Latent Fingerprint Enhancement via Scattering Wavelets Network

Yash Bhalgat
B.Tech Project
Stage-1

Guide: Prof. V.M. Gadre
Department of Electrical Engineering
IIT Bombay

21 November, 2016

Outline

What is a latent fingerprint?

- Literally ‘latent’ means invisible
- Impressions lifted from the surfaces of objects typically at crime scenes
- Latent fingerprints form an important evidence in the identification of criminals
- Uniqueness of fingerprint is determined by the flow of ridges and feature points are formed by the different patterns of the ridges

What is a latent fingerprint?

(a) (b) (c)

Figure: Sample latent fingerprint images from NIST SD27 database

- Latent fingerprints are of poor quality due to the background noise and the unclear ridge structure
- Various techniques are used to enhance these images before passing them to the identification or feature-extraction systems

Aim and Motivation

- Aim of this project is latent fingerprint enhancement by exploiting the properties of cascaded Scattering Wavelets such as invariance to small deformations and high preformance in texture classification.
- To encourage research in this area, NIST has made available the **NIST SD4, SD14 and SD27** databases as benchmarks for rolled, matted and latent fingerprints respectively.

Existing Techniques

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

Orientation-based 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.

Patch-based Techniques

Research	Approach	Comments
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 Technique

- 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 (similar to Liu's work)
- Unlike CNNs, no training required for the scattering network. Minimal training for the classifier.
- Use of non-linear modulus gives invariance to small deformations. Previous research on cascaded scattering wavelets has shown high accuracy for texture classification.

CNNs (a brief overview)

- In recent years, Deep neural networks (DNNs) trained via back-propagation were shown to perform well on image classification.
- Feature representation learnt by these networks works not only on the classification task for which the network was trained, but also on various other visual recognition tasks like scene recognition.
- 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.**
  <explain Cao, Jain's paper>

CNNs (a brief overview)

- CNNs are traditionally made up of convolution layers, pooling layers (max pooling, weighted mean pooling, etc) and fully connected layers.
- Mathematically, a cascade of linear operators, intertwined with a non-linearity which computes a sequence of layers, $\phi_m(x)$,

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

- 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):

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

CNNs (a brief overview)

CNNs in latent fingerprint enhancement:

- The CNNs have provided promising results shown in this area [5].
- The method [5] 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.
- An image of latent fingerprint is divided into overlapping patches and each patch is passed through the **trained** CNN to give a corresponding dictionary element for further filtering.

CNNs (a brief overview)

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.

Need for Deformation-invariant transform

- Whatever transform is used should address the problem of difference in impressions obtained from the same person.
- **Need for small deformation-invariant transform with strong classifying power**.
- A representation is stable to deformation when the difference induced on the representation by a deformation can be bounded as: [2]

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

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

Fourier Transform: Deformation-invariant?

- Translation in x only adds a phase component to $Fx(\omega)$, so magnitude of Fourier transform is translation-invariant. ($C1 = 0$ and $C2 = const$)
- But it is unstable to non-translational deformations as is shown in the example below:

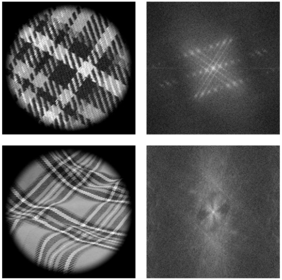


Figure: small deformation in texture and its fourier transform[2]

Building invariance with Wavelets

- Stability to deformations can be obtained if the sine waves are localised. Hence, wavelets is an option!
- But wavelets also build only a slight invariance.
- Consider a 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 wavelet coefficients for the wavelet ψ_j preserve invariance only upto 2^j [9].
- Mallat [6] showed that cascading several layers of wavelet modulus operators provides much better invariance to deformations as well as retaining the information
  <will come back to this after next section>

Building invariance with Wavelets

- Mallat [6] showed that cascading several layers of wavelet modulus operators provides much better invariance to deformations as well as retaining the information
  <will come back to this after next section>

Complex Gabor wavelets

- Need wavelets which are translational as well as rotational invariants
- Family of wavelets generated as dilated and rotated versions of a mother wavelet ψ is denoted by $\{\psi_\lambda\}_\lambda$ where $\lambda = (\theta, j)$
- Most commonly used example of such family of wavelets in Image Processing and filtering applications is the Gabor wavelets
- A Gabor wavelet is a gaussian envelope modulated by a sine wave of frequency f
- In 2D, they are written as:

$$\psi_{gabor}(u, v) = \frac{1}{2\pi\sigma^2} exp\left(\frac{-\|(u, v)\|_2^2}{2\sigma^2} + ifx\right)$$

Complex Gabor wavelets

- The main disadvantage of Gabor wavelets is that they have a non-zero average which makes their responses non-sparse
- Sparsity is important for classification problems
- So, use a variation of Gabor wavelets called the Mortlet wavelets

Complex Mortlet wavelets

- Subtract a gaussian 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
- The expression for an elongated Mortlet wavelet is obtained as:

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

Complex Mortlet wavelets

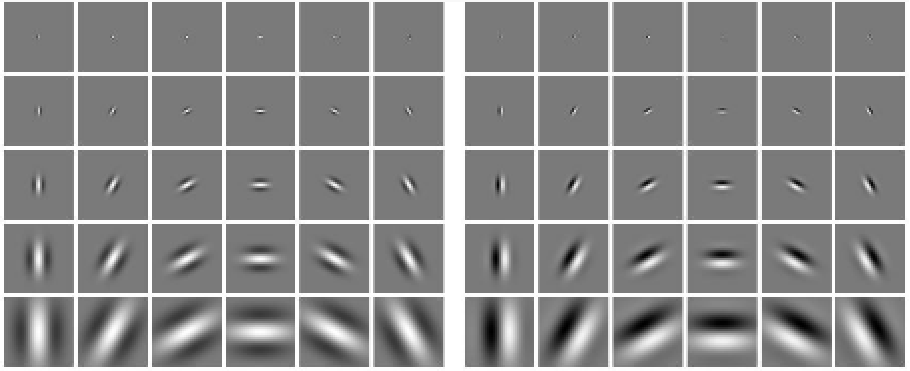


Figure: Mortlet family for 5 scales and 6 orientations (left half is the real part and right half is the imaginary part)

Cascading wavelets and Wavelet Modulus operator

- **Wavelet transform:** For every family of wavelets $\{\psi_l\}$, the wavelet coefficients $\{x * \psi_\lambda\}_\lambda$ 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
- ϕ_J is **a standard gaussian dilated by a scale of** 2^J
- Averaging by ϕ_J creates stability to deformation upto a scale of 2^J. (result proved in [9]) i.e., $\exists C \geq 0$ s.t.:

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

Cascading wavelets and Wavelet Modulus operator

- Problem with single layer is that ϕ_J loses all high frequency components and the high-pass wavelets ψ_j are translational invariant only upto 2^j
- This is where we need the Wavelet modulus operator

Cascading wavelets and Wavelet Modulus operator

- Wavelet modulus operator is defined as:

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

- To enforce invariance to 2^J translations while maintaining stability to deformations of wavelet coefficients $x * \psi_j$, the most obvious thing to do is to average them into $x * \psi_j * \phi_J$.
- But this yields almost zero coefficients since ψ_j and ϕ_J are designed to be orthogonal to each other.
- Therefore, to build non-trivial translation invariant, a possible strategy is to intertwine a modulus non-linearity between the wavelet convolution $*\psi_j$ and the averaging $*\phi_J$.

Cascading wavelets and Wavelet Modulus operator

- Call the non-linear part as: $U_1 x(u, \theta, j) = |x * \psi_{\theta,j}(u)|$
- Now, to build a honogeneous wavelet upto the scale 2^J, U_1 is also averaged to get the next level of scaterring coefficients:

$$S_1 x(u, \theta, j) = U_1 x(u, \theta, j) * \phi_J = |x * \psi_{\theta,j}| * \phi_J$$

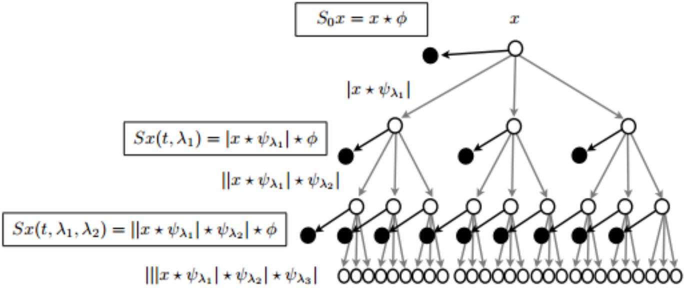
- And the non-linear part at level-2 is given by:

$$U_2 x(u, \theta_1, j_1, \theta_2, j_2) = ||x * \psi_{\theta_1,j_1}| * \psi_{\theta_2,j_2}|$$

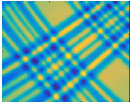
Cascading wavelets and Wavelet Modulus operator

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

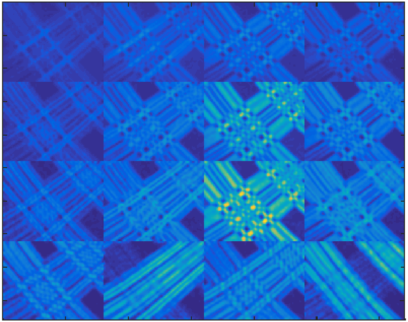
$$S_m x(u, \theta, j) = U_m x(u, \theta, j) * \phi_J$$
$$U_{m+1} x(u, \ldots) = \left| U_m x(u, \ldots) * \psi_{\theta_{m+1}, j_{m+1}} \right|$$



Cascading wavelets and Wavelet Modulus operator



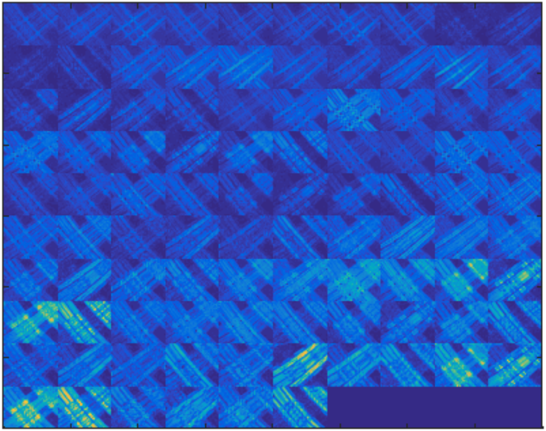
(a) S_0x



(b) S_1x

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

Cascading wavelets and Wavelet Modulus operator



(a) $S_2 x$

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

Cartoon Texture Decomposition

- The presentation is almost over!
- A technique common to all the methods in the literature survey and the proposed method is the T**otal 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:

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

- $F_1(u) \gg F_2(u)$ and $F_1(v) \ll F_2(v)$, i.e. the cartoon component u is penalised by F_1 and the texture v penalised by F_2
- So, large values of λ will give more importance to F_2 and hence give the texture component and *vice versa*

Local **Non-linear** Total Variation model

- The relative reduction rate of LTV is defined as [8]:

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

  where L is the low pass filter
- This is because, for texture component has high local variation which will decrease largely if passes through a lowpass filter. Whereas, LTV of cartoon component doesn't change under lowpass filtering
- $\lambda_\sigma = 0$ 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

Local **Non-linear** Total Variation model

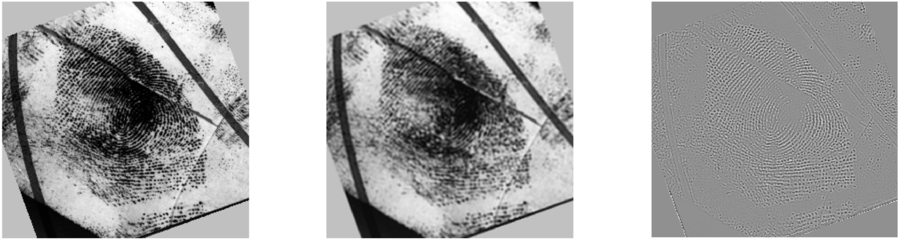


Figure: Original image, cartoon component and texture component

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 $y_\theta = -xsin\theta + ycos\theta$

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

Proposed Work - Training set creation

- 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:

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

Proposed Work - Training

- Using patches extracted from the SD14 database images, we create the training data for this 672-class PCA-affine 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

Proposed Work - Block diagram

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

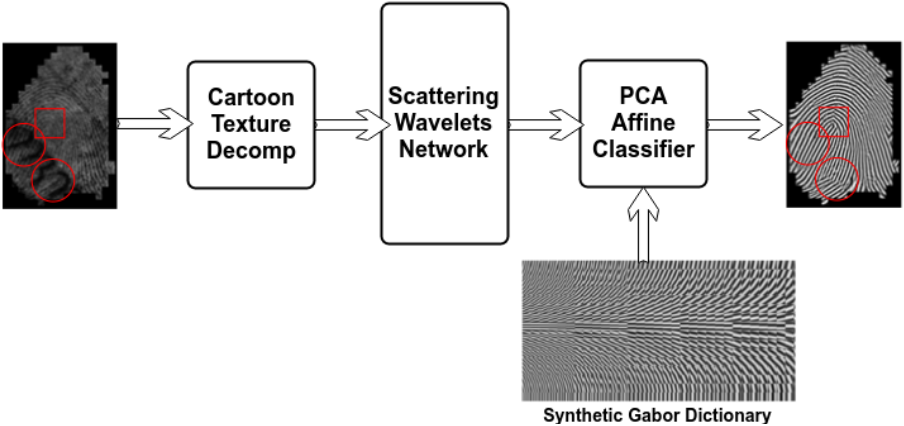


Figure: **A block-diagram of the proposed algorithm.**

References I

- Bruna, J. (2013). Scattering representations for recognition (*Doctoral dissertation, Ecole Polytechnique X*).
- Sifre, L., & Mallat, S. (2014). Rigid-motion scattering for texture classification. *arXiv preprint arXiv:1403.1687*.
- 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.
- 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.

References II

- 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.
- Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1872-1886.
- Le Guen, V. (2014). Cartoon+ texture image decomposition by the TV-L1 model. *Image Processing On Line*, 4, 204-219.
- 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.
- Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10), 1331-1398.

References III

- 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*.
- 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.
- 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.
- 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.

References IV

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).