Latent Fingerprint Enhancement *via* Scattering Wavelets Network

B. Tech Project Report
Stage-1

Yash Bhalgat
13D070014

Guide:
Prof. Vikram M. Gadre

Department of Electrical Engineering,
Indian Institute of Technology, Bombay
Powai, Mumbai - 400 076.

2016-2017

Acknowledgments

First and foremost, I would like to express my gratitude to my guide, Prof. V.M Gadre, for his constant motivation and guidance during the project.

I would also like to thank Prof. Parmeshwar Birajadar with whom I'm working closely for my B. Tech project. For the first month, Parmeshwar sir helped me extensively in the literature survey of the past work done in latent fingerprint enhancement. I also want to thank him for bringing the idea of Scattering Wavelets Networks on the table to solve this problem and for constantly being there to help me throughout the project.

I would also like to thank Utkarsh Sharma for pitching in with the suggestions at crucial stages in this project.

Finally I would like to thank all my TI DSP Lab colleagues.

Background

What is a latent fingerprint?

Latent fingerprints are impressions lifted from the surfaces of objects typically at crime scenes.

(a) (b) (c)

Figure 1: Sample latent fingerprint images from NIST SD27 database

The word "latent" means invisible. These latent fingerprints are of poor quality due to the background noise and the unclear ridge structure. As latent fingerprints form an important evidence in the identification of criminals, various techniques are used to enhance these images before passing them to the identification or feature-extraction systems.

To encourage the research in automatic fingerprint enhancement and identification, NIST has made available excellent databases as benchmarks for rolled, matted and latent fingerprints called NIST SD4, SD14 and SD27 databases. In our application, we use the rolled fingerprint (better quality than latent) images from SD14 database for the training and the testing will be done on the SD27 latent fingerprint images.

Literature survey

The research pursued on this topic for the past few years can be classified into two major parts: Orientation based enhancement and Patch-based enhancement. Both these techniques have a common dictionary-based approach, in which a synthetic or a trained dictionary is used. Further, the fingerprint is enhanced by replacement of the image patches or the orientation-field patches with the dictionary atoms. The method used to choose the dictionary atom in each of these approaches is different as described in the table below:

Table 1: Existing latent fingerprint enhancement techniques

Research	Approach	Comments
Orientation-dictionary based approach		
Kai Cao, Anil Jain, et al.[5]	Trains a Convolutional Neural Network to classify each patch to a corresponding orient-dict atom. Training data created using angular similarity measure	Training requires huge computation time and resources (eg. GPUs). Trained model gives much better results than other techniques
Manhua Liu, et al.[12]	Initial orientation field is learnt by traditional techniques and dict is then used to refine estimate	Satisfactory results without any pre-training required. Dictionary learnt has redundant atoms which add to computation time.
X. Yang, et al.[13]	Proposed a Hough transform-based fingerprint pose estimation algorithm. Uses pose for replacement of noisy patch with dict element.	Incorporating pose estimation gives more accurate results, but hough transform calculation make it impractical for real-time.
C. Chen et al.[10]	Introduces of position-dependent dicts. Avoids impossible occurrence of orientation patches	Encounters the incorrect results around singularities in other techniques.
Patch-dictionary based approach		
Manhua Liu, et al.[3]	Dictionaries are constructed with a set of Gabor elementary functions. Multiscale patch-based sparse representation is iteratively applied to reconstruct high-quality fingerprint image	Iterative optimization of a non-convex function. Doesn't use the global ridge structure in the optimization.
Kai Cao, Anil Jain, et al.[4]	Uses L1 regularisation to remove piecewise smooth noise. Ridge quality of a patch, which is used for latent segmentation, is defined as the structural similarity between the patch and its atom	Dictionary is not synthetic and is created clustering the patches from SD14 database. Doesn't work for very low quality latent images
Proposed	**Need of a transform invariant to small changes but strong classifying power - Scattering Wavelets Networks. Using synthetic Gabor dict as in [3] for replacement of noisy patches.**	**Unlike CNNs, no training required. Use of non-linear modulus gives invariance to small deformations. Previous research has shown high accuracy for texture classification.**

Scattering Wavelet Networks v/s CNNs

CNNs (a brief overview)

In the recent years, Deep neural networks (DNNs) trained via backpropagation were shown to perform well on image classification tasks with lakhs of training images and lot of categories. The feature representation learnt by these networks achieves state-of-the-art performance not only on the classification task for which the network was trained, but also on various other visual recognition tasks, for example: classification on various dataset; scene recognition, etc. Due to this capability to generalize to new datasets, supervised CNN (Convolutional Neural Networks) training was an attractive approach for this generic application in Biometrics.

Each CNN has its own architecture - no. of layers and structure of each layer. They are traditionally made up of convolution layers, pooling layers (max pooling, weighted mean pooling, etc) and fully connected layers. Mathematically, it consists of a cascade of linear operators, intertwined with a non-linearity which computes a sequence of layers, $\phi_m(x)$,

$$\phi_m(x) = f(W_m f(...f(W_1(x))))$$

The W_i represent the filters in the network. These filters are learnt using gradient descent or any other optimization over the error function (on the target value):

$$W_{i,updated} = W_i + \eta \frac{\partial E}{\partial W_i}$$

In Latent fingerprint enhancement, the CNNs have provided promising results shown in [5]. The method first learns an orientation-based dictionary from the SD14 database using clustering. To create the training set, a similarity measure is used to assign a label to a set of 1 lac patches. After the training is done, an image of latent fingerprint is divided into overlapping patches and each patch is passed through the CNN to give a corresponding dictionary element for further filtering.

Limitations:
1. Typically requires 3-4 days for training the network and resources high processing power like GPUs.
2. Same fingerprint might give different impressions, but CNNs doesn't take pose or small deformations into account while training.

Scattering Wavelets Network

Need for deformation-invariant transform

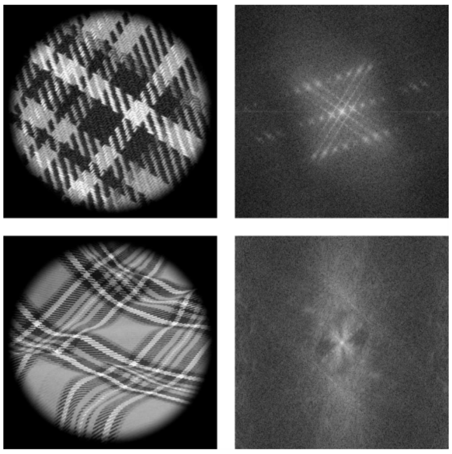
All the methods in the literature survey are based on replacement of a noisy patch by a perfect match from the created dictionary. And some of the methods strive to address the problem of invariance in the impressions obtained from the same person. So, clearly, whatever transform is used should be invariant to (small) deformations along with a strong classifying power.

A representation is stable to deformation when the difference induced on the representation by a deformation can be bounded as: [2]

$$\|\Phi L_{\tau}x - \Phi x\| \leqslant (C1\|\tau\|_{\infty} + C2\|\nabla\tau\|_{\infty})\|x\|$$

where $L_{\tau}x$ is the result of the deformation and Φx is the transform.

The modulus of the Fourier transform is global translation invariant but is unstable to deformations as is shown in the example below:



(a)

Figure 2: We can see that the modulus of the Fourier transform doesn't remain stable to deformations although being translationally invariant [2]

The other option is using wavelets, be cause stability to deformations can be obtained if the sine waves are localised. But wavelets also build only a slight invariance. Mallat [6] showed that cascading several layers of wavelet modulus operators provides much better invariance to deformations than the wavelet transform as well as retaining the information.

Complex Gabor and Mortlet wavelets

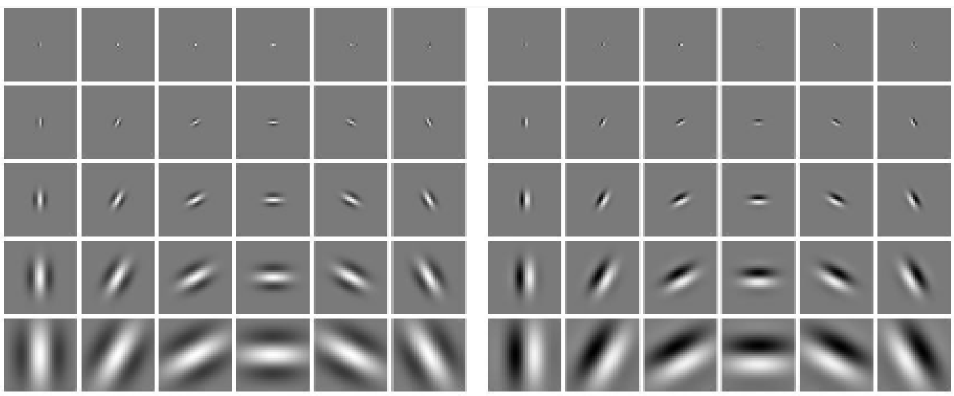
We need wavelets which are translational as well as rotational invariants (we will see in the next section that scattering wavelets are infact invariant to any small deformations). The family of wavelets generated as dilated and rotated versions of a mother wavelet ψ is denoted by $\{\psi_\lambda\}_\lambda$ where $\lambda = (\theta, j)$. The mostly commonly used example of such family of wavelets in Image Processing and filtering applications is the Gabor wavelets. In 2D, they are written as:

$$\psi_{gabor}(u,v) = \frac{1}{2\pi\sigma^2} exp\left(\frac{-\|(u,v)\|^2}{2\sigma^2} + ifx\right)$$

A Gabor wavelet is a gaussian envelope modulated by a sine wave of frequency f. One of the main disadvantages of Gabor wavelets is that they have a non-zero average which makes their responses non-sparse. So, for classification tasks, we use a variation of Gabor wavelets called the Mortlet wavelets. The basic modification is that a gaussian is subtracted from the gabor wavelet to make its average zero. We also want to increase the angular sensitivity of the wavelets, otherwise the rotations of a circular wavelet bring no advantage. So, the circular envelope of the gabor wavelet is replaced with an elliptical one. We get the expression for an elongated Mortlet wavelet as:

$$\psi_{elongatedmortlet}(u,v) = \frac{s}{2\pi\sigma^2} exp\left(\frac{-(u^2 + s^2v^2)}{2\sigma^2}\right) exp\left(ifx - K\right)$$

When increasing the number of orientations and scales per octave, elongated Morlet wavelets have better angular and scale sensitivity and therefore better separate the different frequency components of an image. This can positively impact classification.



(a)

Figure 3: Mortlet family for 5 scales and orientations (left half is the real part and right half is the imaginary part)

Wavelet Modulus operator and cascading wavelets

For every family of wavelets $\{\psi_l\}$, the wavelet coefficients $\{x \circledast \psi_\lambda\}_\lambda$ retain the variations of the signal at scales smaller than 2^J and to retain a coarse approximation of the signal, it is convolved with a lowpass filter ϕ_J.

It was shown in [9] that this averaging created stability to deformation upto a scale of 2^J. Mathematically saying, there's a constant C $\geqslant$ 0 s.t.:

$$\|L_\tau x * \phi_J - x{*}\phi_J\| \leqslant C\|x\|(\|2^J\|\tau\|_\infty + \|\nabla\tau\|_\infty)$$

But the problem here is, the averaging loses all the high frequency components and the wavelet coefficients which capture the high frequencies preserve invariance only upto 2^j (not upto the max scale). This is where the idea of Wavelet modulus operator is induced [2]. It is defined as:

$$|Wx| = \{x * \phi_J, |x * \psi_\lambda|\}_\lambda$$

Now, we can see that the first component (the average) is invariant till the scale of 2^J adn the second **non-linear** component is translation invariant to scale of 2^J and is stable to deformation. For further usage denote the non-linear part by $U_1x(u,\theta,j) = |x * \psi_{\theta,j}(u)|$
So, now, to build a honogeneous wavelet upto the scale 2^J, U_1 is also averaged to get the next level of scaterring coefficients:

$$S_1x(u,\theta,j) = U_1x(u,\theta,j) * \phi_J = |x * \psi_{\theta,j}| * \phi_J$$

and the non-linear part at level-2 is given by:

$$U_2x(u,\theta_1,j_1,\theta_2,j_2) = ||x * \psi_{\theta_1,j_1}| * \psi_{\theta_2,j_2}|$$

At the m^{th} level, the coefficients are given by:

$$S_mx(u,\theta,j) = U_mx(u,\theta,j) * \phi_J$$

$$U_{m+1}x(u,\theta_1,j_1,\theta_2,j_2) = |U_mx(u,...) * \psi_{\theta_{m+1},j_{m+1}}|$$

In practice and in my project, Mortlet wavelets are used, because they decay very fast and hence scattering coefficients only upto order 2 are sufficient. Since the wavelet operator preserves the norm of the input signal, the scattering transform also preserves the norm. (The second part is proved in [9] and I have skipped the proof since it is quite mathematically intensive.)

The feature vectors for the classification (explained later) are created by concatenating the 2 layers of coefficients as a vector:

$$featurevector \leftarrow \{S_0, S_1, S_2\}$$

The figure(5) below shows the scattering coefficients for order 0, 1, 2 for a given texture image.

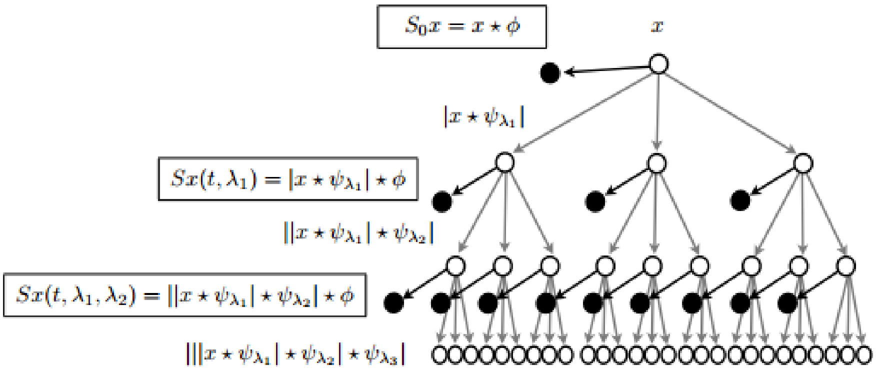
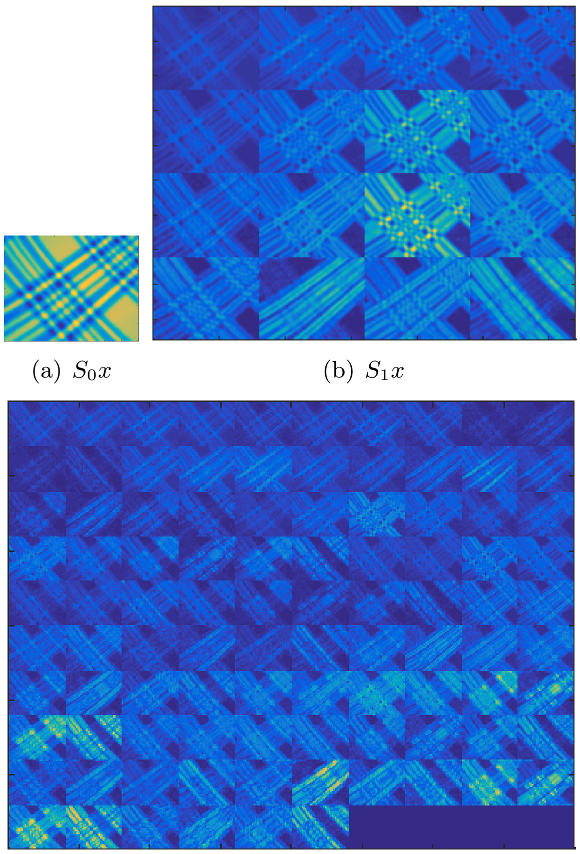


Figure 4: 3 cascaded layers of scattering wavelet transform [1]



(a) S_0x (b) S_1x

(c) S_2x

Figure 5: Scattering coefficients of order m = 0, 1, 2 computed with 4 scales and 4 orientations

Cartoon Texture Decomposition

A technique common to all the methods in the literature survey and the proposed method is the Total Variation model based Cartoon Texture Decomposition. In our application, the friction ridge structure of the fingerprint represents the texture component and the large-scale background noise forms the cartoon component. So a non-linear total variation model is utilized to remove the cartoon component.

Local Non-linear Total Variation model

The aim is to decompose an image f into two components $f = u + v$, where u is the cartoon component and v is the texture component. Adopting the Meyer's model, the general framework is:

$$\underset{u,v}{inf}\{F_1(u) + \lambda F_2(v) | f = u + v\}$$

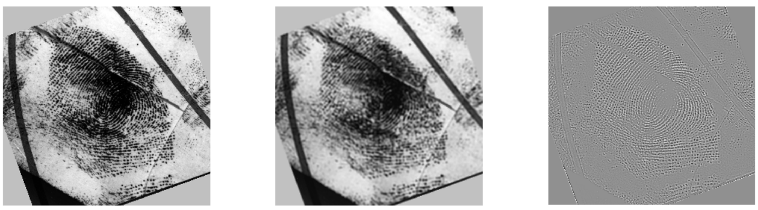
where $F_1(u) \gg F_2(u)$ and $F_1(v) \ll F_2(v)$, i.e. the cartoon component u is penalised by F_1 and texture by F_2. The cartoon region is characterised as its total variation does not decrease by low-pass filtering and the total variation of the texture component is high. But this total variation decreases very fast under low-pass filtering. So, mathematically, the local total variation is defined as [8]:

$$LTV_\sigma f = L_\sigma * |Df|$$

where L_σ is the low pass filter and Df is as defined in [7].
The relative reduction rate of LTV is defined as [8]:

$$\lambda_\sigma \triangleq \frac{LTV_\sigma(f) - LTV_\sigma(L * f)}{LTV_\sigma(f)}$$

This parameter gives us the local oscillatory behaviour of f. So, $\lambda_\sigma = 0$ means that f has only low pass content, hence the point can be classified as a cartoon component. And λ_σ close to 1 means that the low pass content is zero, hence a texture component.



(a)

Figure 6: Original image, cartoon component and texture component

Proposed Work

Similar to the method in [3], we use a synthetic dictionary made up of Gabor basis functions. The 2D Gabor functions have the general form:

$$h(x, y, \theta, f) = exp\left\{-\frac{1}{2}\left[\frac{x_\theta^2}{\delta_x^2} + \frac{y_\theta^2}{\delta_y^2}\right]\right\} cos(2\pi f x_\theta + \varphi_0)$$

where $x_\theta = xcos\theta + ysin\theta$ and $y_\theta = -xsin\theta + ycos\theta$

f is the frequency which should correspond to the ridge frequency. So, f is varied from 7 to 19 in steps of 2. θ varies from 0 to $5\pi/16$ at a step of $\pi/16$ to have 16 values. ϕ varies from 0 to $5\pi/6$ at a step of $\pi/6$ to have 6 values. Finally, we have a dictionary of $7 * 16 * 6 =$ 672 gabor atoms.

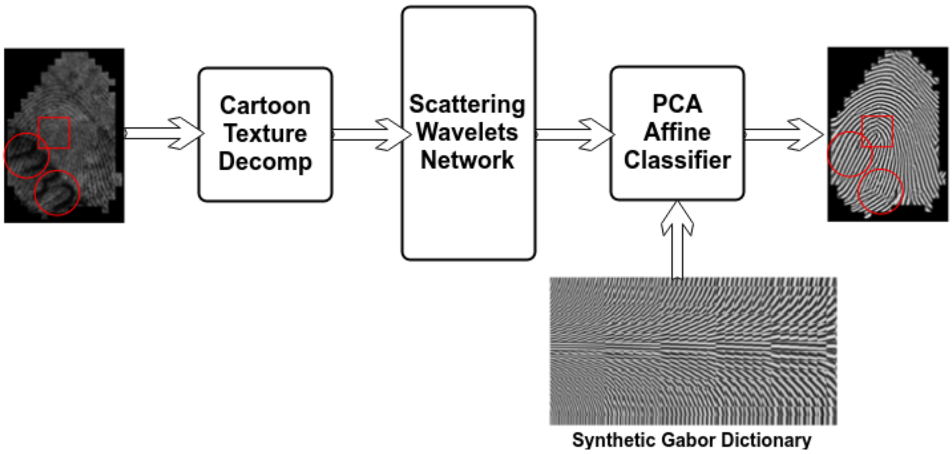
The dictionary atoms serve the purpose as replacement patches. An impure latent input image is divided in overlapping patches (of size equal to the atom size) and each patch is replaced by a corresponding atom. This representative atom is chosen using an affine PCA classifier, which needs to be trained. So, to create the training data for each atom, we will use the structural similarity (SSIM) index. Given, two images x and y, the similarity between them is given by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

Using patches extracted from the SD14 database images, we create the training data for this dictionary. The scattering wavelet network is used for the extraction of a feature vector for each patch. These feature vectors are used to train the classifier which has 672 classes/labels.

Once the classifier is trained, an input latent fingerprint image is divided into overlapping patches. Each patch is passed through the scattering network and the feature vector obtained is passed to the affine PCA classifier. Affine PCA classifier reduces the dimension of the feature vector by projecting it to an eigen-space and this reduced vector is used to classify the patch to one of the dictionary atoms. In this way, each patch is then replaced by the corresponding atom and the final enhanced fingerprint image is formed by filtering with a Gabor filterbank.

The overall block-diagram of the system described above is shown below:



(a)

Figure 7: **A block-diagram of the proposed algorithm.**

Bibliography

[1] Bruna, J. (2013). Scattering representations for recognition (*Doctoral dissertation, Ecole Polytechnique X*).

[2] Sifre, L., & Mallat, S. (2014). Rigid-motion scattering for texture classification. *arXiv preprint arXiv:1403.1687.*

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

[4] Cao, K., Liu, E., & Jain, A. K. (2014). Segmentation and enhancement of latent fingerprints: A coarse to fine ridgestructure dictionary. *IEEE transactions on pattern analysis and machine intelligence*, 36(9), 1847-1859.

[5] Cao, K., & Jain, A. K. (2015, May). Latent orientation field estimation via convolutional neural network. *In 2015 International Conference on Biometrics (ICB)* (pp. 349-356). IEEE.

[6] Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1872-1886.

[7] Le Guen, V. (2014). Cartoon+ texture image decomposition by the TV-L1 model. *Image Processing On Line*, 4, 204-219.

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

[9] Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10), 1331-1398.

[10] Chen, C., Feng, J., & Zhou, J. (2016, June). Multi-scale dictionaries based fingerprint orientation field estimation. *In Biometrics (ICB), 2016 International Conference on (pp. 1-8). IEEE.*

[11] Feng, J., Zhou, J., & Jain, A. K. (2013). Orientation field estimation for latent fingerprint enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 35(4), 925-940.

[12] Liu, M., & Liu, S. (2015, December). Latent fingerprint orientation estimation via sparse representation. *In 2015 10th International Conference on Information, Communications and Signal Processing (ICICS)* (pp. 1-4). IEEE.

[13] Yang, X., Feng, J., & Zhou, J. (2014). Localized dictionaries based orientation field estimation for latent fingerprints. *IEEE transactions on pattern analysis and machine intelligence*, 36(5), 955-969.

[14] Sifre, L., & Mallat, S. (2013). Rotation, scaling and deformation invariant scattering for texture discrimination. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1233-1240).