CHAPTER 6

SHOEPRINT RECOGNITION BASED ON WAVELET TRANSFORMS AND SUPPORT VECTOR MACHINE

6.1 INTRODUCTION

Shoeprint recognition is an important research field of pattern recognition. It’s a typical multi-class pattern recognition problem. Wavelet transform has been developed as powerful analyzing tool for signal processing and successfully applied to pattern recognition. Feature extraction is an important part of pattern recognition, whose main purpose is extracting essential features and dimensionality reduction. In this chapter, three different Wavelet Transforms are used to extract the features of shoeprint images and use SVM as classifier. Further, a general modification for wavelet transform based feature extraction is proposed to find the optimal feature set, its performance then compared with the state of art wavelet transforms and other existing feature extraction approaches. The one-against-one (1A1) classification strategy for multi-class pattern recognition with SVM is employed for classification. The following section presents the feature extraction approaches based on Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT) and Redundant DWT (RDWT), followed by the description of the proposed wavelet transform based optimal feature selection. And the next section discusses the experimental results with SVM classifier.
6.2 FEATURE EXTRACTION USING WAVELET TRANSFORMS

Wavelet based texture analysis inherently regards the problem of describing textures at finer and coarser scales due to multiresolution analysis. It is capable of fetching texture similarities at different scales. Several wavelet based approaches exist in the literature for texture studies. Not only multiscale analysis but also directional selectivity of wavelet decomposition is a favored property of such approaches.

6.2.1 The Wavelet Transform

The driving force behind Wavelet Transforms (WTs) is to overcome the disadvantages of the Time Fourier Transform (STFT), which provides constant resolution for all frequencies since it uses the same window for the analysis of the inspected signal \( x(t) \). On the contrary, WT uses multi-resolution, that is, they used different windows to analyze different frequency bands of the signal \( x(t) \). Different window functions \( \psi(s, b, t) \), which are also called on wavelets, can be generated by dilation or compression of a mother wavelet \( \psi(t) \), at different time frames. Morlet wavelet is an example for mother wavelet as shown in Figure 6.1. Ascale is the inverse of its corresponding frequency. WT can be categorized as discrete WT or continuous WT. A type of Continuous Wavelet Transform (CWT) is applied to the signal \( x(t) \) can be defined as,

\[
w(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left( \frac{t - b}{a} \right) dt
\]  

(6.1)

where

i. \( a \) is the dilation factor,
ii. \( b \) is the translation factor and

iii. \( \psi(t) \) is themother wavelet.

iv. \( 1/|a| \) is a normalization term that makes waveletsofdifferent scales have the same amount of energy.

### 6.2.1.1 Wavelet Theory

A wavelet is a wave form of effectively limited duration that has a nonaveraged value of zero. Instead of shifts and modulates of a prototype function, one can choose shifts and scales, and obtain a constant relative bandwidth analysis known as the wavelet transform. To achieve this, take a real-band-pass filter with impulse response \( \Psi(t) \) and zero mean.

\[
\int_{-\infty}^{\infty} \psi(t) \, dt = \psi(0) = 0 \tag{6.2}
\]

Then, the continuous wavelet transformation as defined in equation (6.1). That is, we measure the similarity between the signal \( f(t) \) and scales of an elementary function, since

\[
CWT_f(a,b) = \left\langle \psi_{a,b}(t), f(t) \right\rangle, \quad \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \tag{6.3}
\]

And the factor \( 1/\sqrt{a} \) is used to conserve the norm. Now, the functions used in the expansion have changing time–frequency tiles because of the scaling. For small \( a < 1 \), \( \Psi_{a,b}(t) \) will be short and of high frequency, while for large \( a > 1 \), \( \Psi_{a,b}(t) \) will belong and of low frequency. Thus, a natural discretization will use target time steps for large \( a \), and conversely, choose finer time steps for small \( a \).
Figure 6.1  Morlet wavelet of arbitrary width and amplitude, with time along the x-axis

6.2.1.2 Wavelet Properties

The domain is compact support, which ensures that the function is fast decaying, and sometimes localization can be obtained. And the admissibility condition

\[ C_\psi = \int \left( \frac{1}{\omega} \right) |\psi(\omega)|^2 d\omega < \infty \quad (6.4) \]

where \( \psi(\omega) = \int \psi(t) e^{-i\omega t} dt \)

This condition means that the waveform of the mother wavelet function must be oscillating; the average value of the wavelet in the time domain must be zero.

6.2.1.3 Scaling of Wavelet

Scaling a wavelet simply means stretching (or compressing) it. Low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signal), whereas high frequencies (low scales) correspond to detailed information of hidden patterns in the signal (that usually last for relatively short time). In terms of mathematical functions, if \( f(t) \) is a given function, \( f(st) \) corresponds to a contracted (compressed) version of \( f(t) \).
if \( s > 1 \) and to an expanded (dilated) version of \( f(t) \) if \( s < 1 \). Figure 6.2 shows scaling of wavelet at four different scales.

![Scaling of Wavelet](image)

**Figure 6.2 Scaling of Wavelet**

### 6.2.2 Computation of the CWT

Interpretation of the equation (6.1) is explained in this section. Let \( x(t) \) be an analyzed mother wavelet. The mother wavelet is chosen to serve as a prototype for all windows in the process. All the windows that are used are the dilated (or compressed) and shifted versions of the mother wavelet. There are a number of functions that are used for this purpose. Once the mother wavelet is chosen, the computation starts with \( s = 1 \) (scale) and the continuous wavelet transform is computed for all values of \( s \), smaller and larger than “1”. The wavelet is placed at the beginning of the signal at the point which corresponds to time = 0. The wavelet function is scaled “1” is multiplied by the signal and then integrated.
overall times. The result of the integration is then multiplied by the constant number $1 / \sqrt{s}$. This multiplication is for energy normalization purposes so that the transformed signal will have the same energy at every scale. The final result is the value of the transformation, i.e., the value of the continuous wavelet transform at time zero and scales $= 1$. In other words, it is the value that corresponds to the point $t = 0, s = 1$ in the time-scale plane. The wavelet at scales $= 1$ is then shifted towards the right by “b” amount to the location $t = b$, and the above equation is computed to get the transform value at $t = b, s = 1$ in the time-frequency plane. The procedure is repeated until the wavelet reaches the end of the signal. One row of points on the time-scale plane for the scales $= 1$ is now completed. Then, “s” is increased by a small value. Since, it is a continuous transform, and therefore, both time (b) and “s” are incremented continuously. However, if this transform needs to be computed, then both parameters are increased by a sufficiently small stepsize. This corresponds to sampling the time-scale plane. The above procedure is repeated for every value of “s”. Every computation for a given value of “s” fills the corresponding single row of the time-scale plane. When the process is completed for all desired values of “s”, the CWT of the signal has been calculated. The Figure 6.3 illustrates the entire process step by step.
Figure 6.3 Computation of the CWT (a) scales = 1 (b) scale s = 5
6.2.2.1 Scalogram

For all \( \psi(t), x(t) \in L^2(\mathbb{R}) \), the inverse wavelet transform of \( x(t) \) is defined as

\[
X(t) = \frac{1}{C_\psi} \int a^{-2} W_x(a, b; \psi) \psi_a^*(t) \, da \, db
\]  

(6.5)

Since the wavelet transform does not lose any information, and the energy is preservative for the transform. So the following equation is tenable:

\[
\langle X(t), X(t) \rangle = \int |X(t)|^2 \, dt = \frac{1}{C_\psi} \int a^{-2} \int |W_x(a, b; \psi)|^2 \, da \, db
\]  

(6.6)

\[ |w_x(a, b; \psi)|^2 \Delta a \Delta b / C_\psi a^2 \]

represents the total energy of a domain centered at \( (a, b) \) with scale interval \( \Delta a \) and time interval \( \Delta b \) and \( |w_x(a, b; \psi)|^2 \) is defined as the wavelet scalogram. It shows how the energy of the signal varies with time and frequency. The wavelet scalogram has been widely used for the analysis of non-stationary signal.

In order to illustrate the characteristics of the scalogram, a signal \( X(t) \) is considered. Figure 6.4 shows the signal \( X(t) \) in the time domain and it can be expressed by

\[
X = 6 \sin 2\pi (10t) + 6 \sin 2\pi (20t) \quad 0 \leq t < 0.5
\]

(6.7)

\[
X = 6 \sin 2\pi (20t) + 6 \sin 2\pi (30t) \quad 0.5 \leq t < 1
\]
Figure 6.4 TestsignalX(t)

Figure 6.5 FFT of numerical test signal X(t)

Figure 6.6 Scalogram for signal X(t)
Although we know that the signal X(t) contains three components whose frequencies are 10, 20 and 30 Hz, respectively, from the spectrum (Figure 6.5) we cannot determine whether the component exists during all the signal lifespan or only part of the lifespan. However, through the signal scalogram, using Morlet Wavelet, we are not only able to recognize the three components but also able to know the time at which they exist. It is observed that for Scalograms, the wavelet scale 'a' is inversely proportional to the frequency, this means that lower scale values show high frequency and vice versa. In Figure 6.6, the vertical-axis is represented by scale factor 'a' which is inversely proportional to the frequency and horizontal-axis is time. Bright region in scalogram shows the presence of that particular frequency at a particular time.

1. \( a = 8 \) represents frequency = 10 Hz which is present in the timespan 0 < t < 0.5.

2. \( a = 4 \) represents frequency = 20 Hz which is present in the timespan 0 < t < 1.

3. \( a = 3 \) represents frequency = 30 Hz which is present in the timespan 0.5 < t < 1.

Such representation is very helpful to reveal the time intervals as well as the frequency range of the occurrence of variations in the signal or image.

6.2.3 The DWT Transform

Wavelets are special functions which, in a form analogous to sines and cosines in Fourier analysis, are used as basal functions for representing signal. For 2-D images, applying DWT corresponds to processing...
the image by 2–D filters in each dimension (Figure 6.7). The filters divide the input image into four non-overlapping multi-resolution sub-bands $LL_1, LH_1, HL_1$, and $HH_1$. The sub-band $LL_1$ represents the coarse-scale DWT coefficients while the sub-bands $LH_1, HL_1$, and $HH_1$ represent the fine-scale of DWT coefficients. To obtain the next coarser scale of wavelet coefficients, the sub-band $LL_1$ is further processed until some final scale $N$ is reached. When $N$ is reached we will have $3N + 1$ sub-bands consisting of the multi-resolution sub-bands $LL_N$ and $LH_X, HL_X$ and $HH_X$ where $X$ ranges from 1 until $N$.

Due to its excellent spatio-frequency localization properties, the DWT is very suitable to identify the areas in the image for feature extraction. In particular, this property allows the exploitation of the masking effect of the human visual system such that if a DWT coefficient is modified, only the region corresponding to that coefficient will be modified. In general, most of the image energy is concentrated at the lower frequency sub-bands $LL_X$. Embedding in the low frequency sub-bands, however, could increase robustness significantly. On the other hand, the high frequency sub-bands $HH_X$ include the edges and textures of the image and the human eye is not generally sensitive to changes in such sub-bands. Figure 6.8 shows the 1D and 2D DWT decomposition of shoeprint images.
Figure 6.7 Discrete Wavelet Transform
(a) 1DDWT analysis and synthesis filter (b) 2D wavelet transformation of an image
Figure 6.8  DWT Composition of Shoeprint Image (a) 1D-DWT and (b) 2D-DWT

6.2.4  Redundant Discrete Wavelet Transform (RDWT)

Unlike the DWT, the redundant discrete wavelet transform (RDWT) is an overcomplete representation for any input sequence and it gives a better approximation to the continuous wavelet transform. The RDWT is shift invariant and its redundancy introduces an overcomplete frame expansion. It has been proved that frame expansions add numerical robustness in case of adding white noise (Goyal et al., 1998 and Daubechies 1990) such in the case of quantization. This property makes RDWT-based signal processing more robust than DWT. It is also observed that RDWT is very successful in noise reduction and feature detection.

The RDWT removes the decimation operators form DWT filter banks and to retain the multiresolution characteristic, the wavelet filters are adjusted accordingly at each scale. Specifically,

\[ h_{j_1}[k] = h[k] \]

where \( J_1 \) is the start scale, \( h_{j_1}[k] \) is the RDWT scaling filter at scale \( J_1 \). \( h[k] \) is a normal DWT scaling filter.

Filters at later scales are upsampled versions from the filter coefficients at the upper stage,

\[ h_j[k] = h_{j+1}[k] \uparrow 2 \]

and similar definitions is applied to \( g_j[k] \), the wavelet filter of the orthonormal DWT.
The RDWT multiresolution analysis is implemented based on the filter bank equations:

Analysis: \[(6.8)\]

\[(6.9)\]

Synthesis: \[(6.10)\]

The lack of down sampling in the RDWT analysis yields a redundant representation of the input sequence; that is, two valid descriptions of the coefficients exist after one stage of RDWT analysis. A two dimensional RDWT is shown in Figure 6.9 and its spatial representation of a two scale is shown in Figure 6.10. The 1-level RDWT representation for the input shoeprint image is shown in Figure 6.11.

![Diagram](image_url)

**Figure 6.9** A two dimensional decomposition using RDWT
Figure 6.10 Spatially coherent representations of a two scale two dimensional RDWT.

Figure 6.11 1-Level RDWT representation of input image
6.2.4.1 Frames

A family of functions \((\psi_i)_{i \in I}\) in a Hilbert space \(\mathcal{H}\) is called a frame if there exist \(A > 0\), and \(B < \infty\) so that, for all \(f \in \mathcal{H}\),

\[
A \| f \|^2 \leq \sum_{i \in I} | < \psi_i, f > |^2 \leq B \| f \|^2 (6.11)
\]

A and B are called the frame bounds. The dual frame \(\hat{\psi}_i\) of \((\psi_i)\) is an expansion set in Hilbert space \(\mathcal{H}\) and for all \(f \in \mathcal{H}\),

\[
\frac{1}{B} \| f \|^2 \leq \sum_i | < \hat{\psi}_i, f > |^2 \leq \frac{1}{A} \| f \|^2 (6.12)
\]

Any function \(f \in \mathcal{H}\) can be expanded as

\[
f = \sum_{i} \alpha_i \psi_i = \sum_{i} < \hat{\psi}_i, f > \hat{\psi}_i (6.13)
\]

\[
= \sum_{i} \beta_i \psi_i = \sum_{i} < \hat{\psi}_i, f > \psi_i (6.14)
\]

If two frame bounds are equal, \(A = B\), the frame is called a tight frame, In a tight frame, for all \(f \in \mathcal{H}\),

\[
\sum_{i \in I} | < \psi_i, f > |^2 = A \| f \|^2 \quad (6.15)
\]

\[
\hat{\psi}_i = \frac{1}{A} \psi_i \quad (6.16)
\]

\[
f = \frac{1}{A} \sum_{i} (\psi_i, f) \psi_i \quad (6.17)
\]

In this case, \(A > 1\), and \(A\) gives the “redundancy ratio”, a measure of the degree of over completeness of the expansion.
6.2.4.2 RDWT and frame expansion

RDWT is a frame expansion with frame bounds $A=2$ and $B=2^J$, where $J$ is the number of levels in the transform. Thus, for one level, the RDWT is a tight frame. As shown in Figure 6.12, for the lowpass coefficients $c_j$, it is composed with two parts $c_j'$ and $c_j''$, each of them is a valid DWT lowpass description of $c_{j+1}$. It is a similar case for the highpass coefficients $d_j$. Thus, Parseval’s theory holds for each of the descriptions:

$$
\sum \|c_j'\|^2 + \sum \|d_j'\|^2 = \sum \|c_{j+1}\|^2
$$

$$
\sum \|c_j''\|^2 + \sum \|d_j''\|^2 = \sum \|c_{j+1}\|^2
$$

---

One – Level RDWT Analysis

$$
\begin{array}{c|c|c|c|c}
  & c_j'[0] & c_j''[0] & c_j'[1] & c_j''[1] \\
c_{j+1}[0] & c_{j+1}[1] & c_{j+1}[2] & c_{j+1}[3] \\
\end{array}
$$

$$
\begin{array}{c|c|c|c|c}
  & d_j'[0] & d_j''[0] & d_j'[1] & d_j''[1] \\
d_{j+1}[0] & d_{j+1}[1] & d_{j+1}[2] & d_{j+1}[3] \\
\end{array}
$$

Figure 6.12 One scale of RDWT decomposition
Thus, for the RDWT coefficients all together, we have

\[ \sum ||c_j||^2 + \sum ||d_j||^2 = 2 \sum ||c_{j+1}||^2 \] (6.18)

Therefore, one-level RDWT decomposition is a tight frame with A=2.

For decomposition level J>1, that is, if the decomposition starts at scale J_1, then

\[ \sum ||c_{j_1}||^2 = \frac{1}{2} \left( \sum ||c_{j_1-1}||^2 + \sum ||d_{j_1-1}||^2 \right) \]

\[ = \frac{1}{2^2} \left( \sum ||c_{j_1-2}||^2 + \frac{1}{2} \sum ||d_{j_1-2}||^2 + \frac{1}{2} \sum ||d_{j_1-1}||^2 \right) \]

\[ = \frac{1}{2^j} \sum ||c_{j_1-j}||^2 + \sum_{j=1}^{j} \frac{1}{2^j} ||d_{j_1-j}||^2 (6.19) \]

and the energy for the RDWT coefficients is given by

\[ E = \sum ||c_{j_1 - j}||^2 + \sum_{j=1}^{j} \sum ||d_{j_1-j}||^2 \] (6.20)

so that,

\[ 2^j \sum ||c_{j_1}||^2 - E = \sum_{j=1}^{j} \sum ||d_{j_1-j}||^2 \]

(since j=1...J-1, we have \(2^{j-j} \geq 2\), so that)

\[ 2^j \sum ||c_{j_1}||^2 - E \geq 0 \]

\[ \Rightarrow E \leq 2^j \sum ||c_{j_1}||^2 \] (6.21)
On the other hand,

\[ E - 2 \sum \|c_{j1}\|^2 = (1 - 2^{1-j}) \sum \|c_{j1-j}\|^2 + \sum_{j=2}^{J} (1 - 2^{1-j}) \sum \|d_{j1-j}\|^2 \]

(since \(J\geq1\) and \(j=2\ldots J\), we have \(2^{1-j} \leq 1\), so that)

\[ E - 2 \sum \|c_{j1}\|^2 \geq 0 \]

\[ \Rightarrow E \geq 2 \sum \|c_{j1}\|^2 \quad (6.22) \]

The bounds of \(A=2\) and \(B = 2^J\) are the tightest bounds which give sequences that meet the bounds. For a constant sequence \(x_1[n] = 1\), only the lowpass coefficients are nonzero and all highpass subbands are zero valued. That is, \(\sum \|c_{j1-j}\|^2 \neq 0\) but \(\sum \|d_{j1-j}\|^2 = 0\), for \(j = 1 \ldots J\).

\[ 2^J \sum \|c_{j1}\|^2 - E = \sum_{j=1}^{J-1} (2^{j-1} - 1) \sum \|d_{j1-j}\|^2 = 0 \]

\[ \therefore E = 2^J \sum \|c_{j1}\|^2 \quad (6.23) \]

For an oscillatory sequence \(x_2[n] = (-1)^n\), only the finest detail coefficients would be nonzero. That is,

\[ \sum \|c_{j1-j}\|^2 = 0 \text{ and } \sum \|d_{j1-j}\|^2 = 0 \text{, for } j = 2 \ldots J. \]

\[ E - 2 \sum \|c_{j1}\|^2 - (1 - 2^{1-j}) \sum \|c_{j1-j}\|^2 + \sum_{j=2}^{J} (1 - 2^{1-j}) \sum \|d_{j1-j}\|^2 = 0 \]

\[ \therefore E = 2 \sum \|c_{j1}\|^2 \quad (6.24) \]
Therefore, for decomposition level $J > 1$, the energy of the decomposition coefficients are well bounded and hence, the RDWT is a frame expansion.

### 6.2.4.3 Robustness of Transforms

The three different transforms, i.e., orthonormal basis, tight frame and frame basis, in their corresponding transform domains are as given below and the robustness is measured as the Mean Square Error (MSE) of the reconstructed signal with the original signal.

$$
MSE = E[\|f - \hat{f}\|^2] = E[<f - \hat{f}, f - \hat{f}>]
$$

$$
= E[<f, f>] - 2E[<f, \hat{f}>] + E[<\hat{f}, \hat{f}>]
$$

where $f$ is the original signal and $\hat{f}$ is the reconstructed signal.

### 6.2.4.4 Orthonormal basis

$$
f = \sum_{i=1}^{N} \alpha_{i} \psi_{i} = \sum_{i=1}^{N} (\psi_{i}, f) \psi_{i}
$$

$$
\hat{f} = \sum_{i=1}^{N} (\alpha_{i} + n_{i}) \psi_{i}
$$

where the Gaussian noise is $n_{i} \sim (0, \mathcal{C})$.

The error signal

$$
e = \hat{f} - f = \sum_{i=1}^{N} n_{i} \psi_{i}
$$
So that,

\[
MSE = E\|e\|^2 = E \left[ \sum_{i=1}^{N} n_i^2 \right]
\]

\[
= \sum_{i=1}^{N} E[n_i^2] = N \varepsilon^2
\]  

(6.25)

### 6.2.4.5 Tight Frame

\[
f = \frac{1}{A} \sum_{\ell=1}^{AN} \langle \psi_{\ell}, f \rangle = \frac{1}{A} \sum_{i=1}^{AN} \alpha_i \psi_i
\]

\[
\hat{f} = \frac{1}{A} \sum_{i=1}^{AN} (\alpha_i + n_i) \psi_i
\]

The MSE is given by,

\[
MSE = E[\langle f, f \rangle] - 2E[\langle f, \hat{f} \rangle] + E[\langle \hat{f}, \hat{f} \rangle]
\]

\[
MSE = E[\|f\|^2] - E[\langle \hat{f}, \hat{f} \rangle] + E[\langle f, \hat{f} \rangle]^2
\]

\[
= \frac{1}{A^2} \sum_{\ell} \sum_{j} E[n_{\ell} n_{j}] \langle \psi_{\ell}, \psi_{j} \rangle
\]

\[
= \frac{1}{A^2} \sum_{\ell} E[n_{\ell}^2] \langle \psi_{\ell}, \psi_{\ell} \rangle
\]

\[
= \frac{1}{A^2} \sum_{\ell=1}^{AN} \varepsilon^2 = N \varepsilon^2 / A
\]  

(6.26)

Since \(A > 1\), \(\frac{N \varepsilon^2}{A} < N \varepsilon^2\), and it yields less distortion to \(f\) than does an orthonormal basis.
6.2.4.6 Frame Expansion

\[ f = \sum_{i=1}^{M} \alpha_i \psi_i = \sum_{i=1}^{M} \langle \psi_i, f \rangle \psi_i \]

\[ \hat{f} = \sum_{i=1}^{M} (\alpha_i + n_i) \psi_i \]

\[ MSE = E[< f, f >] - 2E[< f, \hat{f} >] + E[< \hat{f}, \hat{f} >] \]

Combining all the components:

\[ MSE = E[< f, f >] - 2E[< f, \hat{f} >] + E[< \hat{f}, \hat{f} >] = \sum_i E[n_i^2] \langle \psi_i, \psi_i \rangle \]

\[ = c^2 \sum_i \| \psi_i \|^2 \quad (6.27) \]

6.2.4.7 Decimation in RDWT

The decimators removed redundant coefficients, which are not necessary to perfectly reconstruct the signal. This makes wavelet compression algorithms more computationally efficient. However, in some wavelet techniques, the redundant coefficients are useful. For example, this work interpolates missing pixels of an image, in the wavelet domain, based on the values of the surrounding pixels, as much information as possible is needed to accurately interpolate the missing pixels. Decimation removes potentially valuable information. In cases like this, it is beneficial to remove the decimators. This is known as the Redundant Discrete Wavelet Transform (RDWT), as shown in Figure 6.13.
6.3 PROPOSED WAVELET TRANSFORM BASED FEATURE EXTRACTION

In 2D wavelet transforms, an image is decomposed into one approximation and three detail images. The approximation and the detail images are then decomposed into a second-level approximation and detail images, and the process is repeated. The standard 2D-WT can be implemented with a low-pass filter \( h \) and a high-pass filter. The 2D-WT of an \( N \times M \) discrete image \( X \) up to level \( P+1 \) (\( P \leq \min(\log_2 N, \log_2 M) \)) is recursively defined in terms of the coefficients at level \( p \) as follows:

\[
C_{4k,(i,j)}^{p+1} = \sum_{m} \sum_{n} h(m)h(n)C_{4k,(m+2i,n+2j)}^{p} \quad (6.28)
\]

\[
C_{4k+1,(i,j)}^{p+1} = \sum_{m} \sum_{n} h(m)g(n)C_{4k,(m+2i,n+2j)}^{p} \quad (6.29)
\]

\[
C_{4k+2,(i,j)}^{p+1} = \sum_{m} \sum_{n} g(m)h(n)C_{4k,(m+2i,n+2j)}^{p} \quad (6.30)
\]
\[ C_{4k+3i,j}^{p+1} = \sum_{m} \sum_{n} g(m) g(n) C_{4k+2i+3j}^{p} \]

(6.31)

Where \( C_{0,i,j}^{0} \) is the image X. At each step, the image \( C_{k}^{p} \) is decomposed into four quarter-size images \( C_{4k}^{p+1}, C_{4k+1}^{p+1}, C_{4k+2}^{p+1}, C_{4k+3}^{p+1} \). 2D-WT decomposition allows us to analyze an image simultaneously at different resolution levels and orientations. Several energy functions can be used to extract features from each subband for classification. Commonly used energy functions include magnitude, magnitude square and the rectified sigmoid. Both magnitude and magnitude square are parameter free while the rectified sigmoid function can be adjusted by the parameter \( \alpha \). The mean and the standard deviation of subband coefficients are also extracted as features. Basically, all these definitions of energy are highly correlated. In this research, the definition of energy based on squaring is used. The energy in different subbands is computed from the subband coefficient matrix as

\[ \sigma_{p}^{2}(k) = \sum_{i} \sum_{j} [C_{k}^{p}(i,j)]^{2} \]  

(6.32)

Where \( \sigma_{p}^{2}(k) \) is the energy of the texture projected to the subspace at node (p, k). The energy of each subband provides a measure of the image characteristics in that subband. The energy distribution has important discriminatory properties for texture images and as such can be used as a feature for texture matching. A simple illustration of energy distribution over different scales and orientations is given in Figure 6.14.
Figure 6.14 Energy values from the subbands, (a) Original Shoeprint, (b) Scaled Shoeprint Images

The early research in wavelet packet-based feature extraction focuses on extracting energy values from all subbands and using them for classification. However, it is common knowledge in the area of pattern recognition that proper feature selection is likely to improve classification accuracy. At the same time, the structure of wavelet packet transform, i.e. the
overcomplete tree representation, motivates the proper subband selection for classification. In pattern classification, the goal of subband selection is to obtain a sparse representation of features that can achieve high classification accuracy. It is not required that the texture image is reconstructed from the selected subbands. To generate a sparse representation that can achieve accurate classification results, the energy values for the selected subbands should be as independent of each other as possible. In past research, each subband is evaluated separately based on a pre-determined criterion and the feature selection is based on the individual evaluation results. In this process, it is implicitly assumed that the energy values from different subbands are independent, which rarely holds in practice. In order to have a sparse subband selection process that can achieve good performance for pattern classification; we need to know the dependence between the energy values extracted from different subbands. The dependence between wavelet coefficients that have the ‘parent–child’ or ‘sibling’ relationship has been successfully measured by mutual information, or modeled by hidden Markov model (HMM). However, the energy value is a function of all coefficients in a sub-band. Thus, the dependence between energy values is more complex than the dependence between individual wavelet coefficients. In the wavelet transform based pattern classification, we need to consider the dependence between the energy values from different sets of subbands. If the energy values from a set of subbands are dependent on the energy values from another set of subbands, the latter set of subbands should not be included in the sparse representation for pattern classification. In this research work, we analyze the dependence of energy values from different subbands, which may be from the same wavelet basis or different wavelet bases. Based on the theoretical analysis, we propose a mutual information measure to select subbands for sparse representation of texture for shoeprint image classification, and the proposed method is called as Subband Relevance based Wavelet Transform (SR-WT).
6.3.1 Analysis of Dependence Between Subbands

In this section, covariance is used to analyze the dependence between energy values from any two subbands. Based on Eqs. (6.28)–(6.31), the wavelet coefficients at level (p+1) are obtained by convolving the filters with the wavelet coefficients at level p. The lowest level in the wavelet packet decomposition tree is the original image X. Therefore, the coefficients in any subband can be written as a linear combination of pixel values in the original image X. The weighting coefficients in the linear combination are determined by the properties of the wavelet basis, i.e., the lengths and the coefficient values for filters g and h. Considering the definition of subband energy in Eq. (6.32), the energy of wavelet coefficients in a subband is a second-order polynomial in terms of the pixel values of the image X. Coefficients in the polynomials are determined by the properties of the filters g and h and the decomposition level as follows:

\[
\sigma_p^2(k) = \sum_i \sum_j \sum_{i'} \sum_{j'} f_k^p(i, j, i', j')X(i, j)X(i', j')
\]  

(6.33)

where X is the original image. Due to downsampling, some of the coefficients \(f_k^p(i, j, i', j')\) are zero. Hence, the covariance between two energy values is defined as

\[
\text{Cov} \left( \sigma_{p1}^2(k), \sigma_{p2}^2(k) \right) = E[\sigma_{p1}^2(k)\sigma_{p2}^2(k)] - E[\sigma_{p1}^2(k)]E[\sigma_{p2}^2(k)]
\]

(6.34)

\(\text{Cov} \left( \sigma_{p1}^2(k), \sigma_{p2}^2(k) \right)\) will be a fourth-ordered polynomial in terms of the pixel values in the image X. Using definitions (6.28)–(6.31), the correlation between energy values of two children nodes \(\sigma_1^2(0)\) and \(\sigma_1^2(1)\) is
\[
E[\sigma_1^2(0)\sigma_1^2(1)]
= E \left[ \sum_l \sum_j \left( \sum_m \sum_n h(m)h(n) x(m + 2i, n + 2j) \right)^2 \times \sum_p \sum_q \left( \sum_l \sum_r h(l)g(r) x(l + 2p, r + 2q) \right)^2 \right]
\]

(6.35)

Given that \( g \) and \( h \) are FIR filters, the random variables in Eq. (6.35) are the different pixel values. Therefore, the covariance is a function of the fourth-order statistics of the original image \( X \). If we have a statistical model for the pixels, the correlation value can easily be calculated. To illustrate this, suppose that the image \( X \) is of size \( 4 \times 4 \). The image model can be described using AR-1 Gaussian model, which is simple but often used in digital image processing. A useful feature of AR-1 Gaussian model is that the marginal distribution of each pixel is also Gaussian. The Gaussian AR-1 process with mean \( \mu \) is usually written in terms of a series of white noise innovation processes \( \{E_n\} \):

\[
X_n - \mu = a (X_{n-1} - \mu) + E_n
\]

(6.36)

Where \( E_n \sim N(0, \sigma^2) \) are independent and identically distributed. and \( |a| < 1 \). The marginal distribution is also normal:

\[
X_n \sim N \left( \mu, \frac{\sigma^2}{1 - a^2} \right)
\]

(6.37)

Given the marginal distribution, Eq. (6.35) can be expanded as follows:
\[ E[\sigma_1^2(0)\sigma_1^2(1)] = v_1(h, g)E[X_1^4] + v_2(h, g)E[X_1^3X_2] + v_3(h, g)E[X_1^2X_2^2] + v_4(h, g)E[X_1^2X_2X_3] + v_5(h, g)E[X_1X_2X_3X_4] \] (6.38)

Where \( \{X_1, X_2, X_3, X_4\} \) are independent and identically distributed. Gaussian random variables corresponding to different pixels distributed as given by Eq. (6.37), and \( \{v_1(h, g), v_2(h, g), v_3(h, g), v_4(h, g)\} \) are functions of wavelet filters.

Suppose that the size of the original image X is NM, and define \( K=\min(N, M) \). \( v_1(h, g)E[X_1^4] \) represents those terms on the right-hand side of Eq. (6.38) where the indices of four pixel values are exactly the same. Thus, \( v_1(h, g)E[X_1^4] \) contains exactly \( K \) terms. Similarly, \( v_2(h, g)E[X_1^3X_2] \) contains \( 4K (K - 1) \) terms, \( v_3(h, g)E[X_1^2X_2^2] \) contains \( 3K (K - 1) \), \( v_4(h, g)E[X_1^2X_2X_3] \) contains \( 6K (K - 1) \) and \( v_5(h, g)E[X_1X_2X_3X_4] \) contains \( K (K - 1) (K - 2) (K - 3) \) terms from the right-hand side of Eq. (6.38). Based on the marginal distribution, the four expectation values are given as follows:

\[ E[X_1^4] = \mu^4 + 6\mu^2 \frac{\sigma^2}{1 - a^2} + 3 \left( \frac{\sigma^2}{1 - a^2} \right)^2 \]
\[ E[X_1^3X_2] = \mu^4 + 3\mu^2 \frac{\sigma^2}{1 - a^2} \]
\[ E[X_1^2X_2^2] = \left( \mu^2 + \frac{\sigma^2}{1 - a^2} \right)^2 \]
\[ E[X_1^2X_2X_3] = \mu^4 + \mu^2 \frac{\sigma^2}{1 - a^2} \]
\[ E[X_1X_2X_3X_4] = \mu^4 \] (6.39)
Therefore, we can obtain the covariance between the energy features from two subbands analytically. However, covariance is only defined to describe the dependence between two variables, not between a single variable and a set of variables. Moreover, the correlation coefficient and covariance are inadequate for describing the dependence when the distributions are non-Gaussian. For these reasons, mutual information, which can be used to measure the dependence between a single variable and a set of variables, is considered. Mutual information has been widely used for image registration, feature selection and classification.

Consider two random variables \( X \in \mathcal{R} \) and \( Y \in \mathcal{Y} \) without a joint distribution \( p(x, y) \). The mutual information between \( X \) and \( Y \) is defined by

\[
I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E_{xy} \left[ \log \frac{p(x, y)}{p(x)p(y)} \right]
\]

\[
= D(p(x, y) \mid \mid p(x)p(y))
\]  

(6.40)

Where \( D \) is the relative entropy of Kullback-Leibler distance between two probability mass functions \( p(x, y) \) and \( p(x)p(y) \). When the logarithm function in the definition uses a base of 2, the unit for the \( I(X; Y) \) is bits. The mutual information is symmetric in \( X \) and \( Y \), nonnegative, and is equal to zero if and only if \( X \) and \( Y \) are independent. The mutual information \( I(X; Y) \) indicates how much information \( Y \) conveys about \( X \). Given \( Y \), the extra information required to fully describe \( X \) is given by conditional entropy \( H(X \mid Y) \). Thus, the following equation holds:

\[
I(X; Y) = H(X) - H(X \mid Y)
\]

(6.41)

where \( H(X) \) is the entropy of random variable \( X \). Similar to the definition of conditional entropy, the conditional mutual information between random variables \( X \) and \( Y \) given \( Z \) is defined by
\[
I(X;Y|Z) = H(X|Z) - H(X|Y,Z)
= E_{p(x,y,z)} \left[ \log \frac{P(X,Y|Z)}{P(X|Z)P(Y|Z)} \right]
\] (6.42)

The conditional information has an interpretation similar to that of mutual information. Given the above definitions, the mutual information between a set of random variables \( \{X_1, X_2, \ldots, X_n\} \) and a single random variable \( Y \) can be defined by

\[
I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, X_{i-2}, \ldots, X_1)
\] (6.43)

This definition of mutual information can be used for evaluating the independence between the features extracted from a set of subbands \( U \) and a single subband \( WS_i(k) \). The higher the value of the mutual information, the easier it is to estimate the distribution of \( WS_i(k) \) given \( U \). The mutual information thus provides a criterion for the subband selection process. Based on those subbands, the energy calculation is done with CWT, DWT and RDWT based feature extraction, which are used for classification.

6.4 EXPERIMENTS AND PERFORMANCE ANALYSIS

The same set of shoeprint image database as discussed in the previous chapters is used here for the experiment. Initially the features are extracted from the shoeprint images using CWT, DWT, and RDWT also with the proposed feature extraction approaches SR-CWT (Subband Relevance based CWT), SR-DWT, and SR-RDWT. Then the features are classified with the SVM classifier as described in section 5.3. The following Table 6.1 shows the energy extracted for a shoeprint image with the above said feature extraction approaches.
Table 6.1 Wavelet Energy Features Extracted from the Shoeprint Images

<table>
<thead>
<tr>
<th>Feature Extraction Methods</th>
<th>Energy Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ea</td>
</tr>
<tr>
<td>SR-RDWT</td>
<td>0.1668</td>
</tr>
<tr>
<td>SR-DWT</td>
<td>0.1668</td>
</tr>
<tr>
<td>SR-CWT</td>
<td>0.1667</td>
</tr>
<tr>
<td>RDWT</td>
<td>0.1667</td>
</tr>
<tr>
<td>DWT</td>
<td>0.1666</td>
</tr>
<tr>
<td>CWT</td>
<td>0.1664</td>
</tr>
</tbody>
</table>

Likewise the energy values for the approximation (Ea), horizontal (Eh), vertical (Ev) and diagonal (Ed) coefficients are calculated as features for the shoeprint images. Then the classification is performed with the SVM classifier, the Table 6.2 quantifies the classification results obtained with the proposed wavelet transform based feature extraction, compared against the standard wavelet transform based approaches. Figure 6.15 and 6.16 illustrates the same in terms of ROC curve for full and partial shoeprint recognition systems respectively.

Table 6.2 Classification Performance for the WT based Feature Extraction

<table>
<thead>
<tr>
<th>Feature Extraction Methods With SVM Classifier</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Shoeprint</td>
</tr>
<tr>
<td>SR-RDWT</td>
<td>94.82</td>
</tr>
<tr>
<td>SR-DWT</td>
<td>93.16</td>
</tr>
<tr>
<td>SR-CWT</td>
<td>91.57</td>
</tr>
<tr>
<td>RDWT</td>
<td>91.53</td>
</tr>
<tr>
<td>DWT</td>
<td>91.07</td>
</tr>
<tr>
<td>CWT</td>
<td>90.16</td>
</tr>
</tbody>
</table>
Figure 6.15 ROC Analyses on Wavelet Transform based Feature Extraction for full Shoeprint Recognition

Figure 6.16 ROC Analyses on Wavelet Transform based Feature Extraction for partial Shoeprint Recognition

And the following Table 6.3 summarizes the comparison between the proposed shoeprint recognition systems with other existing approaches.
Table 6.3 Performance Analysis of SR-WT based Shoeprint Recognition

<table>
<thead>
<tr>
<th>References</th>
<th>Feature Extraction</th>
<th>Classification</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed System for Full Shoeprint (FS)</td>
<td>SR-RDWT</td>
<td>SVM</td>
<td>94.82</td>
</tr>
<tr>
<td></td>
<td>SR-DWT</td>
<td>SVM</td>
<td>93.16</td>
</tr>
<tr>
<td></td>
<td>SR-CWT</td>
<td>SVM</td>
<td>91.57</td>
</tr>
<tr>
<td>Proposed System for Partial Shoeprint (PS)</td>
<td>SR-RDWT</td>
<td>SVM</td>
<td>93.36</td>
</tr>
<tr>
<td></td>
<td>SR-DWT</td>
<td>SVM</td>
<td>92.73</td>
</tr>
<tr>
<td></td>
<td>SR-CWT</td>
<td>SVM</td>
<td>91.39</td>
</tr>
<tr>
<td>Proposed System for Full Shoeprint (FS)</td>
<td>PCA</td>
<td>SVM</td>
<td>89.88</td>
</tr>
<tr>
<td>Proposed System for Partial Shoeprint (PS)</td>
<td>PCA</td>
<td>SVM</td>
<td>87.65</td>
</tr>
<tr>
<td>Proposed DCT + FLD</td>
<td>DCT</td>
<td>FLD</td>
<td>87.23</td>
</tr>
<tr>
<td>Jonsson et al., (2000)</td>
<td>PCA</td>
<td>SVM</td>
<td>87.62</td>
</tr>
<tr>
<td>Su et al., (2007)</td>
<td>LIF</td>
<td>kNN</td>
<td>81.22</td>
</tr>
<tr>
<td>Wang et al., (2009)</td>
<td>WT</td>
<td>ANN</td>
<td>86.27</td>
</tr>
<tr>
<td>Bouridane et al., (2000)</td>
<td>Fractal</td>
<td>Euclidian</td>
<td>75.05</td>
</tr>
<tr>
<td>Gharsa et al., (2008)</td>
<td>Hu’s moments</td>
<td>Euclidian</td>
<td>86.59</td>
</tr>
</tbody>
</table>

6.5 SUMMARY

In this chapter a novel wavelet transform approach is proposed for extracting the efficient features from the shoeprint images. Energy values from different subbands of wavelet decomposition are often used for shoeprint classification. In this chapter, the dependence of energy values from different subbands are analyzed, which may be from the same wavelet basis or multiple wavelet bases. Based on this analysis, a novel algorithm is proposed to choose independent subbands to form a sparse representation of texture images for classification. Mutual information is used as the criterion to measure the dependence between energy values from different subbands. Based on the dependency the features are extracted from the subbands, and then the features are classified with SVM classifiers. The experimental result shows that the proposed wavelet based feature extraction outperforms the other approaches.