and $\frac{1}{2} \int \lambda(x) \, |u(x) - f(x)| \, dx$ represent the new total variation and data fidelity terms, and $\vec{a}(x)$ and $\lambda(x)$ denote spatial-dependent orientation and scale parameters at location x, respectively. Parameter $\vec{a}(x)$ gives the direction for texture suppression, namely, the variation is suppressed along the direction of $\vec{a}(x)$ but allowed along its perpendicular direction. Parameter $\lambda(x)$ controls the scale of features extracted or amount of texture component at pixel x. It is set to a large (or small) value in the structured-noise-dominant region (or the fingerprint-ridge-dominant region) to separate the cartoon component from the texture component. An example of latent fingerprint decomposition using ADTV is shown in Fig. 3.3.

Note that, when $\vec{a}(x) = (1, 1)^T$ and $\lambda(x)$ is set to a spatially-independent constant λ, the ADTV model degenerates to the traditional TV model with fidelity regularization in the $L1$-norm. For more details on the proper choice of parameters $\vec{a}(x)$ and $\lambda(x)$, we refer to [119].

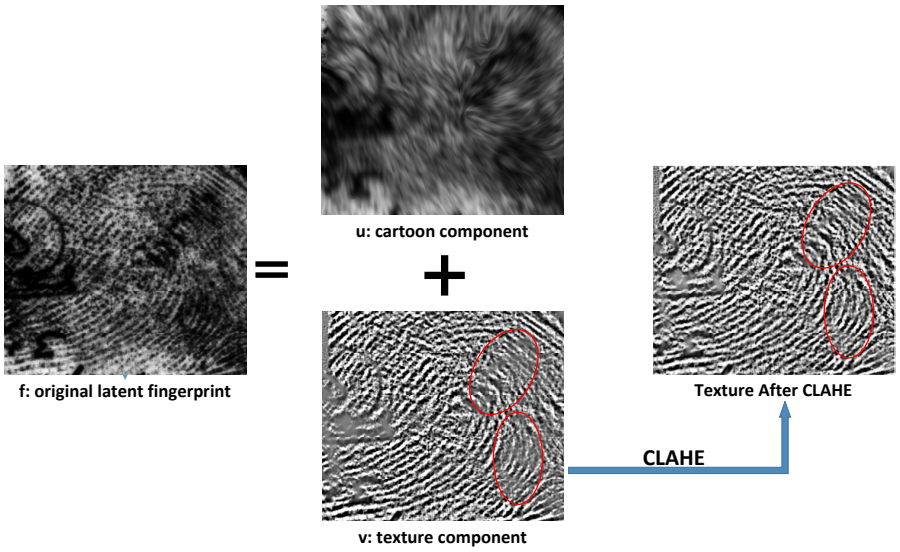
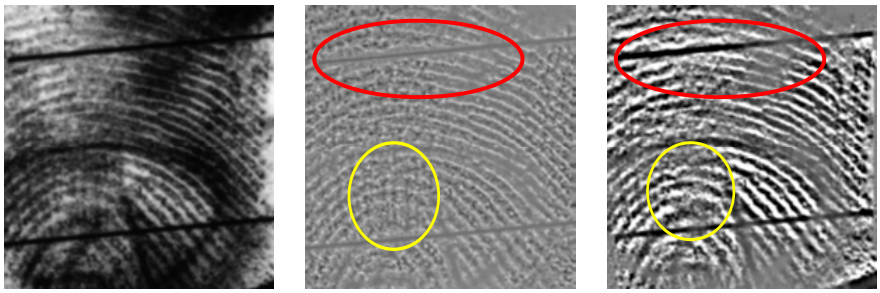


Figure 3.3: Illustration of decomposing a latent fingerprint image, f, into a cartoon image, u, and a texture image v, by ADTV. We see that the structured noise and the fingerprint ridge pattern are split into the cartoon and the texture components, respectively. Then, the CLAHE technique is applied to the texture component for further enhancement.

TV or ADTV

Both TV and ADTV models have been applied to latent fingerprint enhancement successfully. However, it is not always true that ADTV performs better than TV, and vice verse. The adaptive parameters, $\vec{a}(x)$ and $\lambda(x)$, work well in fingerprint regions of good quality since these two parameters can be estimated accurately. However, ADTV fails in fingerprint regions of poor quality where those two adaptive parameters are wrongly estimated. For example, structured noise such as curves and letters will result in wrong estimation of $\vec{a}(x)$ that is related to orientation estimation. Furthermore, structured noise with a scale similar to that of fingerprint ridges will be kept in the texture component. In both cases, the TV model is more robust against inaccurate guidance of adaptive parameters in the ADTV model. An example is illustrated in Fig. 3.4. The red and yellow circles in Fig. 3.4 show poor- and good-quality fingerprint regions, respectively. We see that the TV model performs better in poor quality fingerprint region but worse in the



(a) Original (b)Texture from TV (c)Texture from ADTV

Figure 3.4: Comparison of the texture component selected using the TV model and the ADTV model, where the red and the yellow circles show poor- and good-quality fingerprint regions, respectively.

good quality fingerprint region. We will present a method that selects the TV model or the ADTV model adaptively based on the local fingerprint quality in Sec. 3.3.3.

CLAHE

To further enhance the contrast of the fingerprint ridge pattern in the texture component, we adopt the contrast limited adaptive histogram equalization (CLAHE) technique [89]. It is a useful tool for two reasons. First, the low contrast region can be enhanced by adaptive histogram equalization. Second, it can prevent noise over-amplification in a local neighborhood using a contrast limiting procedure. An example of latent fingerprint decomposition and enhancement using ADTV and CLAHE is shown in Fig. 3.3. The contrast has been improved as indicated in the two red-circled regions in Fig. 3.3.

To summarize, the fingerprint ridge pattern in the texture component becomes cleaner and stronger by applying these pre-processing techniques. Afterwards, the ridge dictionary can be more effectively applied to these cleaned images and, for any given patch, a pool of candidates can be selected for further optimization using MRF. We will discuss dictionary construction in Sec. 3.3.2, and candidates selection and optimization via MRF in Sec. 3.3.4.

3.3.2 Ridge Dictionary Construction

A fingerprint patch can be further enhanced by representing it using elements from a well-designed ridge dictionary, known as the sparse representation [33, 64, 66, 75, 76, 77, 91, 103, 109]. Two ridge dictionaries are examined in this work; namely, a dictionary learned from exemplar fingerprint patches and the Gabor dictionary.

Learned Dictionary

We use a set of high quality patches extracted from exemplar fingerprints in the FVC2002 dataset as the training samples. Besides, we collect training patches from the center of each fingerprint in order to include singular points such as the core as much as possible. Each patch of size $s \times s$ is normalized to be with zero mean and unit l^2-norm. All N training samples can be denoted by a matrix $X = (x_1, x_2, ..., x_N)$ of dimension $s^2 \times N$, where each column is one training sample x_n, $1 \leq n \leq N$.

Sparse dictionary learning is usually formulated as an optimization problem in form of:

$$\min_{D,A} \|X - DA\|_F^2, \quad \text{s.t.} \, \|\alpha_i\|_0^0 \leq L, \forall i, \tag{3.4}$$

where $D = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_M)$ is a dictionary of dimension $s^2 \times K$ with $\mathbf{d}_k$ being an atom, $A = (\alpha_1, \alpha_2, ..., \alpha_N)$ is a $K \times N$ coefficient matrix with α_n being the coefficient of sample x_n projected to D, $\|\cdot\|_0^0$ is the l^0 norm, $\|\cdot\|_F$ is the Frobenius norm, and L is a predetermined number of nonzero entries. This is not a trivial problem since both D and A are unknowns and only training samples X are given. It can however be effectively solved using the K-SVD dictionary learning algorithm [1] through an iterative procedure. The dictionary, D, is often initialized using the Discrete Cosine Transform (DCT) basis to have a good initial point. Then, the K-SVD learning process

iterates between the sparse coding stage and the dictionary update stage until it reaches convergence. These two stages are detailed below.

Sparse coding stage. With a given dictionary D, we apply the Orthogonal Matching Pursuit (OMP) algorithm to find coefficient matrix $\hat{A} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)$, where

$$\hat{\alpha}_n = \arg\min_{\alpha_n} \|D\alpha_n - \mathbf{x}_n\|_2^2, \quad \text{s.t.} \|\alpha_\mathbf{n}\|_0^0 \leq L, \tag{3.5}$$

is the sparse coefficient vector for atom x_n under the constraint that no more than L atoms are selected to represent each sample via linear combination. It is the projection of $\mathbf{x}_n$ onto dictionary D.

Dictionary update stage. After finding coefficient matrix A for given dictionary D, the next step is to update the dictionary. Only one atom is updated at a time. We take the update of atom $\mathbf{d}_k$ as an example. Let $\mathbf{r}_k$ be the set whose elements are indices of samples that use atom $\mathbf{d}_k$ as basis. Mathematically, we have

$$\mathbf{r}_k = \{i \mid \alpha_T^k(i) \neq 0, 1 \leq i \leq N\}, \tag{3.6}$$

where α_T^k is the k^{th} row of A. To update $\mathbf{d}_k$ while preserving sparsity, an overall residual matrix E_k of dimension $s^2 \times N$ without the contribution of atom $\mathbf{d}_k$ can be written as

$$E_k = X - \sum_{j \neq k} \mathbf{d}_j \alpha_T^j, \tag{3.7}$$

and its restricted shrinking form, if only considering samples which use $\mathbf{d}_k$, can be expressed as

$$\tilde{E_k} = \{e_i\}_{i \in \mathbf{r}_k}, \tag{3.8}$$

where $\tilde{E}_k$ is a matrix of dimension $s^2 \times |\mathbf{r}_k|$, $|\mathbf{r}_k|$ is the size of set $\mathbf{r}_k$, and e_i is the i^{th} column of E_k. Then, $\mathbf{d}_k$ and $\tilde{\alpha}_T^k$, which is the restricted shrinking version of $\{\alpha T^k(i)\}_{i \in \mathbf{r}_k}$, are updated by

$$\min_{\mathbf{d}_k, \tilde{\alpha}_T^k} \left\| \tilde{E}_k - \mathbf{d}_k \tilde{\alpha}_T^k \right\|_F^2. \tag{3.9}$$

The singular value decomposition (SVD) decomposition of $\tilde{E}_k$ can be used to solve the optimization problem in Eq. 3.9. Note that only coefficients for samples using $\mathbf{d}_k$ are updated to preserve sparsity. Finally, each atom in the dictionary is normalized again to be with zero mean and unit norm.

Gabor Dictionary

Another ridge dictionary to be used in our system was proposed in [67]. Since local fingerprint ridges can be approximated by 2D sinusoidal waves with specified orientation and frequency parameters, the 2D Gabor functions parameterized by a certain orientation and frequency can be used model a local fingerprint patch. Mathematically, a 2D Gabor function can be expressed as follows [39, 56]

$$g(m, n, \theta, w) = \exp\{-(\frac{m_\theta^2}{2\delta_m^2} + \frac{n_\theta^2}{2\delta_n^2})\} \cos(2\pi w m_\theta + \phi), \tag{3.10}$$

where

$$m_\theta = m\cos(\theta) + n\sin(\theta), \tag{3.11}$$

$$n_\theta = -m\sin(\theta) + n\cos(\theta), \tag{3.12}$$

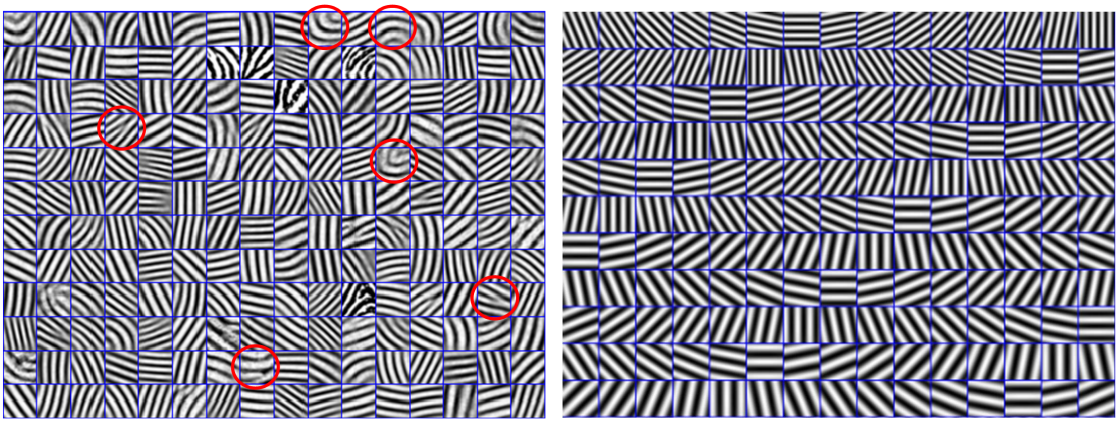


Figure 3.5: Illustration of subsets of (a) the learned dictionary and (b) the Gabor dictionary, where each atom in the dictionaries has a size of 32×32. Ridge patterns with singular points and minutiae in the learned dictionary are circled in red.

where (m, n) are the 2D coordinates in the conventional image coordinate system, θ is the direction perpendicular to the parallel stripes of a Gabor function, w is the wavelength of a sinusoidal plane wave, δ_m and δ_n are spatial constants of the Gaussian envelope along the corresponding axes, and ϕ is the phase offset. We follow the same Gabor function parameter setting in [67]: 16 values for θ, 9 values for w and 6 values for ϕ. Thus, a total of 864 basis atoms is generated to construct the Gabor dictionary for sparse representation. Each basis atom is normalized to be with zero mean and unit norm.

Dictionary Comparison

The above two dictionaries are constructed, respectively, and a subset of them are shown in Fig. 3.5. Note that we cannot show the two entire dictionaries due to the space limitation. We can see their pros and cons from this figure. For the learned dictionary, we can obtain special ridge patterns such as minutiae, singular points and curved ridges, which are circled in red in Fig. 3.5. Although the number of these special ridge patterns is few as compared to ordinary ridge patterns, they are useful in approximating regions in the

neighborhood of singular points. In addition, curve-like ridge patterns are close to actual ridges. On the other hand, due to the limited size of training samples, it is difficult to obtain atoms with rich frequency variations. In contrast, the Gabor dictionary has a rich combination of frequency/orientation patterns due to its parametric form. Nevertheless, ridge patterns in the Gabor dictionary do not look like actual ridges. Besides, they cannot provide a good approximation to the neighborhood of singular points. These two limitations of Gabor dictionary introduce artifacts, feature shifting and over-enhancement. Although the learned and the Gabor dictionaries are studied separately in [12, 67], there is no direct comparison of two dictionaries in the same framework. Therefore, we investigate both the dictionaries in Sec. 5.4 to show their performance.

3.3.3 Local Quality Assessment

As discussed in Sec. 3.3.1, it is desired to select between the TV model and the ADTV model based on local fingerprint image quality. The fingerprint quality map was used to segment background and foreground regions in [12]. By following [12], we assess local fingerprint quality by measuring the structure similarity (SSIM) between a latent fingerprint patch and its reconstruction from sparse representation with a minor difference. That is, a normalized SSIM value was adopted in [12] while the absolute SSIM value is used in our work since it is possible that all local patches are with poor quality.

Recall that the SSIM index [106] is used to measure the similarity of two images (or image patches). It is mathematically defined as

$$SSIM(I_a, I_b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{a,b} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)}, \tag{3.13}$$

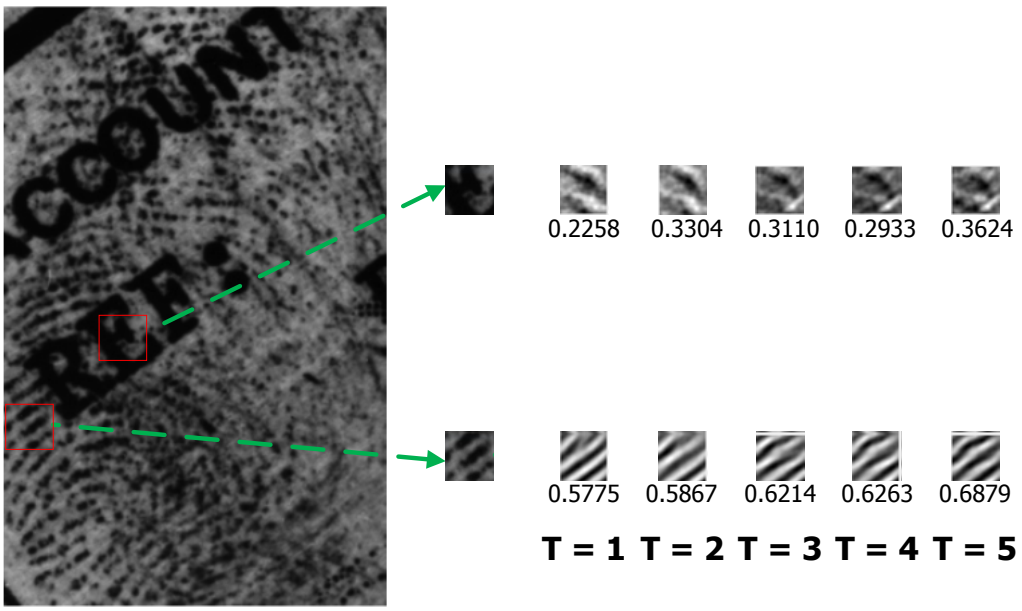


Figure 3.6: Illustration of two fingerprint patches in good and poor quality regions, respectively, where the SSIM values between the original patch and its reconstructed version using the proposed joint sparse representation of sparsity T (i.e., the number of atoms used in reconstruction) are given.

where I_a and I_b are two images (or patches), μ_a and σ_a, μ_b and σ_b are the intensity mean and the standard deviation of I_a and I_b, respectively, $\sigma_{a,b}$ is the covariance between I_a and I_b and C_1 and C_2 are constant parameters to avoid computation instability.

A latent fingerprint that has different quality values in different local regions is shown in Fig. 3.6. In our experiments, we set the quality threshold, T_q, between good and poor quality regions empirically to 0.4. If the SSIM value of a local patch is less than 0.4, the region is viewed as the poor quality region and the TV model is used. Otherwise, it is treated as the good quality region and the ADTV model is applied. In the final processing, two texture components from both TV and ADTV models are generated firstly. Then, based on quality of a local region, the proposed algorithm will select the corresponding region of texture components from TV or ADTV.

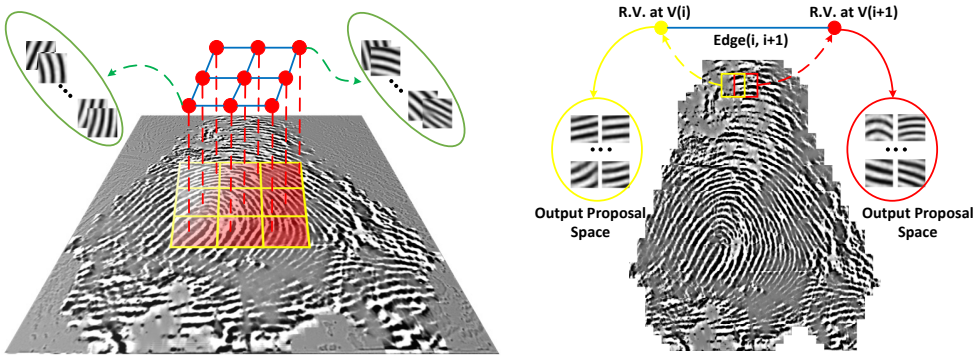


Figure 3.7: MRF modeling for latent fingerprint enhancement with overlapping patches, where each patch is denoted by a red vertex (or node) in a 2D grid and neighboring vertices are connected by a blue edge. The observed data is the texture component of the corresponding patch and its output is selected from a candidate pool defined by the ridge dictionary and circled by green.

3.3.4 MRF-Optimized Sparse Representation

Problem Formulation

Instead of applying sparse coding to each noisy patch directly and independently using the ridge dictionary as the enhancement output, we would like to take the relationship between neighboring patches into account. This can be achieved by casting the latent fingerprint enhancement problem in the MRF modeling and optimization framework as shown in Fig. 3.7. That is, we divide a fingerprint image into overlapping patches, where each patch corresponds to a vertex (or node) in a 2D grid and two adjacent vertices are connected by an edge.

Given observed latent fingerprint data and a ridge dictionary, D, consisting of K atoms (with indices from 1 to K), we would like to minimize the following energy function:

$$E(s_1, ..., s_P) = \sum_{i \in V} \varphi(\mathbf{p}_i, \mathbf{y}_{i,s_i}) + \lambda \sum_{(i,j) \in E} \phi(\mathbf{y}_{i,s_i}, \mathbf{y}_{j,s_j}), \tag{3.14}$$

where $\mathbf{p}_i$ is the i^{th} normalized patch of the texture component obtained by TV or ADTV depending on local image quality, s_i is the associated atom index, P is the total number of patches in this latent fingerprint image, φ is the unary potential function used to impose the data fidelity constraint, ϕ is the interaction potential function used to impose the smoothness constraint, λ is a parameter that offers a trade-off between the data fidelity term and the neighborhood smoothness term, and

$$\mathbf{y}_{i,s_i} = \alpha_{i,s_i}\mathbf{d}_{s_i} \tag{3.15}$$

denotes a possible enhanced output for the i^{th} patch by selecting atom $\mathbf{d}_{s_i}$ from the ridge dictionary, D, with coefficient α_{i,s_i}. It is worthwhile to point out that the realization output of optimized variables $s_i, i = 1, \cdots, K$, is an atom from the ridge dictionary. We set $\lambda = 1$ in our experiment for simplicity. It is however possible to adjust λ adaptively if we can tell which term is more important at a certain spatial location.

The conventional sparse representation is optimized against the MSE measure, which is however not robust in latent fingerprint enhancement. For example, for a heavy-noise-corrupted region, it is possible that the closest atom measured in the MSE sense approximates structured noise more than the fingerprint component. Due to the recent success of the SSIM index in image quality assessment [106] and fingerprint quality measure [12], we compute the SSIM value between the texture component of a fingerprint patch and its candidate output and use it as the quality measure. We will define unary potential φ and interaction potential ϕ below.

Unary Potential

We define the unary potential in Eq. (3.14) for patch $\mathbf{p}_i$ as

$$\varphi(\mathbf{p}_i, \mathbf{y}_{i,s_i}) = 1 - SSIM(\mathbf{p}_i, \mathbf{y}_{i,s_i}), \tag{3.16}$$

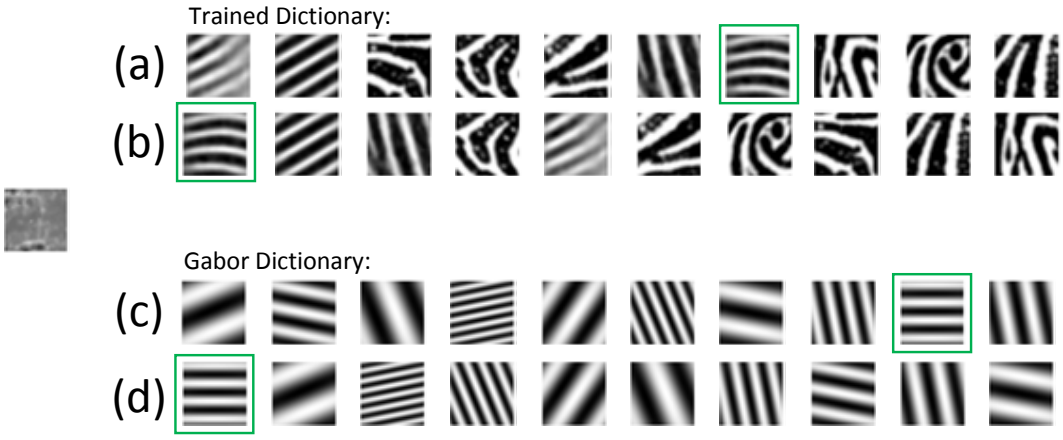


Figure 3.8: Advantage demonstration using the defined unary potential: the same example in Fig. 3.2 is used. (a), (b) are the reassigned order based on MSE and defined unary potential respectively for trained dictionary;(c), (d) are similar but for Gabor dictionary. Order in this figure is from the closest to the farthest.

which indicates the dissimilarity between a candidate solution and texture component of the fingerprint patch, based on the corresponding SSIM.

The well known SSIM is defined as

$$SSIM(I_a, I_b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{a,b} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)}, \tag{3.17}$$

where I_a and I_b are two images or image patches; μ_a and σ_a, μ_b and σ_b are the intensity mean and standard deviation of I_a, I_b respectively; $\sigma_{a,b}$ is the covariance between I_a and I_b; C_1 and C_2 are constant parameters to avoid the computation instability. Fig. 3.8 shows a order reassignment if we only consider our defined unary potential using the same example in Fig. 3.2. Compared (b) with (a), (d) with (c) in Fig. 3.8, for both of the dictionaries, we can see that the expected atom pops up to the top even though it seldom has chance to be selected from conventional sparse coding strategy. This shows the advantage of our defined unary potential that makes it more robust to select the indeed correct atom from the ridge pattern dictionary.

Interaction potential

Concordance in orientation is required in the overlapping region of neighboring patches. Since candidates are obtained from the ridge dictionary with a clean ridge structure, their orientation estimation is relatively precise. We use the interaction potential to impose the compatibility of candidates in a local neighborhood in form of:

$$\phi(\mathbf{y}_{i,s_i}, \mathbf{y}_{j,s_j}) = 1 - \frac{1}{N} \sum_{n=1}^{N} \left| \cos O_{i,n}^{j} \right|. \tag{3.18}$$

It denotes the orientation potential, where $O_{i,n}^{j}$ represents the pixel orientation difference of the candidate $\mathbf{y}_{i,s_i}$ with respect to $\mathbf{y}_{j,s_j}$ in their corresponding overlapping region, similar to [36]. The pixel orientation map is calculated using the gradient method in [46] and N is the size of the overlapping area that contributes to the interaction potential.

Candidate Pool and Energy Minimization

For a given patch, it is not necessary to test all atoms in the ridge dictionary since it is time-consuming. Instead, we use sparse coding via OMP with the MSE measure to get the top n_c candidates.

$$\alpha_i = \arg\min_{\alpha} \|\mathbf{p}_i - D\alpha\|_2^2, s.t. \|\alpha_i\|_0^0 = n_c, \tag{3.19}$$

$$\alpha_{i,\mathbf{c}_i} = \{\alpha_{i,k} \neq 0\}_{k=1}^{K} = \{\alpha_{i,c_i^1}, \alpha_{i,c_i^2}, ..., \alpha_{i,c_i^{n_c}}\}, \tag{3.20}$$

where α_i is the coefficient vector when selecting n_c atoms from the dictionary for patch $\mathbf{p}_i$; $\alpha_{i,\mathbf{c}_i}$ is the sub-vector of α_i in which the elements are non-zero; $\mathbf{c}_i = \{c_i^1, c_i^2, ..., c_i^{n_c}\}$

is the indice list of the non-zero coefficient, which also represents indice of the corresponding dictionary atoms as possible candidates. Therefore, the candidate enhancement output pool $\mathbf{Y}_{i,\mathbf{c}_i}$ respectively for patch $\mathbf{p}_i$ can be expressed as

$$\begin{aligned}\mathbf{Y}_{i,\mathbf{c}_i} &= \{\mathbf{y}_{i,c_i^1}, \mathbf{y}_{i,c_i^2}, ..., \mathbf{y}_{i,c_i^{n_c}}\} \\ &= \{\alpha_{i,c_i^1} \cdot \mathbf{d}_{c_i^1}, \alpha_{i,c_i^2} \cdot \mathbf{d}_{c_i^2}, ..., \alpha_{i,c_i^{n_c}} \cdot \mathbf{d}_{c_i^{n_c}}\}.\end{aligned} \tag{3.21}$$

With Candidates Pool as restriction for each patch rather than picking up any atom from the whole dictionary, Eq. 3.15 will become to

$$\mathbf{y}_{i,s_i} = \alpha_{i,s_i}\mathbf{d}_{s_i}, s_i \in \mathbf{c}_i = \{c_i^1, c_i^2, ..., c_i^{n_c}\}. \tag{3.22}$$

From the nature and process of OMP, texture component of latent fingerprint will be sliced into n_c layers. The sum of the candidates or linear combination of the related n_c atoms will be a close approximation of the texture component by sparse representation. More importantly, the actual fingerprint signal can be in any layer but not necessary to be in the first few layers.

Finally, the loopy belief propagation algorithm is employed here to minimize the cost function in Eq. (3.14). As a message passing algorithm, belief propagation can be applied to any graphical model with any defined potential, and it will provide exact inference when there is no loop in the graph. It can be well approximated by "loopy" belief propagation [8, 107] for ease of implementation and parallelization.

3.4 Experimental Results

Experiments are conducted to demonstrate the effectiveness and robustness of the proposed MRF-SR method on latent fingerprint image enhancement and OF estimation in

this section. The datasets and the experimental setup will be introduced first. Then, the effect of various TV models and ridge dictionaries on the matching performance of latent fingerprints of three quality categories is investigated. Finally, the performance of the proposed latent fingerprint enhancement method with a fusion strategy is presented and compared with other state-of-the-art algorithms.

3.4.1 Experimental Datasets and Setup

Datasets FVC2002[1], NIST SD27[2] and NIST SD14[3] were used in our experiments [37, 84, 83]. To construct the learned dictionary via K-SVD, we selected 430 high quality fingerprints and collected high quality patches from FVC2002. The number of training patches varies with the patch size and the dictionary size. In our case, for a patch size of 32×32, we used 351,905 patches to train a ridge dictionary X of size 3072. For the Gabor dictionary, we applied the same dictionary in [67], which had 864 atoms.

The latent fingerprint dataset, NIST SD27, is often used to evaluate the performance of fingerprint enhancement, OF estimation and identification (or matching) by researchers. It consists of 258 latent fingerprint images and their mated rolled fingerprints. Most fingerprints in NIST SD27 are highly corrupted due to various noise sources, including complicated background, unclear ridge patterns, overlapping fingerprints, etc. This dataset is subjectively classified into three quality categories "Good", "Bad" and "Ugly" by fingerprint experts. There are 88, 85 and 85 images in the good, bad and ugly categories, respectively. We perform our analysis experiments only using the 258 corresponding rolled fingerprints as true mates in 3.4.2. To make latent fingerprint identification more practical and challenging, we used 27,000 rolled fingerprints

[1]FVC2002: http://bias.csr.unibo.it/fvc2002/

[2]NIST SD27: http://www.nist.gov/srd/nistsd27.cfm

[3]NIST SD14: http://www.nist.gov/srd/nistsd14.cfm

from the NIST SD14 dataset as additional matching templates (called the background dataset) in our final experiments. Thus, by including 258 mated rolled fingerprints from NIST SD27, we had 27,258 registered rolled fingerprints in total as the final identification dataset to evaluate the performance in Sec. 3.4.3.

The performance of latent fingerprint enhancement can be subjectively evaluated by the human visual system (HVS). For example, clear ridge patterns, arch flows and reliable observed features indicate better enhancement performance. It is however challenging to develop an objective metric to evaluate the enhancement performance. Since fingerprint identification is the ultimate goal, we evaluate the identification performance with the help of a commercial fingerprint identification software known as VeriFinger SDK 6.2 [4]. That is, we compare the identification performance in terms of the cumulative match characteristic (CMC) curve using different enhancement methods. The CMC curve plots the rank-k identification rate by varying the value of k, which represents the percentages of true mates' appearance in the top k matches. The CMC curve is the most widely used measure for fingerprint identification, and we plot it against the NIST SD27 dataset as our performance measure.

For fair comparison, we used manually marked ROI's for the identification task by following the same practice in [36, 113, 67, 117, 11]. Each latent fingerprint image with its manual ROI was used as the query to search over the rolled fingerprint dataset. When performing fingerprints identification, the corresponding rolled fingerprints as true mates are not enhanced and only latent fingerprints are enhanced.

3.4.2 Effects of TV Models and Dictionaries

As mentioned in Subsection 3.3.2, the learned and the Gabor dictionaries have their own advantages and disadvantages. Thus, there is no obvious reason to choose one over

[4] VeriFinger: http://www.neurotechnology.com/verifinger.html

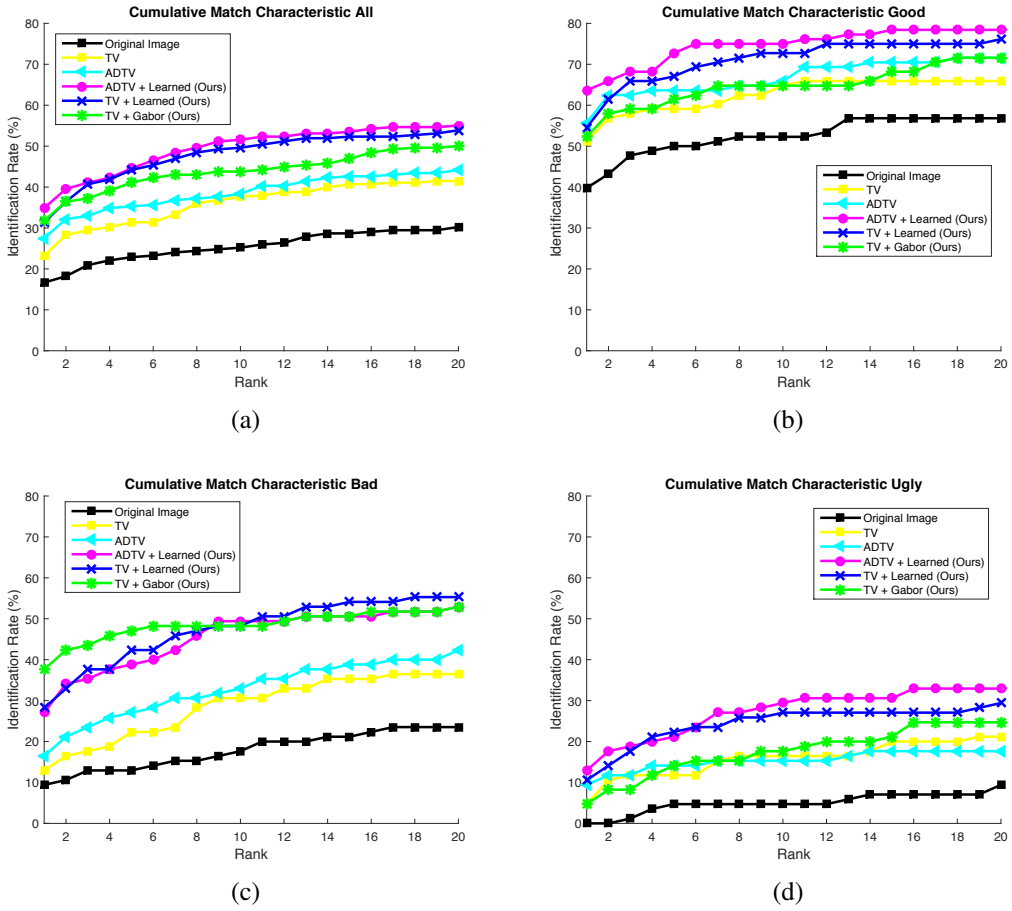


Figure 3.9: The CMC perfomance curves with various TV models and dictionaries for (a) the whole NIST SD27 dataset, (b) the good category, (c) the bad category and (d) the ugly category.

the other. Actually, Liu *et al.* [67] adopted the Gabor ridge dictionary while and Cao *et al.* [12] used their learned ridge dictionary. Similarly, the traditional TV-L1 model was used in [12, 67, 11] while the ADTV model was adopted in [119] to extract the texture component. We are interested in knowing whether there is a clear winner or they provide complementary strengths. To answer these questions, we conducted latent fingerprint identification experiments based on the original input without any enhancement (ORIG) and the enhanced input with the five enhancement schemes: 1) the ADTV decomposition plus the learned dictionary (ADTV+L), 2) the TV-L1 decomposition plus

the learned dictionary (TV+L), 3) the TV-L1 decomposition plus the Gabor dictionary (TV+G), 4) the ADTV decompostion only (ADTV), and 5) the TV-L1 decomposition only (TV). Their performance was compared for the following four scenarios: 1) the whole NIST SD27 dataset, 2) the good category, 3) the bad category and 4) the ugly category in the NIST SD27 dataset. The results are shown in Figs. 3.9 (a)-(d), respectively.

The CMC curves of several enhancement methods against the whole NIST SD27 dataset are shown in Fig. 3.9(a). We see that the ADTV+L scheme achieves the best performance. The second one is TV+L, whose first rank identification rate drops by around 3.5% in comparison with ADTV+L. Then, the performance becomes poorer in the order of TV+G, ADTV, TV and, finally, ORIG. The proposed ADTV+L scheme outperforms the ADTV scheme by 7.4% in the Rank-1 identification rate. Generally speaking, the ADTV decomposition offers better results than the TV decomposition, and the learned dictionary provides better performance than the Gabor dictionary in the simplified MRF-SR framework. However, there are exceptions when referring to different quality categories. Matching performance for the ugly case is too poor to conclude useful analysis. However, performance analysis on good and bad categories using different combination of TV models and dictionaries shows the complementary strengths.

For example, for the good category in Fig. 3.9(b), ADTV+L outperforms TV+L and TV+G. There are two major reasons. First, latent fingerprints in the good category is relatively strong, and the parameters required by ADTV can be estimated more accurately, which results in better texture decomposition in the ADTV model. Second, because of the relatively clear ridge structures for good latent fingerprints, the learned dictionary is more suitable to recover ridges while preserving features such as minutiae and singular points.

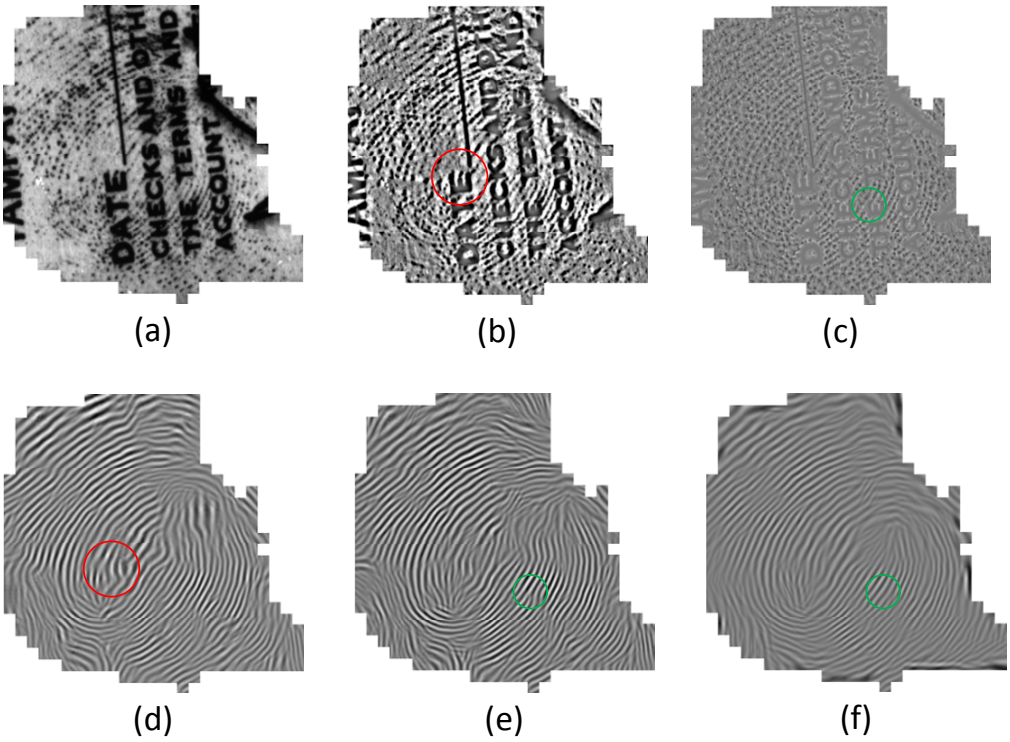


Figure 3.10: An example to explain the superior performance of TV+G for the bad category: (a) the original image of B127, (b) the texture component using the ADTV decomposition, (c) the texture component using the TV-L1 decomposition, (d) the enhanced image using ADTV+L, (e) the enhanced image using TV+L, (f) the enhanced image using TV+G.

However, for the bad category in Fig. 3.9(c), TV+G offers better performance than TV+L and ADTV+L, especially for the rank-1 identification rate. This can be explained as follows. First, poor image quality results in wrongly estimated parameters, which in turn degrades the performance of the ADTV decomposition. Second, structured noise in some bad images is nip and tuck with ridge patterns so that it stays in the texture component, which offers an erroneous query to dictionary search. The TV model offers similar strength for both structured noise and the fingerprint signal, which provides a higher probability in reconstructing correct ridges form the dictionary. Third, bad latent fingerprints often have false minutiae, which tend to mislead the learned dictionary. The Gabor dictionary is more robust to this type of errors.

We use an example to explain the better performance of the TV+G scheme for the bad category of the NIST SD27 dataset in Fig. 3.10(a). Its texture components extracted by the ADTV decomposition and the TV decomposition are shown in Figs. 3.10 (b) and (c), respectively. We observe more structured noise left with the ADTV decomposition as compared to that with the TV decomposition. The learned dictionary can be influenced by structured noise left in the texture component. For example, the letter "E" circled in red in ADTV's texture component is negatively affected for ridge recovery using the learned dictionary as shown in Fig. 3.10(d). For more comparison, we can examine another letter "E" circled in green as shown in Fig. 3.10(c). The application of the learned dictionary leads to a false bifurcation type minutiae as shown in Fig. 3.10(e). In contrast, the application of the Gabor dictionary yields an excellent result as shown in Fig. 3.10(f).

To summarize, we show different results by considering different combinations of TV/ADTV models and dictionaries. They are consistent with the analysis given in Sec. 3.3. Moreover, performance analysis on different quality categories shows the importance of integrating local quality assessment as discussed in Sec. 3.3.3. The performance of the final MRF-SR method will be shown in Sec. 3.4.3 to illustrate the effectiveness of the proposed enhancement method.

3.4.3 Performance Benchmarking and Fusion with the State-of-the-Art Method

Due to the effect of different TV models and dictionaries, it is a natural choice to switch between ADTV/TV dynamically based on local image quality and to adopt the learned dictionary. The learned dictionary will provide good performance in general as compared with the Gabor dictionary. Being similar with [12], we adopt gradient enhancement as the post-processing technique to yield the final enhancement results.

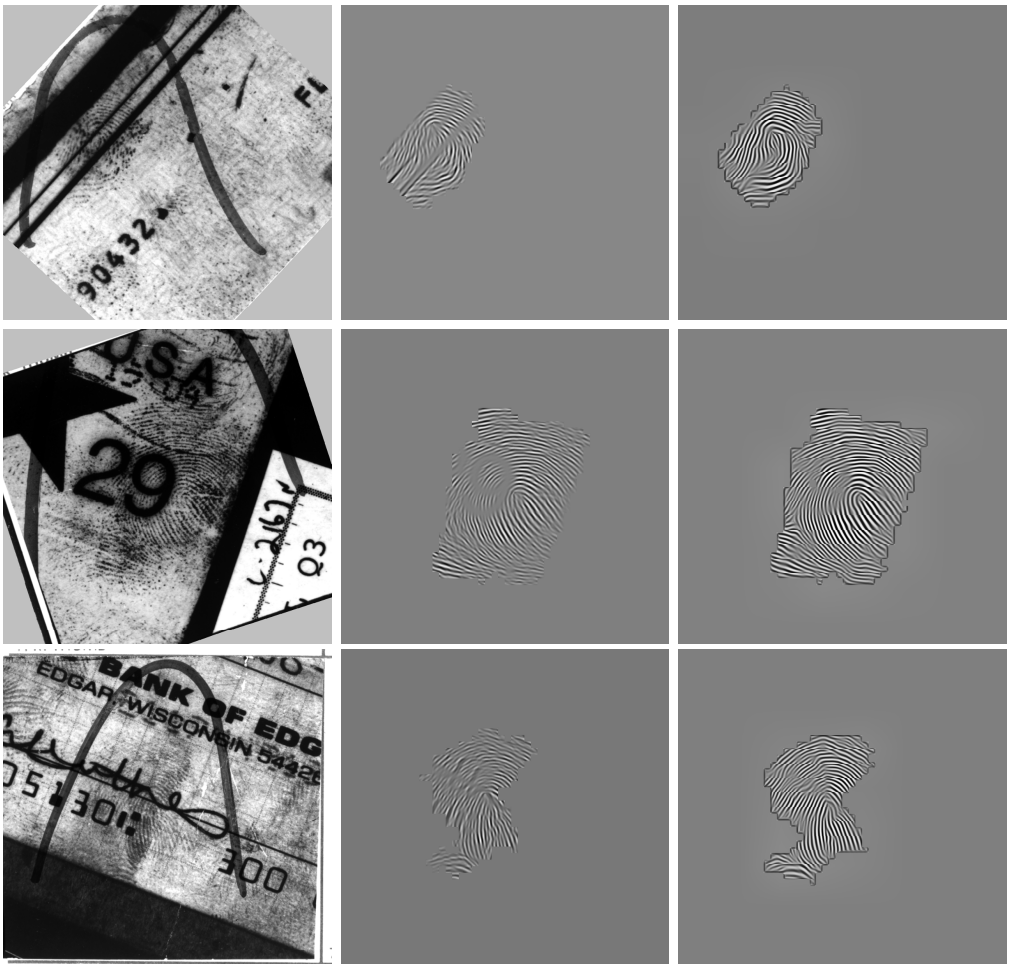


Figure 3.11: Enhanced latent fingerprints visual comparison: (a) the original latent fingerprints, (b) the enhancement results from Localized Dictionary method, and (c) our final enhancement results using the proposed MRF-SR method.

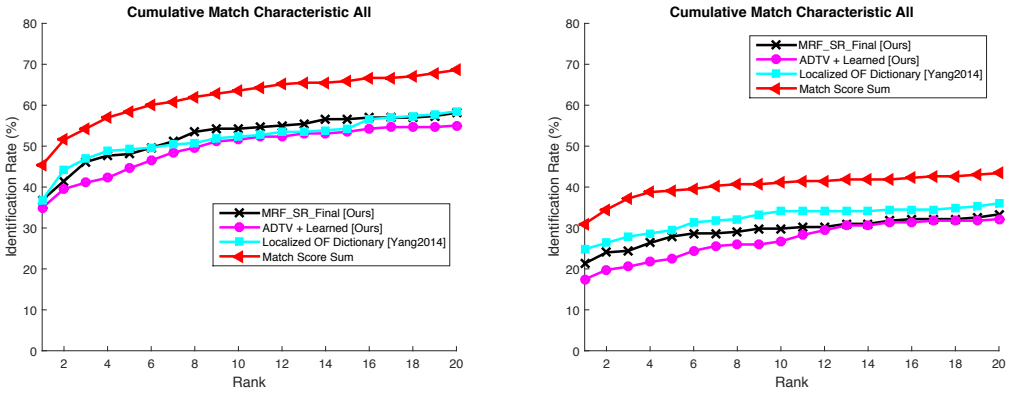
We first evaluate the proposed method visually against the state-of-the-art method proposed in [113], especially the capability of removing structured noise with better ridge connectivity. Several visual examples of the enhanced results as compared with the localized OF dictionary method in [113] are shown in Fig. 3.11. We see that the proposed MRF-SR method has stonger capability in removing structured noise in these examples. Besides, we observe better ridge connectivity in regions where ridges are

missing due to structured noise. Moreover, since our enhancement is based on atoms from the ridge dictionary, we see a clearer ridge/valley contrast.

To have comprehensive performance comparison, we show the CMC curves for cases with and without 27,000 background fingerprints in Fig. 3.12. In particular, we compare the following algorithms: (1) the proposed MRF-SR method, (2) the simplified MRF-SR algorithm using ADTV+L, (3) the localized OF dictionary method in [113] and (4) the matched score sum of (1) and (3). It is worthwhile to point out that the performance of (3) in our evaluation is different from their reported results since they enhanced the rolled fingerprints as well in the matching process. For fair comparison, we only enhance latent fingerprints but not their mated fingerprints, and conduct all experiments under this setting.

As shown in Fig. 3.12, the performance of the MRF-SR method outperforms the simplified ADTV+L with or without adding background fingerprints. This demonstrates the effectiveness of the proposed dynamic switching between texture components from TV/ADTV. It also shows the importance of local quality assessment when handling latent fingerprints.

The localized OF dictionary method [113] offers the state-of-the-art matching performance. We see from Fig. 3.12(a) that the proposed MRF-SR method slightly outperforms it when no background noise is included. This is especially true for the identification rate when the rank goes beyond 8. For latent fingerprint matching, human experts do not rely only on top-1 results since the overall rank-1 identification rate is very low. Therefore, a higher rank-8 identification rate is still meaningful since it can reduce the human labor effort by providing top 8 matching results. After adding 27,000 background rolled fingerprints from NIST SD14, the CMC curves are shown in 3.12(b). The localized dictionary method performs slightly better for the rank-1 and rank-2 identification



(a) (b)

Figure 3.12: Performance comparison against the whole dataset with the MRF-SR method, ADTV+L, the localized OF dictionary method, and the fusion of the MRF-SR and the method in [113]: (a) CMC curves without 27,000 background noise; and (b) CMC curves with 27,000 background fingerprints.

rates. However, our method does not need the precise orientation field estimation as the input. Furthermore, it does not demand the global pose estimation either.

Finally, since there exists diversity in the ridge dictionary and the orientation dictionary, there is an advantage in fusing the localized OF dictionary method and the proposed MRF-SR method for better performance. The result of the fused solution is labeled by the "Match Score Sum" in Fig. 3.12. The fusion idea was also conducted in [12, 11]. We see that the fusion of the two methods offers substantial improvement over any one of the two. With adding background fingerprints, the score sum still outperforms the localized method by 6.2% for the rank-1 identification rate. This score fusion result shows the complementary strength of the proposed MSF-SR and the localized OF dictionary method.

3.5 Conclusion

Despite rapid developments in the automated fingerprint identification system, the identification rate in latent fingerprint matching is still far from satisfaction. In this work, we proposed an effective and robust solution by utilizing the sparse ridge dictionary and the MRF model to enhance latent fingerprints. Extensive studies were conducted on the pros and cons of TV/ADTV models and learned/Gabor ridge dictionaries. Without focusing on estimating the orientation field and/or the fingerprint pose, we developed direct enhancement methods for improved fingerprint matching in terms of the identification rate. The fusion of the proposed MRF-SR method and the localized OF dictionary method offers the state-of-the-art performance. For further performance improvement, we would like to find a robust way to estimate the global information of a local patch and, then, incorporate this information in the local patch enhancement process. Moreover, our method only uses the single-scale ridge dictionary and it is worthwhile to extend it to multi-scale dictionaries in the near future.

Chapter 4

A Fusion Approach to Orientation Field Estimation

4.1 Introduction

As one of the important fingerprint features, the orientation field (OF) plays a fundamental role in fingerprint analysis and recognition. It can be used to assist fingerprint enhancement (using contextual filtering), fingerprint types classification, and singular points detection. Although the proposed MRF-SR method can achieve good performance for latent fingerprint enhancement without necessity of estimating orientation field, it is still worthwhile to provide a robust solution for orientation field estimation.

The proposed MRF-SR method not only enhances latent fingerprint, but also has the ability to estimate orientation field. Since the enhancement results are obtained by stitching ridge dictionary atoms, the orientation field can also be estimated stitching those atoms' orientation. One advantage of this strategy is that the orientation estimated for ridge atoms is accurate due to high quality. However, certain modification of MRF-SR needs to be involved in order to fit better to the problem of orientation estimation. With the modified MRF-SR (MRF-SR-Modified), we evaluate the performance especially on orientation estimation and compare with other existing approaches. Our proposed method is demonstrated to outperform the existing ridge dictionary based method such as [12].

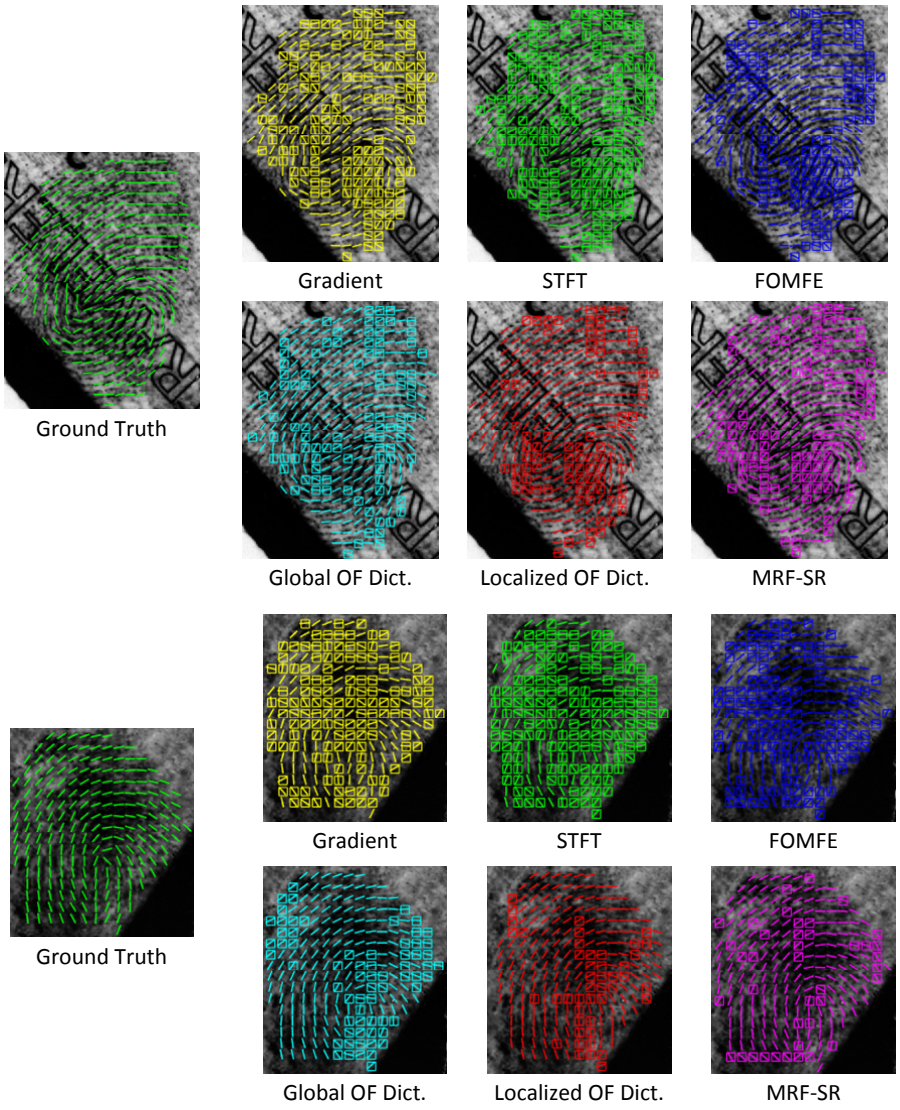


Figure 4.1: Orientation Field estimated by different methods: Gradient-based, STFT, FOMFE, Global OF Dict., Localized OF Dict., and MRF-SR. Wrong local orienations compared with groudnt truth are marked with squares.

We further propose a fusion based approach for orientation field estimation. This strategy is inspired by the fact that different approaches concentrate on different aspects of fingerprint orientation. For example, gradient based method focuses much on local derivative of spatial domain; STFT method handles orientation in the Fourier domain; the global and localized OF dictionary based method rely on compatibility of orientation

elements at local and global; the MRF-SR method emphasizes on the importance of guidance of the actual fingerprint ridges for orientation estimation. Fig. 4.1 shows an illustration about the diversity of different methods. In general, the advanced techniques Global OF Dict., Localized OF Dict. and MRF-SR methods predict orientation better than others. It can be shown that different approach successes at different locations, which implies the potential to fuse them.

4.2 Related Previous Work

Different from the approaches mentioned in Chapter 1 that are mainly proposed for exemplary fingerprints, more and more work are proposed particular for latent fingerprint such as [11, 12, 36, 113].

4.2.1 Global OF Dictionary

In [36], Feng *et al.* proposed a global orientation field dictionary based method to make better smoothing of orientations of fingerprint. This work used an orientation patch dictionary to smooth the initial OF, where clean reference orientation patches are used to create a global dictionary as the prior of the fingerprint orientation. For every initial OF patch, they looked up this dictionary for potential candidates and, then, minimized an energy function to estimate the final OF using loopy belief propagation, where the energy function is defined by orientation similarity and neighboring compatibility. The main steps of this method as given below.

1. Dictionary construction: in the off-line stage, a global orientation field dictionary is constructed by the greedy algorithm using orientation field patch (10×10 orientation elements) of different good quality fingerprints.

2. OF initialization: orientation field estimated by local Fourier analysis is used as the initial orientation field for a latent fingerprint [51].

3. Dictionary lookup: the initial OF is divided into overlapping orientation patches. Each initial orientation patch's six nearest neighbors in the OF dictionary are selected as possible candidates to replace the initial patch.

4. Contextual correction: a combination of candidates for initial orientation patch is optimized by considering compatibility of neighboring orientation patches.

The optimization equations are given below.

$$E(r_1, ..., r_P) = E_s(r_1, ..., r_P) + \lambda_c E_c(r_1, ..., r_P), \tag{4.1}$$

$$E_s(r_1, ..., r_P) = \sum_{i \in V} \{1 - S(\Theta_i, O_{i,r_i})\}, \tag{4.2}$$

$$E_c(r_1, ..., r_P) = \sum_{(i,j) \in N} \{1 - C(O_{i,r_i}, O_{j,r_j})\}, \tag{4.3}$$

where $r_i (1 \leq r_i \leq n_c)$ means the index that the orientation patch i select among n_c candidates, Θ_i is the initial orientation for patch i, O_{i,r_i} denotes the candidate orientation when candidate r_i is chosen for patch i, V is the set of all fingerprint foreground patches, and N represents the set of 4-connected neighboring foreground fingerprint patches. When the candidates are selected, E_s means the similarity energy and E_c means the compatibility energy with those candidates, and λ_c is the weight to balance these two energy functions.

The similarity S is the proportion of similar orientation elements within a threshold between initial orientation patch and the candidate patch. The compatibility function C is defined as

$$C(O_{i,r_i}, O_{j,r_j}) = \frac{1}{N_0} \sum_{n=1}^{N_0} \left| \cos(\alpha_{r_i,n} - \beta_{r_j,n}) \right|, \tag{4.4}$$

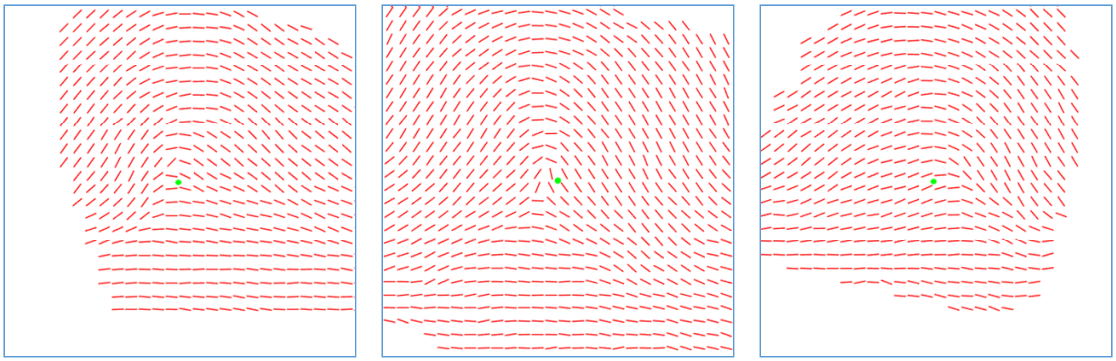


Figure 4.2: Examples of showing registered orientation field for NIST SD4. Green dots are the mannually marked pose and the orientation fields are registered to upright direction in the image center.

where $\alpha_{r_i,n}$ and $\beta_{r_j,n}$ denote orientations of the overlapping orientation elements between patch O_{i,r_i} and O_{j,r_j}. Loopy belief prorogation is employed to minimize the final energy function.

4.2.2 Localized OF Dictionary

However, since this global dictionary does not consider the location-dependent information, the estimated OF of a patch may not be consistent with its spatial location in the fingerprint. To improve this, Yang *et al.* proposed a localized OF dictionary based method to further smooth the local estimated orientation [113]. The two major difference between global OF dictionary based method and localized OF dictionary based method are as follows.

1. Unlike a global OF dictionary, the localized OF dictionary is a set of dictionaries which depends on the location to fingerprint pose. The localized OF dictionaries are learned from a set of registered training orientation fields. Fig. 4.2 shows examples of the registered orientation field as training samples.

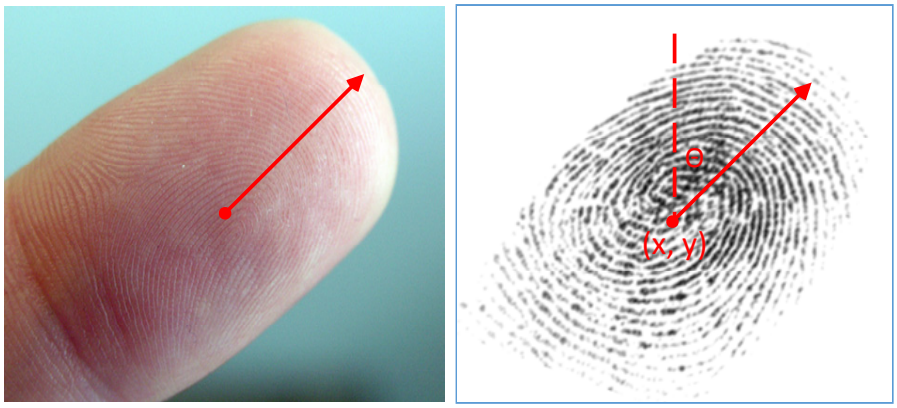


Figure 4.3: An illustration about finger pose (x, y, θ)

2. Pose of fingerprint is applied in this work. Pose of a fingerprint is illustrated in Fig. 4.3. With pose estimated for a fingerprint, the fingerprint orientation field is registered firstly and then each orientation patch will look up the corresponding localized OF dictionary based on its location in the finger instead of a global singe OF dictionary.

The other procedures such as energy optimization, contextual correction are similar with the Global OF dictionary based method. The results show an improvement on the orientation field estimation.

4.3 Proposed Methods for OF Estimation

4.3.1 MRF-SR-Modified Orientation Field Estimation

The proposed MRF-SR method can also estimate orientation field while enhancing latent fingerprints. The methodology is simply to make use of the facts that it is easy and

effective to estimate orientation for atoms in the ridge dictionary. Both the learned dictionary and Gabor dictionary in Chapter 3 include ridge atoms in a well-defined ridge-valley structure. Therefore, it makes possible to utilize stitching of the optimized atom to obtain the final estimated orientation field.

However, to make the MRF-SR method even more proper particular for the orientation field estimation application, several modification can be done to the original MRF-SR, and we call it MRF-SR-Modified.

Firstly, for orientation estimation, it is not necessary to emphasize too much on the data fidelity. In the other hand, candidates atoms with more diversity for a fingerprint patch tend to provide more possible ways when reconstructing a new orientation field. Furthermore, since it is only for orientation field estimation, preserving details such as minutiae is not as critical as fingerprint enhancement, which means a Gabor dictionary is possible to be effective for orientation field reconstruction. Therefore, TV-L1 decomposition with Gabor dictionary will be the proper setting for MRF-SR-Modified in order to estimate orientation.

Secondly, the fingerprint pose is embedded in the MRF-SR method by defining a new unary potential:

$$\varphi_{new}(\mathbf{p}_i, \mathbf{y}_{i,s_i}) = \varphi(\mathbf{p}_i, \mathbf{y}_{i,s_i}) + \varphi_{pose}(\mathbf{p}_i, \mathbf{y}_{i,s_i}), \tag{4.5}$$

$$\varphi_{pose}(\mathbf{p}_i, \mathbf{y}_{i,s_i}) = 1 - p(\mathbf{O}_i|u_c), \tag{4.6}$$

where $\varphi(\mathbf{p}_i, \mathbf{y}_{i,s_i})$ is defined in Eq. 3.16, $p(\mathbf{O}_i|u_c)$ is the probability of the estimated orientation of patch $\mathbf{p}_i$ located in the whole fingerprint given the estimated pose of fingerprint u_c. After this modification, we follow the MRF-SR procedure in Ch. 3 but to generate the estimated orientation.

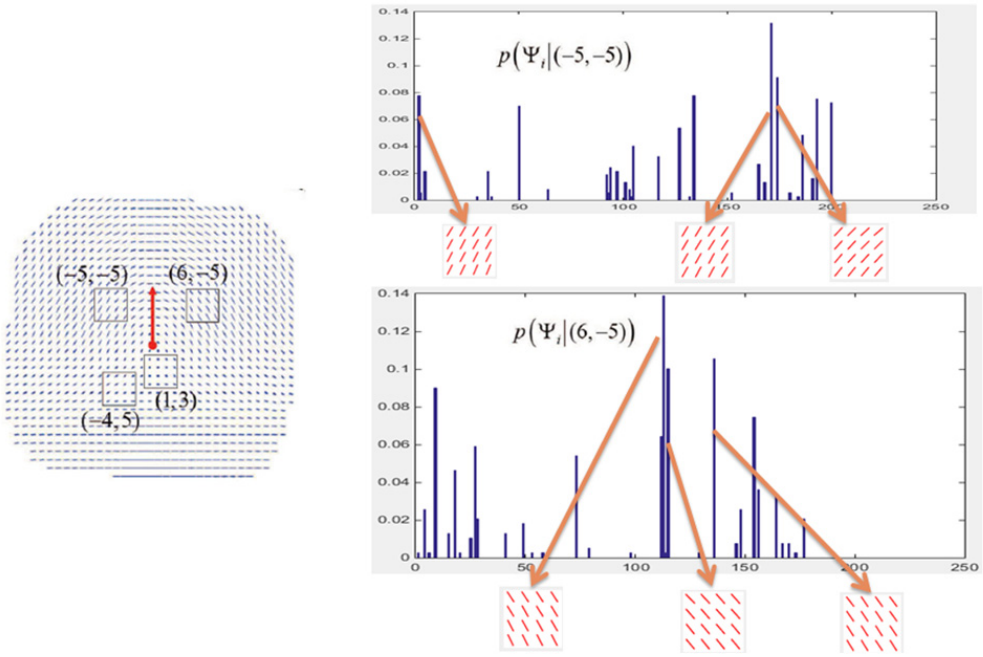


Figure 4.4: An example of probability distribution of prototypes: at location $(-5, 5)$, the prototypes with high probability tend to be the correct orienatin patches appear in this location. Similary for location $(6, 5)$.

Therefore, fingerprint pose estimation will be a key factor her and it is estimated by the Hough Voting schema as follows in [113]:

1. The k-medoids clustering algorithm is applied to the orientation training patches so that 200 prototype orientation patches $\{\Psi_i\}, i = 1, ..., 200$ can be obtained.

2. The probability distribution of prototype patch Ψ_i at location u, $p(\Psi_i|u)$ can be approximated by the frequency of patches in cluster i occurring at location u as follows:

$$p(\Psi_i|u) = \frac{N_{u,i}}{N_f}, \tag{4.7}$$

where $N_{u,i}$ is the number of orientation patches at location u in the training samples belonging to cluster i, and N_f represents the number of training fingerprints. An example of this probability distribution of prototypes is shown in Fig. 4.4.

3. With the probability distribution $p(\Psi_i|u)$, the probability that fingerprint pose is $c = (x, y)$ given an initial orientation path o appearing at location v, $p(c|o, v)$, can be approximated as:

$$p(c|o, v) \approx p(c|\Psi^*, v) = p(u_c|\Psi^*), \tag{4.8}$$

where Ψ^* is the nearest neighbor for o among all the prototypes, $u_c = v - c$. Note that $p(u_c|\Psi^*)$ is the posterior probability of Eq. 4.7 that can be estimated by Bayes' theorem easily.

4. The initial orientation field can be divided into overlapping patches and then $p(c|o, v)$ can be estimated for each patch. Then, a Hough Voting will be applied to determine the final pose of a fingerprint.

With pose estimated, the probability of a ridge atom occurs at its location for a fingerprint patch can be estimated. If the probability is smaller than a pre-defined threshold, the unary potential for this atom will be set to 1, which means the optimization tends not to choose this atom eventually.

4.3.2 A Learning based Fusion Strategy

For fingerprint community, few methods are proposed using machine learning techniques. The reason is that feature for fingerprint("feature" here means the machine learning feature vector, different from fingerprint feature mentioned in previous chapters) is hard to design. Zhu *et al.* designed 11 features based on ridge analysis for fingerprint in [124], however, it is hard to use for latent fingerprint because of the low quality ridge structure. Features are usually corrupted so that they can not represent fingerprint well.

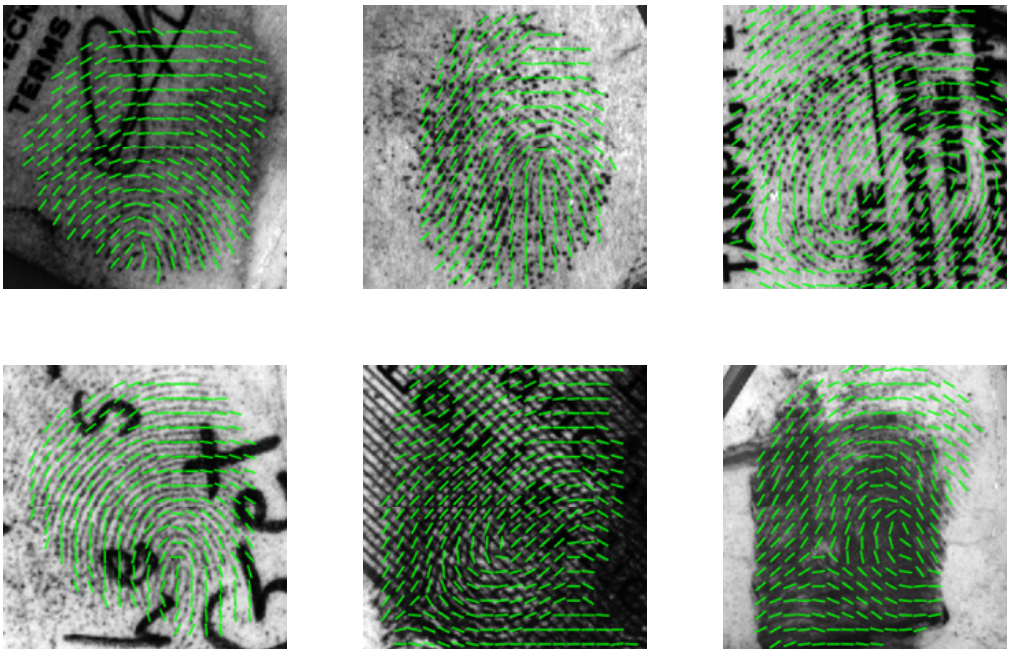


Figure 4.5: Examples of Ground Truth orientation field for latent fingerprint in NIST SD27.

Therefore, instead of using hand engineered features for fingerprint, we propose a decision fusion strategy for latent fingerprint orientation estimation as follows:

1. Six experts such as Gradient based method, STFT, FOMFE, Global OF based method, Localized OF based method and the MRF-SR-Modified based method are used firstly to estimate orientation field.

2. Estimated orientation for each block is considered as a decision for each expert. Then, there are 6 orientation decisions for each fingerprint block.

3. Treat decisions as features and a supervised learning technique such as Support Vector Regression, Linear Regression and Neural Network is applied to train the model.

4. Final testing can be conducted on the unlabeled dataset.

Table 4.1: Fingerprint Datasets Summary

Dataset	Description	Major Purpose
NIST SD4	2,000 pairs of rolled fingerprint http://www.nist.gov/srd/nistsd4.cfm	OF modeling, background dataset,other purpose
NIST SD14	27,000 pairs of rolled fingerprint http://www.nist.gov/srd/nistsd14.cfm	background dataset
NIST SD27	258 pairs of latent fingerprints and the mated rolled fingerprint http://www.nist.gov/srd/nistsd27.cfm	performance evaluation
Tsinghua Overlapped Latent	100 latent fingerprints with overlapped, 12 mated plain fingerprints http://ivg.au.tsinghua.edu.cn	overlapped fingerprint separation
FOE-STD-1.0	10 good quality and 50 bad quality fingerprints	OF evaluation

4.4 Experimental Results

4.4.1 Datasets and Performance Evaluation

For latent fingerprint, the NIST SD27 dataset is still used for the orientation field estimation. However, a summary of several dabasets existed is provided in Table 4.1. Several datasets in this table are used in this work.

For orientation estimation, the ground truth orientation was established manually marked based on 16×16 block in [100]. Fig. 4.5 shows some examples of the manually marked orientation ground truth. The performance of orientation field estimation algorithms is quantitatively measured by the average Root Mean Square Deviation (RMSD) of the estimated orientation fields from the ground truth orientation fields, as suggested in [100].

$$RMSD(D_k, G_k) = \sqrt{\frac{\sum_{i \in F} d\phi(\theta_i^k, g_i^k)^2}{|F|}}, \tag{4.9}$$

$$d\phi(\theta_1, \theta_2) = \begin{cases} \theta_1 - \theta_2, & \text{if } -\frac{\pi}{2} \leq \theta_1 - \theta_2 < \frac{\pi}{2} \\ \pi + \theta_1 - \theta_2, & \text{if } \theta_1 - \theta_2 < -\frac{\pi}{2} \\ \pi - \theta_1 + \theta_2, & \text{if } \theta_1 - \theta_2 \geq \frac{\pi}{2} \end{cases}, \tag{4.10}$$

where G_k, D_k is the ground truth orientation field and estimated orientation field for the k^{th} ($1 \leq k \leq N$) fingerprint, $F = \{i\}_k$ denotes the set of foreground orientation elements, and g_i^k, θ_i^k means the ground truth and estimated orientation of the i^{th} foreground element in the k^{th} fingerprint. The final average RMSD is defined as:

$$AgeErr = \frac{1}{N} \sum_{k=1}^{N} RMSD(D_k, G_k). \tag{4.11}$$

4.4.2 Results Comparison

The results of different algorithms are shown here in Table 4.2, including the proposed modified MRF-SR. It shows that the local estimation methods are far from satisfactory for orientation estimation on latent fingerprints. The proposed modified MRF-SR provides the second best performance and the localized OF Dictionary method is the state-of-the-art method for orientation estimation. However, the proposed MRF-SR-Modified outperforms the global OF dictionary, which still shows the effective of the proposed method.

For the fusion strategy, the 258 latent fingerprint images in NIST SD27 is divided into two parts: training and testing. 129 latent fingerprint images are randomly selected as training and the others are used for testing. As about orientation elements, there are 41,032 training elements and 40,499 testing elements in total. the results are shown in Table 4.3 with comparing with individual approaches and the fusion approach with different classifier. Here, linear regression, SVM regression and the Neural Network fitting have been applied to evaluate the expert fusion on orientation field. The results

Table 4.2: Average Error (in degree) of Different methods on NIST SD27

Method	**All**	**Good**	**Bad**	**Ugly**
Gradient	32.45	27.30	33.79	36.45
STFT	32.51	24.27	34.10	36.63
FOMFE	28.12	22.83	29.09	32.63
RidgeDict.	19.53	15.34	20.70	22.68
GlobalOFDict.	18.44	14.40	19.18	21.88
LocalizedOFDict.	14.35	11.15	15.15	16.85
MRF-SR-Modified	17.62	13.97	18.35	20.69

Table 4.3: Average Error (in degree) on testing dataset

Method	**All**
Gradient	33.13
STFT	32.84
FOMFE	29.42
GlobalOFDict.	19.79
LocalizedOFDict.	16.53
MRF-SR-Modified	18.23
Neural Network	18.63
Linear Regression	22.23
SVM Regression	21.56

show that Neural Network fitting provides better performance than most of the individual approaches in a large margin except Localized OF Dict. method and our proposed MRF-SR-Modified method. The fusion strategy does not perform as the best, it maybe caused by several reasons such as lack of representative training data, over-fitting issues and so on. However, it does show great potential as an ensemble techniques as more strong individual methods explored.

4.5 Conclusion

In this chapter, the orientation field estimation is studied thoroughly. The proposed MRF-SR-Modified can provide competitive results for orientation estimation, which outperforms most of the current algorithms, especially the existing similar ridge dictionary method. Moreover, to further boost the performance of orientation estimation, a supervised learning based fusion strategy is proposed. In this fusion scheme, decisions from different experts are treated as feature vector and then multiple techniques such as linear regression, SVM regression and the Neural Network fitting are applied to train the model. The final performance on testing shows the potential to fuse other experts in order to boost the orientation estimation performance.

Chapter 5

Deep Convolutional Neural Network for Latent Fingerprint Enhancement

5.1 Introduction

Nowadays, deep learning technique particularly convolutional neural network (CNN) has been proved to be very successful on both low-level and high-level computer vision applications [26, 45, 63, 72, 74, 93]. In this chapter, we have a detailed study about applying CNN on latent fingerprint application to explore capability of deep neural network on this topic.

We propose a novel latent fingerprint enhancement method called "FingerNet" based on recent development of Convolutional Neural Network (CNN). Although CNN is achieving superior performance in many computer vision tasks from low level image processing to high level semantic understanding, limited attention has been paid in the fingerprint community. Our proposed neural network has three major parts: one common convolution part shared by two different deconvolution parts, that are the enhancement branch and orientation branch. The convolution part is to extract fingerprint features particularly for enhancement purpose. The enhancement deconvolution branch is employed to remove structured noise and enhance the fingerprints as its task. The orientation deconvolution branch performs the task of guiding fingerprint enhancement

through a multi-task learning strategy. The network is learned in the manner of pixel-to-pixel and end-to-end learning, that can directly enhance latent fingerprint as the output. We also study some implementation details such as single task learning, multi-task learning, and residual learning. Experimental results of our FingerNet system on latent fingerprint dataset NIST SD27 demonstrate effectiveness and robustness of the proposed method.

As mentioned in Chapter 3, Cao *et al.* proposed a CNN based OF estimation method, that treated it as a classification problem [11]. However, there are two major limitations for this method: First, only 128 representative orientation patterns are used to stitch the final orientation field. Each fingerprint patch will be classified into one of the 128 representative pattern using CNN. Therefore, the estimated orientations fall in lack of rich variation due to the natural of classification modeling. This might be also the reason that it needs re-estimate the orientation field again using the enhanced fingerprints. Second, the noise types added in order to generate training samples are majorly limited to lines and zero blocks. This is not sufficient enough to simulate the actual latent fingerprint patches, especially ignoring one of the most important structured noise in latent case - characters.

Our proposed FingerNet is inspired by the recent development of CNN on image processing applications, particularly using the pixel-to-pixel and end-to-end learning manner [74, 28, 29, 60, 61, 80, 81]. We propose a deep network architecture directly targeting on fingerprint enhancement in the aforementioned pixel-to-pixel and end-to-end fashion. Different from [11] that applied CNN on orientation classification in an indirect way of enhancing fingerprint, our method is a solution in new angle to solve enhancement more directly. In this work, training data is firstly generated via distorting clear fingerprint patches by adding structured noise such as lines and characters. The

distorted fingerprint patch is then used as input and its enhanced version by Gradient-based method is used as output of the proposed FingerNet. Our network architecture is consist of multiple convolutional layers and their corresponding symmetrical deconvolutional (or, transposed convolutional) layers. This is also within the framework of encoder-decoder network. Instead of increasing depth of the network to enlarge the receptive field [60, 61, 80, 81], we apply pooling the first time for image processing applications. Two branches of orientation classification and fingerprint enhancement are involved via multi-task learning. In order to compensate the possible image details lost when recovering caused by pooling, skip connections are introduced to pass image details of different resolution.

The major contributions of our FingerNet include four folds: First, to our best knowledge, we are the first one to apply pixel-to-pixel and end-to-end learning for latent fingerprint enhancement; Second, we propose a better data preparation/augmentation to have characters, lines, and other structured noise; third, we successfully train the encoder-decoder network with pooling and striding included. This is different from all other image processing networks without pooling layers; Fourth, the mutli-task learning combining orientation and enhancement proposed in this chapter is proved to be successful.

The rest of this chapter is organized as follows. Related work in the literate is introduced in Sec. 5.2. Then, our proposed method including several implementation details is introduced in Sec. 5.3. Experimental results are shown and analyzed in Sec. 5.4. Finally, we summarize this chapter and point out possible future work in last section.

5.2 Related Previous Work

In this section, we are mainly focusing on related work applying CNN on image processing and latent fingerprint enhancement. Although CNN has shown great success on

computer vision tasks such as image classification, image segmentation, object detection and so on [25, 43, 62, 97, 98], limited attention has been paid into lower level applications, e.g. image processing such as image super resolution, denoising, deblurring, etc.

5.2.1 CNN on Image Processing

Convolutional neural network has achieved grand success not only on high-level semantic vision applications, but also on low-level image processing problems. Since fingerprint enhancement naturally belongs to low-level image processing category, deep learning research on image processing especially on image denoising, super-resolution and restoration is more related here. Instead of using the pre-defined image priors, the deep learning methods conduct learning directly from pixel to pixel in the end-to-end style.

Xie *et al.* proposed Stacked Sparse Denoising Auto-Encoders (SSDA) for image denoising [111]. Denoising Auto-Encoders are stacked to form the network, and then the pre-trained technique with sparse regularization is used to solve the network. Recently, several CNN based methods were presented on image super-resolution. Dong *et al.* proposed a fully convolutional neural network SRCNN that includes 3 layers convolution [28]. It shows representative and success to learn high resolution images from low resolution ones. Kim *et al.* successfully trained a very deep convolutional network (20 weight layers) with residual learning for super-resolution. Each convolutional layer uses a simple $3 \times 3 \times 64$ filter similar with the popular VGG net [60]. Besides, Kim *et al.* also presented a deeply-recursive convolutional network (DRCN) as one solution to super-resolution [61]. This method can increase the network depth by recursion without having additional parameters. To make learning efficient in DRCN, the recursive layers are also under supervision and the similar one path skip connection from input to output exists. Mao *et al.* proposed a fully convolutional Encoder-Decoder network for

image restoration [80]. The symmetrical skip connections to link convolution layer and deconvolution layer in this chapter, can make a faster and better convergence.

5.2.2 CNN on Latent Fingerprint Enhancement

Although the end-to-end learning shows superior performance as state-of-the-art approach, it has not been applied successfully on latent fingerprint application to our best knowledge. The most recent applying of CNN on latent fingerprint was proposed in [11]. However, different from the pixel-to-pixel and end-to-end learning manner, Cao *et al.* applied CNN to estimate the orientation field first and then enhance the latent fingerprint using Gabor filter banks. The OF estimation via CNN is treated as the classification problem. 128 representative orientation patches (160×160 pixels, 10×10 orientation elements) are first obtained as 128 orientation classes by statistic analysis on 2,000 file fingerprints. Then, 1.28 million 160×160 fingerprint patches with line noise added are used as training data, and their corresponding ground truth are labeled as the closest orientation class. After this, one AlexNet like network is employed to train the classification problem. Finally, the orientation patches estimated from CNN are quilted to be the final orientation field. With this estimated orientation field, latent fingerprint can be enhanced using Gabor filtering technique. Then, it requires re-estimating orientation again using the enhanced latent fingerprint. Finally, the re-estimated orientation field is used to enhance fingerprints and provide their final enhancement results.

Although this is the first work to apply CNN on latent fingerprint application, it has the following three major issues:

1. It only applies CNN to estimate the orientation field from the manner of classification task. It is still indirect to enhance fingerprints without making use of advantages of applying CNN directly on enhancement in a regression-like task.

2. Since it treats orientation estimation as the traditional classification problem. Whether 128 classes is enough to represent the whole orientation field space is an potential issue. If it is not enough, there will be potential issues when enhancing fingerprints using the estimated orientation field.

3. It requires re-estimate the whole orientation field again using the enhanced results from their network estimated orientation. This shows that the direct estimation from their network is not very good enough, that means their network training probably is not learning well enough.

5.3 Proposed FingerNet Method

5.3.1 Training Data Preparation

A big challenging to apply CNN on latent fingerprint application is indeed the lack of training data, especially for the latent case. The NIST SD27 latent dataset has only 258 latent fingerprint images, and is short of the correspondence between latent fingerprints' region of interest (ROI) and their true mates. Therefore, it is more appropriate to use NIST SD27 as the final inference rather than training straightly from it. Instead of NIST SD27, we use NIST SD4 to prepare training data but for the latent case. NIST SD4 has 2,000 rolled fingerprints with equally distributed fingerprint types (arch, tented arch, right loop, left loop and whorl). This special property helps to cover varieties of ridge patterns as much as possible when generating training data. Fig. 5.1 shows the illustration of the process to generate meaningful training data from NIST SD4 for latent fingerprint application. This procedure generally includes creating training patches, generating the corresponding enhancement ground truth image, and estimating the orientation ground truth map.

Augmented Fingerprint Patches

Instead of using the whole fingerprint image from NIST SD4 as one training sample, we create multiple overlapping 109×109 fingerprint patches as training samples. By overlapping cropping further at resolution of 61×61, the same fingerprint region appears in different sampling patches, which means different contextual information for it. This is indeed a data augmentation process that generates much more meaningful training data. Meanwhile, we also check reliability map of a fingerprint using algorithms in [46] to exclude the background region and low quality patches when creating the training data. A quality mask can be produced based on the reliability map with thresholding. In Fig. 5.1, the region circled in red shows bad quality and will be excluded when selecting training patches. This is important in order to avoid wrong ground truth generated because of poor quality fingerprint regions. By quality control, it ensures that the ground truth labels (can be enhancement images or orientation classes) estimated for both enhancement and orientation are reliable to supervise the training.

Ground Truth Labels

Fingerprint patches from Sec. 5.3.1 often have relative good quality (*e.g.*, clean ridge structure, few structured noise). This makes it possible to obtain reliable ground truth labels for both the enhancement and orientation estimation. We adopt the simple but classical Gradient-based method proposed in [46] to generate these ground truth labels.

With gradient of each pixel calculated, the orientation field can be estimated in the range of 0° to 180°. To eliminate possible negative influence of incorrect orientation estimation, we quantize each pixel's orientation with 9° as the quantization step. Thus, we finally use 20 classes as the ground truth labels of orientation map in pixel-wise manner.

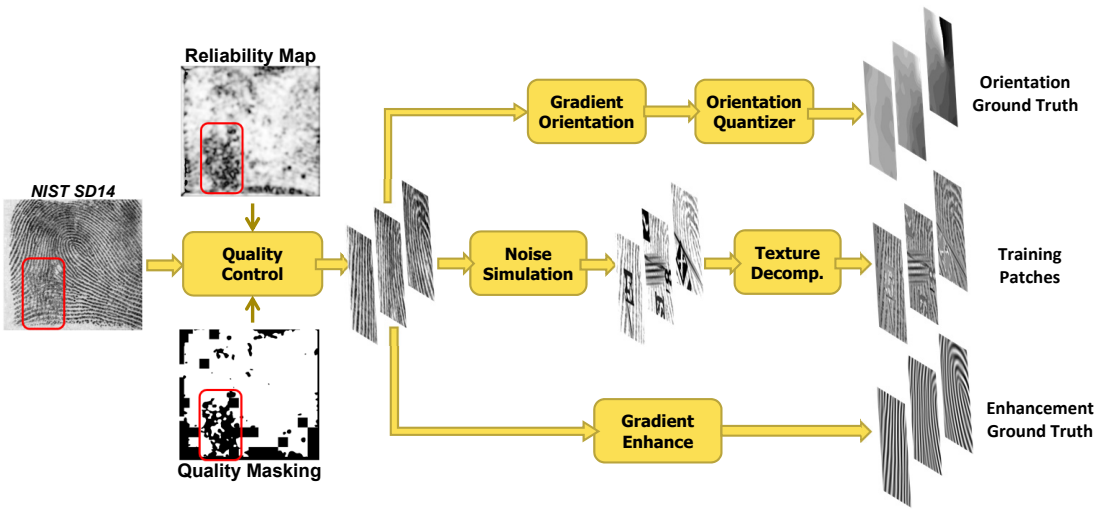


Figure 5.1: Illustration of training data preparation: overlapping patches in good quality are extracted firstly; Then, enhancement ground truth images are created from Gradient-based method; Orientation ground truth maps are generated based on Gradient orientation followed by quantizing; Finally, training patches are produced by a noise simulation module and texture decomposition.

About the enhancement ground truth, the proposed contextual filtering method with estimated orientation field in [46] is applied. The enhancement results from it are actually used as "ground truth enhancement" to be learned from the convolutional neural network.

Training Patches in Latent Case

Structured Noise Simulation: The structured noises are added on the patches of rolled fingerprints to simulate the latent situation, in order to generate training data of the network. Different from [11] that lacked of characters structured noise, we also distort rolled fingerprints with characters. Given a fingerprint patch P, its distorted version P_d with lines and characters can be obtained as following:

Lines: Due to strong similarity with fingerprint ridge structure, line-like noise makes algorithms tend to fail easily. We utilize the Gabor functions to simulate multiple sets of lines as [11]. One set of lines L is determined by one Gabor function with the corresponding parameters, and added to fingerprint patch P by

$$P_l = min(\alpha P, L), \tag{5.1}$$

where α controls the strength of fingerprint signal. The min operation is used here to have a better overlaying effect by adding line noise. We generate up to 3 sets of lines in random by a recursive version of Eq. (5.1).

Characters: Besides lines, another very common structured noise appearing in latent fingerprint is the text including various characters. This type of noise will cause wrong orientation estimation and enhancement if not handling well. Thus, we make use of the case-insensitive character dataset [1] from [50] to add character noise. This dataset includes 163k character images from ICDAR 2003, 2005, 2011, 2013 training images, KAIST and Chars74k with cropping to 24×24 characters. We overlay one fingerprint patch by random selecting characters with the expectation of one quarter region covered. Besides, we arrange the characters in a random direction and scale them to double size with probability 0.5. The final distorted patch P_d with characters as structured noise will be

$$P_d = min(P_l, P_c), \tag{5.2}$$

where P_c is the characters noise added and P_l is fingerprint patch distorted with lines.

[1] https://bitbucket.org/jaderberg/eccv2014_textspotting

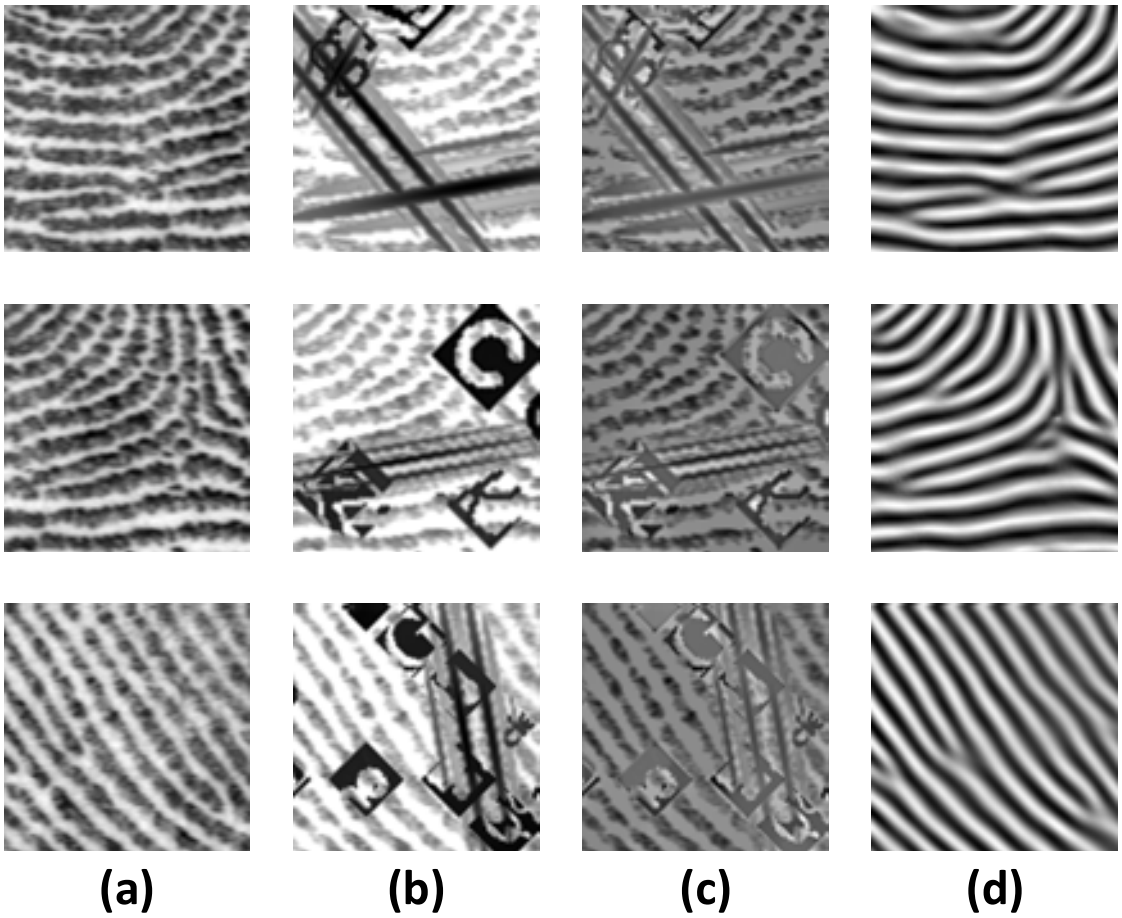


Figure 5.2: Examples showing how to prepare training data: (a) Original patch P; (b) Adding noise and weaken fingerprint strength to be distorted P_d; (c) Texture component P_d^t from (b); (d) Ground truth enhancement P_e from (a).

Pre-processing: Total variation (TV) decomposition has been proved to be a proper pre-processing for latent fingerprint [119, 12, 67]. A latent fingerprint can be decomposed into cartoon and texture components via TV decomposition. In general, the cartoon component consists of piece-wise smooth background noise while the texture component keeps oscillatory textured patterns (*e.g.*, fingerprint ridges). Although there are still structured noises contained in texture component, the strength of structured noises

is relatively suppressed comparing with original latent fingerprint. Therefore, texture component of a latent fingerprint is often used as the actual image to be enhanced.

Finally, better structured noise simulation for fingerprint, the TV decomposition preprocessing and the augmented latent patches, can help provide abundant training data to learn how to enhance fingerprint from the distorted patch to its enhanced patch in an end-to-end strategy. The following notations are used in this paper: original fingerprint patch P, its distorted version P_d, the texture component P_d^t decomposed from P_d, and the enhanced patch P_e directly applied on P, and the quantized orientation field P_{OF} estimated from P. In the training process, P_d^t will be used as input, while P_e and P_{OF} is the corresponding outputs of the network. Fig. 5.2 shows some examples of these patches.

5.3.2 Network Architecture

The proposed network architecture contains convolutional/deconvolutional layers, pooling/unpooling layers, and several skip connections. There is the ReLU operation after each convolutional/deconvolutional layer. Fig. 5.3 is an illustration about the detailed configuration of our network. To simplify the overall network structure, we have the same size $c = 64$ for channel size of each feature map, which is proved to be an effective in experiments.

Convolutional Layer

Cascading convolutional layers are fist applied to extract fingerprint features, which is an important motivation that we apply CNN on latent fingerprint applications. For computer vision problems, *e.g.*, semantic understanding, there are very successful low level features such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). However, it is short of superior low level features particular for latent fingerprints in the literature. Zhu *et al.* proposed a hand-engineered feature for

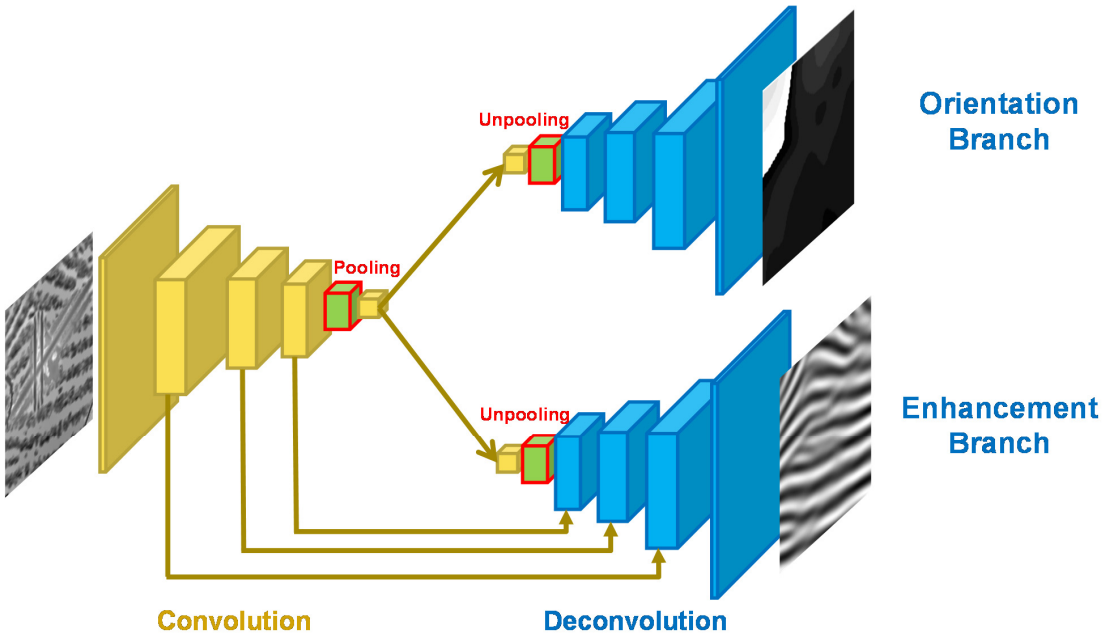


Figure 5.3: Illustration of the overall network architecture we proposed. The network contains one common convolution part and two deconvolution parts: orientation branch and enhancement branch. Skip connections are presented symmetrically to connect the convolution part and enhancement barnch.

exemplar fingerprint based on signal analysis [124]. However, it is not robust for latent situation because of appearance of the structured noise. Therefore,the trained deep features for latent fingerprint could play a crucial role as one important supplement in fingerprint community.

Deconvolutional Layer

Recently, deconvolution related techniques have been adopted into CNN to yield the Encoder-Decoder like network. It shows success on semantic segmentation [5, 85] and image restoration [81]. Deconvolution referred here can also be called "transpose convolution" or "fractal stride convolution", which is not the same as "inverse of convolution" defined in mathematics or signal processing. However, to be consistent and simplified, we keep using this term "deconvolution" in this work for convenience. For

latent fingerprint enhancement, the convolutional procedure will extract features that can eliminate the structured noise while keeping the fingerprint ridge content. Then, the deconvolution layers will recover image details using the extracted CNN features.

Indeed, deconvolution can also be treated as convolution. However, convolution bundles multiple inputs with a single output, yet deconvolution bundles one input with multiple outputs. Therefore, deconvolution layer is able to recover the resolution shrinking that happens during convolution procedure. For image processing applications, it is necessary to keep the same resolution for both the input and output images. Thus, many networks include image padding in order to keep the resolution [28, 60]. Since deconvolution has the property of enlarging resolution, it is possible to ignore padding during convolution, and rely on the deconvolution to recover the same image resolution. Note that we use the same caffe implementation as [5, 80, 81, 85] for deconvolution layers.

Pooling and Unpooling Layer

Most of the neural network for image processing applications (*e.g.* super-resolution, restoration and denoising), abandon pooling or unpooling layers in their network architecture. The major concern is that pooling will discard useful image details, that will be inappropriate for the aforementioned tasks.

However, we enable pooling and unpooling in our network structure based on several reasons. First, fingerprint enhancement is notably different from the low-level image processing task. The learning procedure need to distinguish fingerprint ridges and structured noise, which is indeed in semantic level of image content. Therefore, pooling can help remove noisy activations and obtain better abstraction. Second, pooling layer can increase the receptive field since the down-sampling property. Large receptive field is important for image processing since enough contextual information could be counted in. By pooling, the network architecture does not need go to very deep in order to have the same large receptive field. Third, even though it is possible that pooling will cause

image details lost, we can apply skip connection between input of pooling and output of unpooling to compensate the image details.

About the unpooling, we employ the similar unpooling operation in [85] since the spatial indexing can be used to reverse pixels back to their original location before pooling. This unpooling schema is able to keep spatial information when enhancing fingerprints through reconstruction, while expanding the resolution.

Skip Connection

Similar with [80], we also introduce skip connections in our proposed architecture. The major motivation is to compensate the possible details lost when having deep convolutions or pooling layers. Skip connections are occurring between two symmetrical convolution and deconvolution layers. Particularly, the input of pooling layer is connected with the output of unpooling layer, which will help much as to overcome possible details lost caused by pooling/unpooling operations. By connections, feature maps at different resolutions through convolutions can be passed on to the reconstruction/enhancement stages via deconvolutions in the forward procedure.

Additionally, gradients during the back-propagation procedure can directly flow to previous layers going through those connections. This is demonstrated to make training deeper architecture easier than without skip connections. Similar ideas are also applied successfully in highway network and residual network [44].

Receptive Field Analysis

In the proposed network structure, the first convolution layer conv1 has the filter size 9×9, and the second convolution layer conv2 uses 5×5 as filter size. Then, conv3 and conv4 have the same filter size 3×3. Between conv3 and conv4, there is the 2×2 max pooling layer. This forms the encoder part, and the decoder part has the symmetric settings according to it. Encoder part is for extracting features and then the receptive

field is a critical factor for it. Different from other networks, we also include $stride = 4$ for conv1 in order to widen the receptive field.

For each "pixel" in the feature maps, the receptive field can be calculated on different level or resolution. The feature map after conv1 has the 9×9 receptive field. Similarly, we can calculate that the receptive field sizes for layers after conv2, conv3, pooling and conv4, are 25×25, 33×33, 37×37 and 53×53 respectively. Then, one "pixel" in the final feature layer after conv4 is corresponding to a 53×53 patch of the fingerprint input. Usually, 16×16 pixels are used to compute one direction as an orientation element in the local orientation analysis. Therefore, one ''pixel" in the final feature maps is actually the abstraction of 9 orientation elements of the input fingerprint, which provides much contextual information. Due to skip connections, both coarse and fine resolutions are taking into account to restore an enhanced fingerprint.

Multi-Task Learning

Traditionally as mentioned in Section 5.2, researchers directly target on the orientation field estimation and then apply it to enhance latent fingerprints. Similarly but in a different way, we integrate fingerprint orientation information in a multi-task learning manner. As shown in Fig. 5.3, there are two tasks learning independently with sharing the common convolution features. One branch is for the target of directly enhancement, yet another branch is to predict the orientations pixel-wisely. With the guidance of orientation branch, the enhancement branch can be regularized in a better strategy with orientation constraints.

For the orientation branch, we adopt a coarse orientation learning through orientation quantization. Instead of using accurate orientation in the range of $180°$, we quantize pixel orientations into 20 groups where each group is a class having $9°$. This converts the orientation branch to the classification problem which has 20 classes. This quantization procedure employing classification for orientation branch is helpful and necessary to the

whole learning process. First, orientation itself is higher-level semantic clue for fingerprint. Thus, it is natural to make use of classification for orientation branch. Second, with large contextual region, a coarse estimation is good enough to eliminate structured noise pollution to a certain degree. Moreover, this can make learning easier with relative limited class number on the other hand.

With sharing the convolutional layers in common, representative deep features are expected to perform good in both fingerprint enhancement and orientation estimation tasks. Therefore, multi-task learning from end to end has the capability to abstract more meaningful features than single task. We also evaluate this in the experimental parts.

5.3.3 Training and Inference

The loss functions in our proposed method include two objectives: the mean squared error/euclidean loss for enhancement branch L_{enh} and the multinomial logistic loss for the orientation branch L_{OF}. Given fingerprint patches $\{P_i\}_{i=1}^{N}$, we have $\{P_d^{t(i)}, P_e^{(i)}, P_{OF}^{(i)}\}_{i=1}^{N}$ as the texture component of the distorted patch, the direct enhancement of P_i, and the quantized orientation field of P_i, respectively. Then, the final loss function will be:

$$\begin{aligned} L\{P_d^t, P_e, P_{OF}\} = w_{enh} L_{enh}\{P_d^t, P_e\} \\ + w_{OF} L_{OF}\{P_d^t, P_{OF}\}, \end{aligned} \tag{5.3}$$

where w_{enh}, w_{OF} is the balance loss weights for enhancement branch and orientation branch, respectively. We set them to 1 in the experiments.

The encoding part extracts fingerprint features by convolution layers, then two mappings learned are able to recover an enhanced fingerprint image through enhancement branch, and to provide a coarse estimation on orientation through orientation branch. Although the orientation prediction is in a coarse manner after quantization, it helps guide the enhancement branch with semantic contextual clues. During training, we

have skip connections for the enhancement branch in order to compensate image details. However, we treat the orientation branch in a more traditional way without skip connections.

About the training strategy for multi-task, we apply two stage training. That is, we first train the single enhancement task independently. Then, with this pre-trained model as initial of the enhancement model, the whole multi-task network is trained again from end to end. We take advantage of the Caffe framework developed by Berkeley Vision and Learning Center (BVLC) [2] to have the implementation. The euclidean loss and softmax loss are applied for enhancement branch and orientation branch, respectively. We use Adam solver with base learning rate of 10^{-4} to minimize the loss function by back propagation, since Adam optimizing converges faster than the popular stochastic gradient descent (SGD) method [80].

Given a latent fingerprint image, we can forward the entire image into our network to generate the enhanced fingerprint. The network is trained based on smaller resolution, however, this fully convolution and deconvolution network makes it possible to fit images with arbitrary resolution.

5.4 Experimental Results

In this section, we show the experimental results and analysis conducted on latent fingerprint dataset. First, fingerprints datasets used in this work are introduced. Then, the experimental analysis including deep features for fingerprints, multi-task learning, and residual learning are presented. We also compared with other algorithms in the literate to show the effectiveness and robustness of our proposed final ensemble results. Finally,

[2] http://caffe.berkeleyvision.org/

the computation and speed analysis are stated to show the real-time capability of our method.

5.4.1 Experimental Setup

Datasets and Evaluation

In this chapter, NIST Special Database 4 (SD4), NIST Special Database 14 (SD14), and NIST Special Database 27 (SD27) are mainly used to conduct the related experiments.

As mentioned in Sec. 5.3.1, NIST SD4 is used to generate training data for the proposed network. This dataset has 2,000 rolled fingerprint images with equally distributed fingerprint types, that makes it rich and representative to generate training data. There are 400 fingerprint images for each of these five types: whorl, left loop, right lop, arch and tented arch. We finally generate 224,860 patches at resolution of 109×109 to train the network, in which one fifth portion is for validation purpose and the rest is for training.

NIST SD27 is the public latent fingerprint dataset that is used to evaluate algorithms for latent fingerprint applications such as enhancement, orientation estimation and fingerprint identification. This dataset is composed of 258 latent fingerprint images and their corresponding rolled fingerprints as true mates. Latent fingerprints in this dataset are collected from crime scene and distorted with low quality issue in different degree. Fingerprint experts classify them into three categories based on overall quality: "Good" with 88 fingerprints, "Bad" with 85 fingerprints and "Ugly" with 85 fingerprints. The noises include different degradation types especially structured noise such as lines, hand written/printed letters, stains, etc. Besides, fingerprints appear with overlapping and in partial are also presented much in this dataset. These facts make latent fingerprint enhancement a very challenging problem in order to increase identification rate.

NIST SD14 is used as additional background dataset in order to show robustness of algorithms when perform matching. It has $27,000$ rolled fingerprints and will be mixed with the 258 true mates to make fingerprint identification much more challenging. Note that we only add NIST SD14 in Sec. 5.4.5 when comparing with state-of-the-art algorithm in the final matching evaluation. For other experiments, NIST SD14 is not included in order to analyze the proposed method better with different settings.

The performance evaluation for latent fingerprint enhancement is based on the final identification rate using the enhanced fingerprints. Although this is an indirect way to evaluate the enhancement performance, it is an objective metric to fit the ultimate goal of enhancing fingerprints. We evaluate the matching performance using one commercial fingerprint software that is well-known as VeriFinger SDK 6.2 version. Cumulative Match Characteristic (CMC) curve is employed to compare the identification performance of fingerprint matching. This curve plots the rank-k identification rate when varying values of k, usually from 1 to 20 as a common setting. As the most widely used measurement, we conduct the performance comparison by showing the CMC curves similar with others.

Besides, all experiments are based on the identification scoring of the enhanced latent fingerprints matching with the original rolled fingerprints without enhancement. This is to make the comparison fair with other algorithms.

5.4.2 CNN Feature and Validation Result

Result analysis is firstly conducted on the validation dataset visually. For the validation set, the patches are from NIST SD4 and follow the same procedure of data processing in 5.3.1. However, no validation data is used for training but only for the testing purpose. We visualize features from CNN for some patches and also the enhancement results performed on validation set.

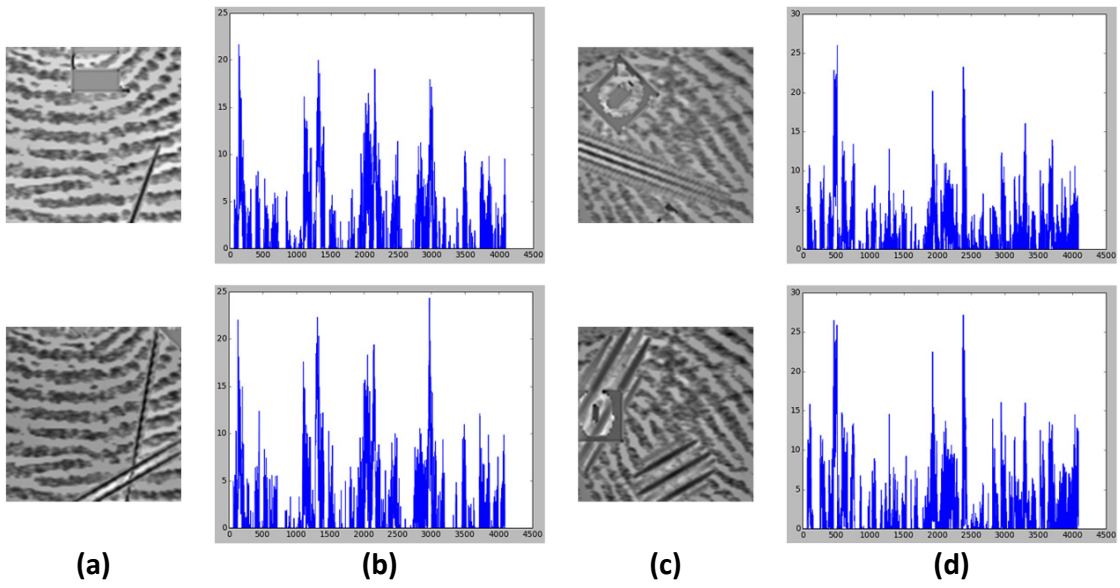


Figure 5.4: CNN features after last convolution conv4: (a) Same patches with different structured noise added; (b) CNNN features for the corresponding patches; (c) and (d) are another set of examples.

Fig. 5.4 shows the CNN features after last convolution layer (conv4) for fingerprint patches with different structured noise added. Here, the CNN features are extracted after the convolutional part but before the deconvolutional part. It can be shown that the CNN features are very similar for the same fingerprint patch, even though distorted with very different structured noise. It demonstrates the robustness and effectiveness of CNN features in terms of describing latent fingerprint with eliminating influence of structured noise.

Additionally, we perform enhancement over the validation set to compare with the ground truth enhancement, as shown in Fig. 5.5. Comparing the enhancement results in Fig. 5.5 (d) with the ground truth enhancement in Fig. 5.5 (e) visually, FingerNet shows its strength to remove structured noise and achieve close enhancing effect with the ground truth enhancement.

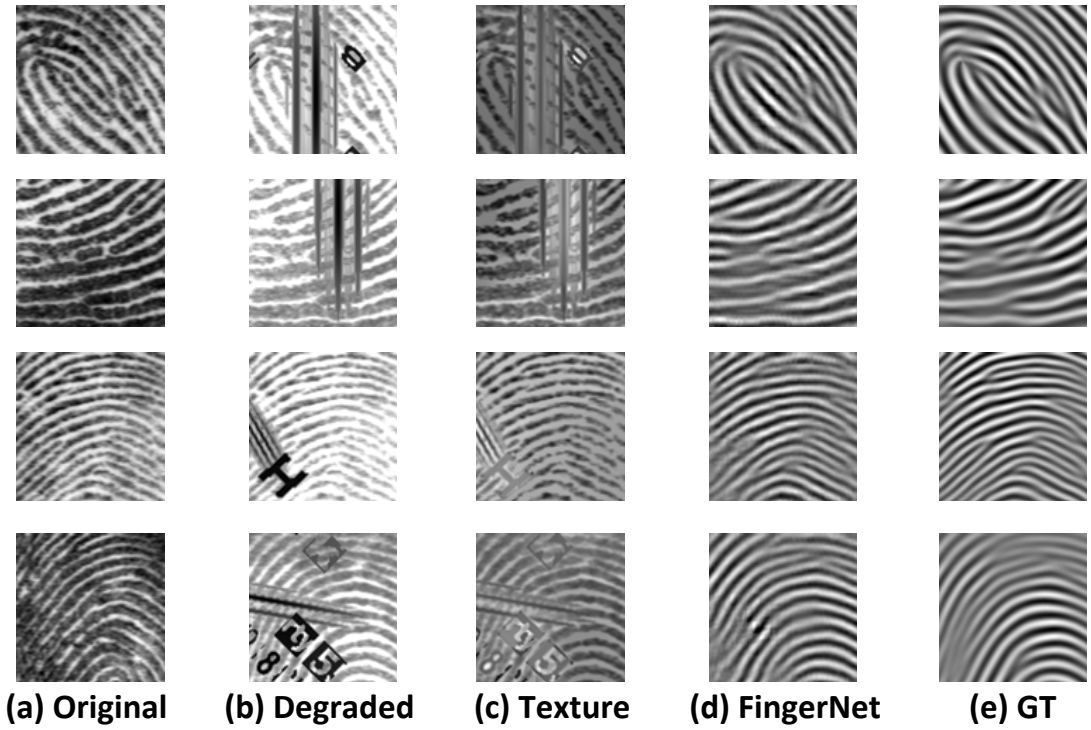


Figure 5.5: Enhancement example results over validation set: (a) Original fingerprint patch; (b) The degraded patch with structured noise; (c) Texture component from (b); (d) Enhancement result using FingerNet; (e) Ground truth enhancement directly from (a).

5.4.3 Single-task or Multi-task

We then evaluate the effectiveness with multi-task learning compared with single-task. By single-task, it means that we only consider the enhancement branch and remove the orientation branch. This can still provide reasonable enhancement results. Moreover, when conducting multi-task learning, there are two learning ways: one is to learn the entire network from scratch and the other one is to utilize the pre-trained model from single-task to fine tune the whole multi-task network. Two learning strategies are also compared in this section. Fig. 5.6 shows the CMC curves of the above mentioned experiments. Fig. 5.7 shows some enhancement examples of different models so that visually difference can also be compared.

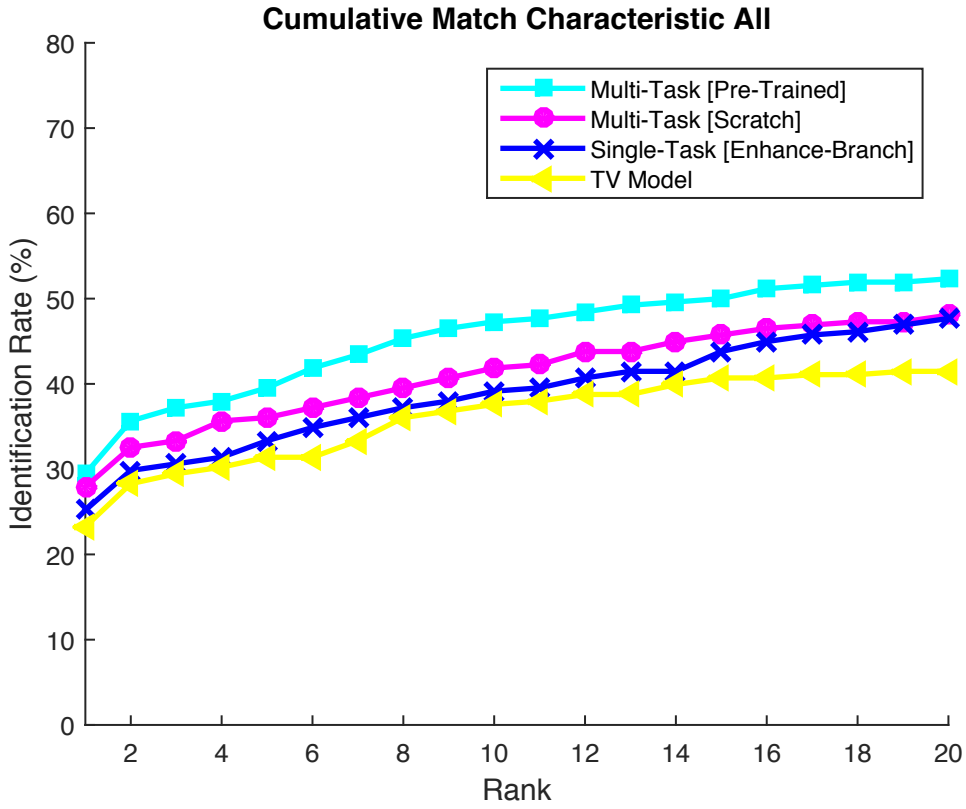


Figure 5.6: CMC curves comparing single-task, multi-task training from scratch and multi-task learning with pre-trained model fine tuning

As we can see from Fig. 5.6, all the proposed CNN models perform better than the benchmark TV model. Multi-task models perform better than the single enhancement task model. This is expected since the orientation branch can play the role as regulating the enhancement. It shows the effectiveness on helping latent fingerprint enhancement with the orientation branch guidance.

Comparing two learning strategies for multi-task models, it can be shown that network fine-tuning performs better than training from scratch. This is consistent with the conclusion of some other multi-task applications. The reason is that these two tasks share the common convolution features, and will try to make a trade-off between two branches when training together. Then, the optimum solution maybe trapped in some

local minimum earlier to have a good balance between two tasks. However, with the pre-trained model from single enhancement task, the optimized parameters are already in a good point for enhancement. Starting from this point, both the two tasks learning will contribute a better solution that will slightly tend to the enhancement task.

Fig. 5.7 shows some enhancement examples of different models so that visually difference can also be compared. Red circles mark regions where multi-task model trained from scratch performs better than single enhancement task. Visually observing from Fig. 5.7, we can see that some ambiguous regions for single-task model can be improved by multi-task models, especially the model with pre-trained fine tuning. Additionally, we can see that multi-task model with fine-tuning can perform better enhancement around some curvature regions as circled in green color.

5.4.4 Residual Learning or Non-Residual Learning

Learning the residual between the input and output is proved to be efficient to improve super resolution performance. That means, it is not directly learning the output supervised by ground truth image, yet the residual to compensate the input is learned instead. This strategy is reasonable since low resolution image lacks of image details compared with the high resolution image. Naturally, residual learning is a good way to compensate this lost of details.

Similarly, we explore whether learning the residual helpful for latent fingerprint enhancement in our experiments. The comparison experiment is designed with exactly the same network configuration such as filter number, filter size, and network depth in order to have a fair comparison. Additionally, we compare the two multi-task learning strategies (learn with pre-trained model and learn directly from scratch) when evaluating the influence of residual learning. This will make our analysis more robust and avoid bias. The only difference in the comparing experimental pair is that one model

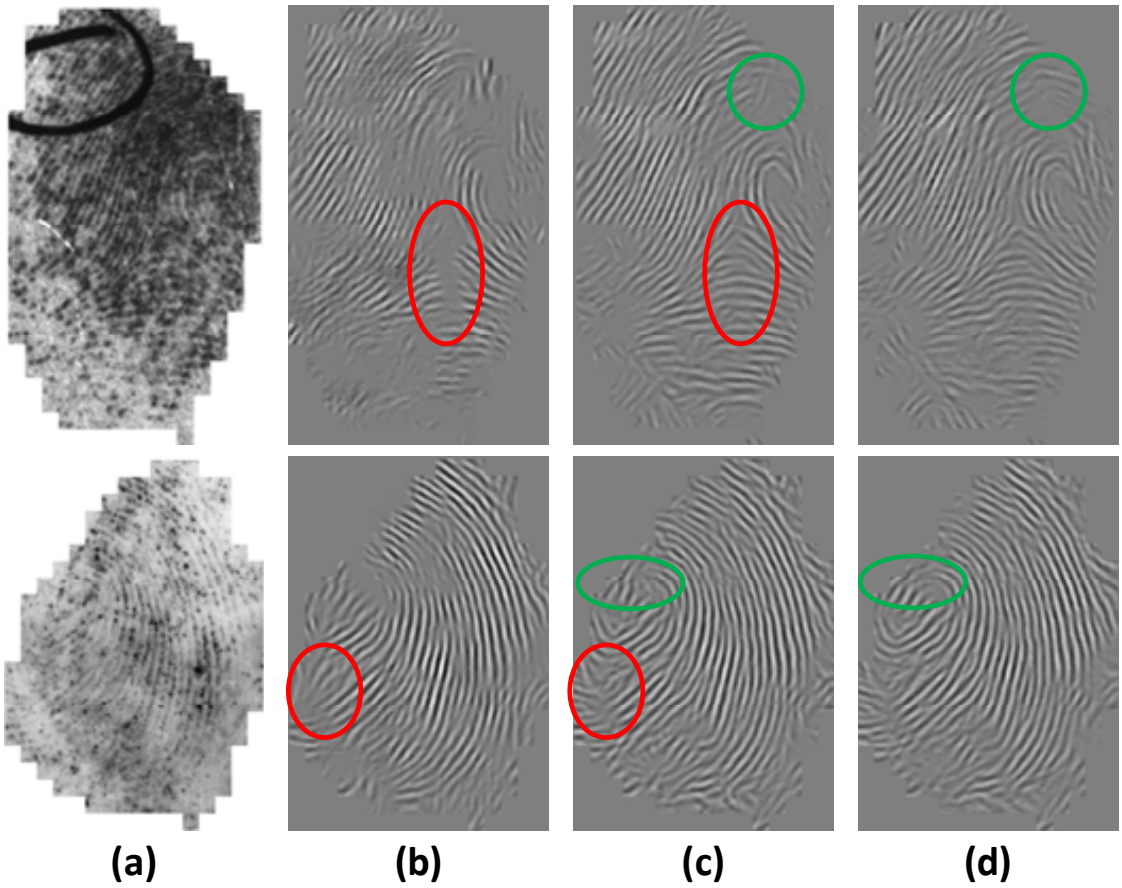


Figure 5.7: Visual examples to show multi-task better performance: (a) original latent image; (b) enhancement from single-task model; (c) enhancement from multi-task training from scratch; (d) enhancement via multi-task training from fine tuning pre-trained model

applies residual learning and the other not. Figure 5.8 shows the CMC curves for their corresponding matching performance.

From Fig. 5.8 (a), it can be shown that the non-residual learning models achieve better performance overall than models with residual learning, for both pre-trained fine-tuning and training from scratch. If we explore more into different quality categories, we can see that sometimes involving residual learning can result in slightly better performance than non-residual learning, especially for "Good" quality category as shown in (b). For "Bad" and "Ugly" quality categories in (c) and (d), the trend is similar with

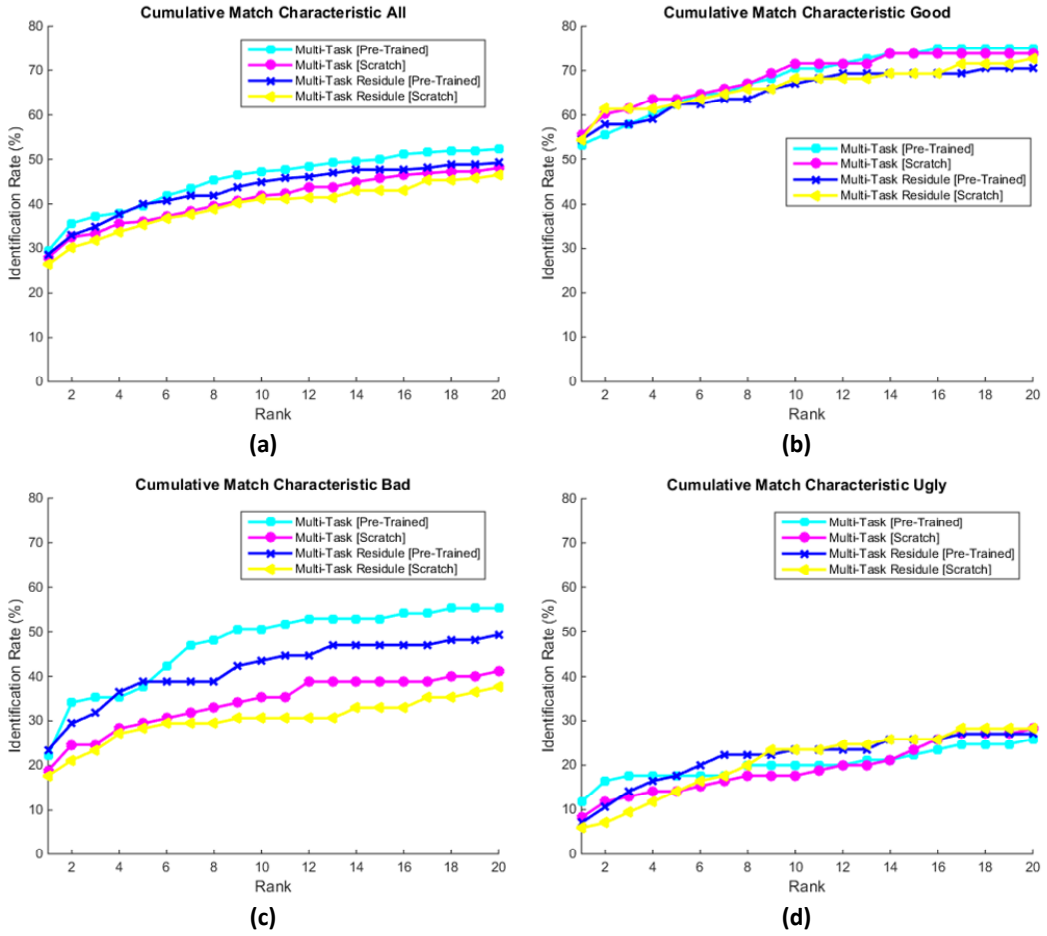
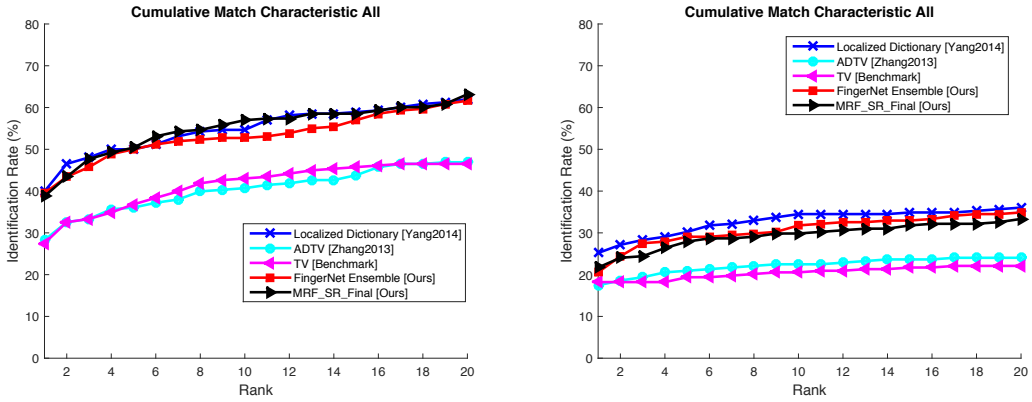


Figure 5.8: CMC curves comparing Multi-Task with residual learning and without residual learning, in both pre-trained fine-tuning and from scratch.

the overall performance in (a). Residual learning means adding input to the network forward prediction to generate final prediction. However, for "Bad" and "Ugly" quality latent fingerprints, adding input back makes the network hard to learn its residual due to the highly distorted property of the input. For "Good" quality images, it is reasonable since the similarity between input and output of the network is higher, which is more like the super resolution problem. However, a balance among different quality categories should be considered. Thus, we will apply non-residual learning since most of the latent fingerprints have poor quality. This also indicates that we should design



(a) (b)

Figure 5.9: Comparison with other algorithms: Localized Dictionary[Yang2014], TV[benchmark] and ADTV[Zhang2013]. Note that Localized Dictionary is the state-of-the-art method. (a) is the CMC curves without including NIST SD14 as background noise images. (b) is the CMC curves including NIST SD14 as background noise images.

the network particular fitting for latent fingerprint applications, but not directly applying existing algorithms for other vision tasks such as image super resolution.

5.4.5 Compare with State-of-the-Art Methods

Our proposed FingerNet can generate enhancement result directly as output of the network. The final matching performance of FingerNet system is evaluated by forwarding with four different convergence stages and then averaging to get the final CMC curve. Here we compare the proposed results with state-of-the-art method [113]. Note that NIST SD14 is included as background noise images to make the identification much more challenging in this subsection.

Figure 5.9 shows the comparison results with different methods when evaluating with background images and not. Comparing with both TV and ADTV methods, the proposed network outperforms them in large margin. Moreover, the proposed method outperforms the state-of-the-art method [113], especially on the top-1 rank identification

Method	Inference Speed
Localized Dict. [Yang2014]	8.6
ADTV [Zhang2013]	58.2
MRF-SR [Ours]	50.3
FingerNet [Ours]	0.7

Table 5.1: Computation Speed Comparison (in Seconds)

rate if no background images involved. When adding background images, performance of all algorithms drop significantly. However, the proposed method have similar performance especially when rank goes higher than 3. Rank-1 identification rate is important but not the only factor to evaluate a system. Good identification rate in higher rank means that human labor can be reduced a lot when checking further by human experts.

Compared with our proposed MRF-SR in Chapter 3, we can see similar performance observed about the CMC curves. This indicates that the proposed FingerNet can be used as a good alternative of MRF-SR, which needs high cost in computation. This is similar with some research observations that CNN approach can replace traditional methods or traditional methods can be interpreted using CNN framework such as [28].

Moreover, one big advantage of the proposed FingerNet is that the inference speed for each latent fingerprint is very fast. The proposed FingerNet can be trained in one day on GPU Titan X. Since we have pooling involved in our network, the inference speed is much faster than without pooling. We compare the computation speed for fingerprints in 800×760 as shown in Table 5.1. The computation times is for one fingerprint in the unit of seconds.

As we can see that the proposed method has superior advantage in inference speed which can make the matching time much more efficient, especially when having larger scale of fingerprints. Due to the property of calculations in convolution/deconvolution,

paralleling computing make it possible to have fast computation speed up. It also makes the ensemble of different network possible to improve the matching performance.

5.5 Conclusion

In this chapter, we present the CNN based method that is FingerNet to solve latent fingerprint enhancement problem. We design the network particular for latent fingerprint applications. FingerNet includes an encoding convolutional part and two decoding deconvolutional parts. Multi-task learning is applied to improve the performance. We also investigate several implementation details such as training multi-task efficiently by fine tuning, whether residual learning helps, and compare with different methods. We have demonstrated that the proposed method can outperform the existing algorithms with no background fingerprints involved, additionally with much faster inference speed than others. For possible future work, more network structures could be studied and how to train more efficiently based on fingerprint property could also be another interesting extension.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Although the development for the automated fingerprint identification system is rapid, the identification rate for latent fingerprint matching is still far from expectation and its practical application is still limited. This dissertation offers a thorough research on latent fingerprint applications, especially on pre-processing for matching, latent fingerprint enhancement and orientation field estimation. We provide solutions using a wide range of techniques, varying from the traditional image processing techniques to modern CNN-based methods.

The proposed MRF-SR method investigated the combination of TV models, ridge dictionary types, and the local patch quality assessment guidance to enhance latent fingerprints. Without the need of estimating the orientation field and the fingerprint pose, a direct enhancement result was developed for fingerprint matching. It was shown by experimental results that the MRF-SR method can remove structured noise, recover low quality regions and improve the identification rate. It was shown by experiments that the MRF-SR method can boost the identification rate significantly.

The MRF-SR method was generalized with some modification by including the orientation field estimation. Furthermore, a supervised-learning-based fusion strategy was adopted to provide better performance. This strategy was inspired by the fact that different methods concentrate on different aspects of fingerprint orientation estimation. Decisions from different experts are treated as feature vectors, and multiple techniques

such as linear regression, SVM regression and the neural network fitting can be used to train the model. The performance on the test dataset showed the effectiveness of the decision fusion scheme.

Finally, we developed an CNN-based solution called the FingerNet for latent fingerprint enhancement. The FingerNet was trained using a pixelwise end-to-end manner. It included a convolutional network as the encoder and two deconvolutional networks as the decoder using a multi-task learning strategy to boost the overall performance. We also investigated techniques such as network fine-tuning, residual learning, etc. It was shown by experimental results that the proposed FingerNet outperforms benchmarking algorithms with a fast computational speed.

6.2 Future Work

Although several issues about latent fingerprint enhancement and orientation field have been studied in this dissertation, there are still some open problems for future investigation.

- **Automated Latent Fingerprint Segmentation.** This is one important research topic in latent fingerprint community. Most algorithms on latent fingerprint enhancement and orientation field estimation are proposed based on the assumption that fingerprint segmentation is given (e.g., directly use the ground truth for NIST SD27). This assumption is adopted since it is a good starting point for researchers and people can focus on actual enhancement or orientation field estimation techniques. Manually providing fingerprint foreground mask is easier than minutiae extraction by human experts, which demands much less human labor. However, it is still worthwhile to develop algorithms for automated fingerprint segmentation. Both the MRF-SR method and the FingerNet cen be potentially

tailored to this purpose. For example, for the MRF-SR method, since the background in latent fingerprints usually need a higher number of atoms than the foreground region in it reconstruction, it is possible to expand the MRF-SR method to fingerprint segmentation. The FingerNet can provide robust and effective features for fingerprints. Thus, it can learn to predict the fingerprint background and foreground.

- **Contactless 3D Fingerprint Identification.** Recently, there is rapid development in contactless fingerprint systems due to the high demand of security on mobile applications. Contactless fingerprint ridge patterns can be acquired by cameras without any physical contact between fingers and the sensor surface. Conventional exemplar fingerprints can be easily degraded in quality due to facts such as skin deformation, finger moisture, dry fingers, and even sensor noise. Contactless 2D fingerprint images often have even lower quality as compared with traditional fingerprints. To make contactless fingerprint systems practical, several contactless 3D fingerprint recognition techniques were introduced recently [3, 88, 104]. Many challenges, including 3D fingerprint acquisition, representation fingerprint, feature extraction, are not well addressed at this point. This is expected to be a main topic on fingerprint research in the near future.

Bibliography

[1] M. Aharon, M. Elad, and A. Bruckstein. k -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov 2006.

[2] A. Almansa and T. Lindeberg. Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale selection. *Image Processing, IEEE Transactions on*, 9(12):2027–2042, 2000.

[3] S. S. Arora, K. Cao, A. K. Jain, and N. G. Paulter. 3D fingerprint phantoms. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 684–689. IEEE, 2014.

[4] D. R. Ashbaugh. *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. CRC press, 1999.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.

[6] A. M. Bazen and S. H. Gerez. Segmentation of fingerprint images. In *Proc. Workshop on Circuits Systems and Signal Processing (ProRISC 2001)*, volume 276280. Citeseer, 2001.

[7] A. M. Bazen and S. H. Gerez. Systematic methods for the computation of the directional fields and singular points of fingerprints. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):905–919, 2002.

[8] A. Blake, P. Kohli, and C. Rother. *Markov random fields for vision and image processing*. Mit Press, 2011.

[9] A. Buades, T. M. Le, J.-M. Morel, L. Vese, et al. Fast cartoon+ texture image filters. *Image Processing, IEEE Transactions on*, 19(8):1978–1986, 2010.

[10] X. Can and Y. Lin. An adaptive algorithm for smoothing fingerprint orientation fields. In *Computational Intelligence and Natural Computing, 2009. CINC'09. International Conference on*, volume 1, pages 70–72. IEEE, 2009.

[11] K. Cao and A. K. Jain. Latent orientation field estimation via convolutional neural network. 2015.

[12] K. Cao, E. Liu, and A. Jain. Segmentation and enhancement of latent fingerprints: A coarse to fine ridgestructure dictionary. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(9):1847–1859, Sept 2014.

[13] R. Cappelli, A. Lumini, D. Maio, and D. Maltoni. Fingerprint classification by directional image partitioning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):402–421, 1999.

[14] R. Cappelli, D. Maio, and D. Maltoni. Semi-automatic enhancement of very low quality fingerprints. In *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*, pages 678–683. IEEE, 2009.

[15] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[16] C. Champod, C. J. Lennard, P. Margot, and M. Stoilovic. *Fingerprints and other ridge skin impressions*. CRC press, 2004.

[17] T. F. Chan and S. Esedoglu. Aspects of total variation regularized l 1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[18] C. Chen, J. Feng, and J. Zhou. Multi-scale dictionaries based fingerprint orientation field estimation. In *Biometrics (ICB), 2016 International Conference on*, pages 1–8. IEEE, 2016.

[19] F. Chen, J. Feng, A. K. Jain, J. Zhou, and J. Zhang. Separating overlapped fingerprints. *Information Forensics and Security, IEEE Transactions on*, 6(2):346–359, 2011.

[20] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9):1519–1524, 2006.

[21] X. Chen, J. Tian, J. Cheng, and X. Yang. Segmentation of fingerprint images using linear classifier. *EURASIP Journal on Applied Signal Processing*, 2004:480–494, 2004.

[22] S. Chikkerur, A. N. Cartwright, and V. Govindaraju. Fingerprint enhancement using stft analysis. *Pattern Recognition*, 40(1):198–211, 2007.

[23] H. Choi, M. Boaventura, I. A. Boaventura, and A. K. Jain. Automatic segmentation of latent fingerprints. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 303–310. IEEE, 2012.

[24] S. A. Cole et al. *Suspect identities: A history of fingerprinting and criminal identification*. Harvard University Press, 2009.

[25] J. Dai, K. He, Y. Li, S. Ren, and J. Sun. Instance-sensitive fully convolutional networks. *arXiv preprint arXiv:1603.08678*, 2016.

[26] J. Dai, Y. Li, K. He, and J. Sun. R-fcn: Object detection via region-based fully convolutional networks. *arXiv preprint arXiv:1605.06409*, 2016.

[27] S. C. Dass. Markov random field models for directional field and singularity extraction in fingerprint images. *Image Processing, IEEE Transactions on*, 13(10):1358–1367, 2004.

[28] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014.

[29] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[30] I. E. Dror, K. Wertheim, P. Fraser-Mackenzie, and J. Walajtys. The impact of human–technology cooperation and distributed cognition in forensic science: Biasing effects of afis contextual information on human experts*. *Journal of forensic sciences*, 57(2):343–352, 2012.

[31] V. Dvornychenko and M. D. Garris. *Summary of NIST latent fingerprint testing workshop*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2006.

[32] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[33] M. Elad, M. A. Figueiredo, and Y. Ma. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*, 98(6):972–982, 2010.

[34] E. Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM report*, 9:31, 2009.

[35] FBI. The FBI's next generation identification (NGI). `http://www.fbi.gov/about-us/cjis/fingerprintsbiometrics/ngi`.

[36] J. Feng, J. Zhou, and A. K. Jain. Orientation field estimation for latent fingerprint enhancement. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(4):925–940, 2013.

[37] FVC. Fvc2002. `http://bias.csr.unibo.it/fvc2002/`.

[38] S. Greenberg, M. Aladjem, D. Kogan, and I. Dimitrov. Fingerprint image enhancement using filtering techniques. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 3, pages 322–325. IEEE, 2000.

[39] S. E. Grigorescu, N. Petkov, and P. Kruizinga. Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions on*, 11(10):1160–1167, 2002.

[40] J. Gu, J. Zhou, and C. Yang. Fingerprint recognition by combining global structure and local cues. *Image Processing, IEEE Transactions on*, 15(7):1952–1964, 2006.

[41] J. Gu, J. Zhou, and D. Zhang. A combination model for orientation field of fingerprints. *Pattern Recognition*, 37(3):543–553, 2004.

[42] L. Haber and R. N. Haber. Error rates for human latent fingerprint examiners. In *Automatic fingerprint recognition systems*, pages 339–360. Springer, 2004.

[43] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[44] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[45] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. *arXiv preprint arXiv:1603.05027*, 2016.

[46] L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: algorithm and performance evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8):777–789, 1998.

[47] S. Huckemann, T. Hotz, and A. Munk. Global models for the orientation field of fingerprints: an approach based on quadratic differentials. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(9):1507–1519, 2008.

[48] M. Indovina, R. Hicklin, and G. Kiebuzinski. Evaluation of latent fingerprint technologies: Extended feature sets. Technical report, Technical Report NISTIR 7775, NIST, 2011.

[49] M. D. Indovina, V. Dvornychenko, E. Tabassi, G. Quinn, P. Grother, S. Meagher, M. Garris, et al. *ELFT Phase II: An Evaluation of Automated Latent Fingerprint Identification Technologies*. US Department of Commerce, National Institute of Standards and Technology, 2009.

[50] M. Jaderberg. *Deep learning for text spotting*. PhD thesis, University of Oxford, 2015.

[51] A. K. Jain and J. Feng. Latent palmprint matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(6):1032–1047, 2009.

[52] A. K. Jain and J. Feng. Latent fingerprint matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(1):88–100, 2011.

[53] X. Jiang. Fingerprint image ridge frequency estimation by higher order spectrum. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 1, pages 462–465. IEEE, 2000.

[54] X. Jiang, M. Liu, and A. C. Kot. Fingerprint retrieval for identification. *Information Forensics and Security, IEEE Transactions on*, 1(4):532–542, 2006.

[55] S. Jirachaweng, Z. Hou, W.-Y. Yau, and V. Areekul. Residual orientation modeling for fingerprint enhancement and singular point detection. *Pattern Recognition*, 44(2):431–442, 2011.

[56] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1233–1258, 1987.

[57] T. Kamei. Image filter design for fingerprint enhancement. In *Automatic Fingerprint Recognition Systems*, pages 113–126. Springer, 2004.

[58] S. Karimi-Ashtiani and C.-C. Kuo. A robust technique for latent fingerprint image segmentation and enhancement. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1492–1495. IEEE, 2008.

[59] M. Kass and A. Witkin. Analyzing oriented patterns. *Computer vision, graphics, and image processing*, 37(3):362–385, 1987.

[60] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. *arXiv preprint arXiv:1511.04587*, 2015.

[61] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. *arXiv preprint arXiv:1511.04491*, 2015.

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[63] C.-C. J. Kuo. Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation*, 2016.

[64] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2006.

[65] S. Z. Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[66] X.-C. Lian, Z. Li, C. Wang, B.-L. Lu, and L. Zhang. Probabilistic models for supervised dictionary learning. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2305–2312. IEEE, 2010.

[67] M. Liu, X. Chen, and X. Wang. Latent fingerprint enhancement via multi-scale patch based sparse representation. *Information Forensics and Security, IEEE Transactions on*, 10(1):6–15, Jan 2015.

[68] M. Liu, X. Jiang, and A. C. Kot. Fingerprint reference-point detection. *EURASIP J. Adv. Sig. Proc.*, 2005(4):498–509, 2005.

[69] M. Liu, X. Jiang, and A. C. Kot. Efficient fingerprint search based on database clustering. *Pattern Recognition*, 40(6):1793–1803, 2007.

[70] M. Liu and P.-T. Yap. Invariant representation of orientation fields for fingerprint indexing. *Pattern Recognition*, 45(7):2532–2542, 2012.

[71] S. Liu and M. Liu. Fingerprint orientation modeling by sparse coding. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 176–181. IEEE, 2012.

[72] Z. Liu, X. Li, P. Luo, C. C. Loy, and X. Tang. Deep learning markov random field for semantic segmentation. *arXiv preprint arXiv:1606.07230*, 2016.

[73] P. Z. Lo and Y. Luo. Method and apparatus for adaptive hierarchical processing of print images, July 11 2006. US Patent App. 11/456,622.

[74] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[75] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[76] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *Image Processing, IEEE Transactions on*, 17(1):53–69, 2008.

[77] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

[78] D. Maltoni. A tutorial on fingerprint recognition. In *Advanced Studies in Biometrics*, pages 43–68. Springer, 2005.

[79] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[80] X.-J. Mao, C. Shen, and Y.-B. Yang. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*, 2016.

[81] X.-J. Mao, C. Shen, and Y.-B. Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *arXiv preprint arXiv:1603.09056*, 2016.

[82] K. A. Nagaty. Fingerprints classification using artificial neural networks: a combined structural and statistical approach. *Neural Networks*, 14(9):1293–1305, 2001.

[83] NIST. Nist special database 14. `http://www.nist.gov/srd/nistsd14.cfm`.

[84] NIST. Nist special database 27. `http://www.nist.gov/srd/nistsd27.cfm/`.

[85] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. *arXiv preprint arXiv:1505.04366*, 2015.

[86] L. O'Gorman and J. V. Nickerson. An approach to fingerprint filter design. *Pattern recognition*, 22(1):29–38, 1989.

[87] M. Oliveira and N. J. Leite. A multiscale directional operator and morphological tools for reconnecting broken ridges in fingerprint images. *Pattern Recognition*, 41(1):367–377, 2008.

[88] G. Parziale, E. Diaz-Santana, and R. Hauke. The surround imagertm: A multi-camera touchless device to acquire 3d rolled-equivalent fingerprints. In *Advances in Biometrics*, pages 244–250. Springer, 2005.

[89] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

[90] S. Prabhakar et al. Probabilistic orientation field estimation for fingerprint enhancement and verification. In *Biometrics Symposium, 2008. BSYM'08*, pages 41–46. IEEE, 2008.

[91] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[92] S. Setzer. Split bregman algorithm, douglas-rachford splitting and frame shrinkage. In *Scale space and variational methods in computer vision*, pages 464–476. Springer, 2009.

[93] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1605.06211, 2016.

[94] B. Sherlock, D. Monro, and K. Millard. Fingerprint enhancement by directional fourier filtering. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 141, pages 87–94. IET, 1994.

[95] B. G. Sherlock and D. M. Monro. A model for interpreting fingerprint topology. *Pattern recognition*, 26(7):1047–1055, 1993.

[96] N. J. Short, M. S. Hsiao, A. L. Abbott, and E. A. Fox. Latent fingerprint segmentation using ridge template correlation. 2011.

[97] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[98] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[99] X. Tao, X. Yang, K. Cao, R. Wang, P. Li, and J. Tian. Estimation of fingerprint orientation field by weighted 2d fourier expansion model. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1253–1256. IEEE, 2010.

[100] F. Turroni, D. Maltoni, R. Cappelli, and D. Maio. Improving fingerprint orientation extraction. *Information Forensics and Security, IEEE Transactions on*, 6(3):1002–1013, 2011.

[101] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*, 108(19):7733–7738, 2011.

[102] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PloS one*, 7(3):e32800, 2012.

[103] X. Wang and M. Liu. Fingerprint enhancement via sparse representation. In *Biometric Recognition*, pages 193–200. Springer, 2013.

[104] Y. Wang, L. G. Hassebrook, and D. L. Lau. Data acquisition and processing of 3-D fingerprints. *Information Forensics and Security, IEEE Transactions on*, 5(4):750–760, 2010.

[105] Y. Wang, J. Hu, and D. Phillips. A fingerprint orientation model based on 2d fourier expansion (fomfe) and its application to singular-point detection and fingerprint indexing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):573–585, 2007.

[106] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[107] Y. Weiss and W. T. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *Information Theory, IEEE Transactions on*, 47(2):736–744, 2001.

[108] C. Wilson, R. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto, and C. Watson. Fingerprint vendor technology evaluation 2003: Summary of results and analysis report, nistir 7123. *NIST, Gaithersburg, MD*, 2004.

[109] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.

[110] C. Wu, X.-C. Tai, et al. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. *SIAM J. Imaging Sciences*, 3(3):300–339, 2010.

[111] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems*, pages 341–349, 2012.

[112] J. Yang, L. Liu, T. Jiang, and Y. Fan. A modified gabor filter design method for fingerprint image enhancement. *Pattern Recognition Letters*, 24(12):1805–1817, 2003.

[113] X. Yang, J. Feng, and J. Zhou. Localized dictionaries based orientation field estimation for latent fingerprints. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(5):955–969, 2014.

[114] W. Yin, D. Goldfarb, and S. Osher. A comparison of three total variation based texture extraction models. *Journal of Visual Communication and Image Representation*, 18(3):240–252, 2007.

[115] W. Yin, D. Goldfarb, and S. Osher. The total variation regularized l^1 model for multiscale decomposition. *Multiscale Modeling & Simulation*, 6(1):190–211, 2007.

[116] S. Yoon, J. Feng, and A. K. Jain. On latent fingerprint enhancement. In *SPIE Defense, Security, and Sensing*, pages 766707–766707. International Society for Optics and Photonics, 2010.

[117] S. Yoon, J. Feng, and A. K. Jain. Latent fingerprint enhancement via robust orientation field estimation. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.

[118] J. Zhang, R. Lai, and C.-C. Kuo. Latent fingerprint segmentation with adaptive total variation model. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 189–195. IEEE, 2012.

[119] J. Zhang, R. Lai, and C.-C. Kuo. Adaptive directional total-variation model for latent fingerprint segmentation. *Information Forensics and Security, IEEE Transactions on*, 8(8):1261–1273, 2013.

[120] Q. Zhao and A. K. Jain. Model based separation of overlapping latent fingerprints. *Information Forensics and Security, IEEE Transactions on*, 7(3):904–918, 2012.

[121] Q. Zhao, L. Zhang, D. Zhang, W. Huang, and J. Bai. Curvature and singularity driven diffusion for oriented pattern enhancement with singular points. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2129–2135. IEEE, 2009.

[122] J. Zhou and J. Gu. A model-based method for the computation of fingerprints' orientation field. *Image Processing, IEEE Transactions on*, 13(6):821–835, 2004.

[123] J. Zhou and J. Gu. Modeling orientation fields of fingerprints with rational complex functions. *Pattern Recognition*, 37(2):389–391, 2004.

[124] E. Zhu, J. Yin, C. Hu, and G. Zhang. A systematic method for fingerprint ridge orientation estimation and image segmentation. *Pattern Recognition*, 39(8):1452–1472, 2006.