Wavelet Analysis and its Applications to Pattern Recognition

Tien D. BUI,
Concordia University
Montreal, Quebec, Canada H3G 1M8

Abstract: In this paper we will describe some of the research done in using wavelet transforms for pattern recognition at the Centre for Research in Pattern Recognition and Machine Intelligence at Concordia University. This includes road signs recognition, face detection and recognition, oriental characters recognition, hand-written numerals and characters, image compression, and image denoising, etc. We have developed a new metric called the Hellinger-Kakutani metric for measuring the distance between two descriptors for classification. This metric is filter-invariant, hence can be used on noisy images.

Key-Words: Pattern recognition, image processing, multiwavelets, translation-invariance, classification metric.

1 Introduction

Wavelet analysis is one of the latest and most significant mathematical tools developed in the last century. Both wavelet analysis and its applications have become fastest growing research areas in recent years. There are already hundreds of patents in which wavelet plays a key role. Wavelet theory has been employed in many areas of research such as numerical analysis, signal and image processing, communications, biomedical imaging, electromagnetics, radars and acoustics....However the research on applying wavelets to pattern recognition is still at its infancy; only a few publications deal with this topic at the present. In recent past, significant advances have been made in the field of pattern recognition and computer vision, but computers still fall far short of humans and animals in their visual performance. Many scientists and engineers devote their great efforts to solve this difficult problem. Indeed, machine recognition of different patterns such as printed and hand-written characters, fingerprints, biomedical images, human faces and expressions, road signs and many others has been intensely researched by scientists in different countries around the world. Although a lot of achievements have been made in this area, many problems are still unresolved.

A pattern recognition system commonly consists of two key components: feature extraction and pattern classification. The success of such a system depends not only on the effectiveness of each component, but also on the coordination of their execution. The feature extraction process has two major objectives: Determination of certain attributes of the pattern classes which are invariant to as many kinds of distortion as possible; and reduction of the dimensionality of the feature vector by setting the most discriminatory characteristics of these patterns. On the other hand, an effective pattern classification system requires a distance measure which is independent of the linear filtering process (to obtain a desired descriptor) and is robust to corruption noise. A number of methods extract 1-D features from 2-D patterns. This approach reduces the size of the database and the time for matching through the database, but suffers a drawback that the recognition rate may not be high. Hence, in this paper, we introduce a 2-D descriptor for pattern recognition.

2 An invariant Fourier descriptor

An invariant Fourier descriptor was developed by (Zhang, Suen and Bui (1996)). This descriptor is invariant to translation, rotation and scaling. The computation of this descriptor is given below:

1. Obtain a 2-D Fourier transform of the original binary image \( \{ X(s, t); 1 \leq s \leq M, 1 \leq t \leq N \} : F(u, v) = \mathcal{F}(X(s, t)) \).

2. Transform \( |F(u, v)| \) in the Cartesian coordinates into an image in the polar coordinates...
3. Apply a 1-D Fourier transform on $G(r, \theta)$ along the axis of the polar angle $\theta$ and obtain its spectrum: $H(r, \phi) = |F_{\theta}(G(r, \theta))|$

3 An invariant Fourier-Wavelet descriptor

Another invariant descriptor developed by (Bui and Chen (1999)) which has become very useful in recent works. The steps of the algorithm called PFW can be summarized as follows:

1. Find the centroid of the pattern $f(x, y)$ and transform $f(x, y)$ into polar coordinate system to obtain $g(r, \theta)$.

2. Conduct 1-D Fourier transform on $g(r, \theta)$ along the axis of polar angle $\theta$ and obtain its spectrum:

$$G(r, \phi) = |F_{\theta}(g(r, \theta))|.$$ 

3. Apply 1-D wavelet transform on $G(r, \phi)$ along the axis of radius $r$:

$$WF(r, \phi) = W_{r}(G(r, \phi)).$$

4. Use the wavelet coefficients to query the pattern feature database at different resolution levels.

4 The Hellinger-Kakutani metric

Given two spectral densities $f$ and $g$, representing the random fields $\{X(s, t); s, t \in \mathbb{Z}\}$ and $\{Y(s, t); s, t \in \mathbb{Z}\}$ respectively, a commonly used metric to measure the distance between $f$ and $g$ is the Euclidean metric:

$$d_1(f, g) = \left( \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} (f(\lambda, \mu) - g(\lambda, \mu))^2 d\lambda d\mu \right)^{\frac{1}{2}}$$

(1)

Another popular measure is the Kullback-Leibler discrimination information rate:

$$d_2(f, g) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left( \frac{f(\lambda, \mu)}{g(\lambda, \mu)} - 1 - \ln \frac{f(\lambda, \mu)}{g(\lambda, \mu)} \right) d\lambda d\mu.$$ 

(2)

Statistical estimation of random fields based on the information rate $d_2(f, g)$ is known to yield the asymptotic maximum likelihood estimators of the parameters of the random fields (Anh and Lunney (1995)). However $d_2$ is not a metric as it does not satisfy the triangle inequality (Parzen (1983)). In this paper, we suggest the following metric based on the Hellinger-Kakutani integral:

$$d_3(f, g) = \left( \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \left( \frac{f(\lambda, \mu) + g(\lambda, \mu)}{2\sqrt{f(\lambda, \mu)g(\lambda, \mu)}} \right) d\lambda d\mu \right)^{\frac{1}{2}}$$

(3)

assuming the usual condition

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln f(\lambda, \mu) d\lambda d\mu > -\infty.$$ 

(4)

It is seen that $d_3$ is symmetric, that is,

$$d_3(f, g) = d_3(g, f).$$

(5)

Also,

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \left( \frac{f(\lambda, \mu) + g(\lambda, \mu)}{2\sqrt{f(\lambda, \mu)g(\lambda, \mu)}} \right) d\lambda d\mu =$$

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \left( \frac{f(\lambda, \mu) - g(\lambda, \mu)}{2\sqrt{f(\lambda, \mu)g(\lambda, \mu)}} + 1 \right) d\lambda d\mu \geq 0;$$ and $d_3(f, g) = 0$ iff $f = g$ a.e.. We can prove that $d_3$ behaves as a metric asymptotically.

Let us now look at the effect of linear filtering on $d_1$, $d_2$ and $d_3$. The coordinate transformation of the above algorithms for the invariant Fourier and Fourier-Wavelet descriptors, the moment Fourier transform (Wang et al. (1994)) and the wavelet transform (Chang and Kuo (1996)) are special cases of linear filtering. In a general setting, a linear filter is characterised by a transfer function $a(\lambda, \mu)$. Then, passing an image with spectrum $f_i(\lambda, \mu)$ through the linear filter will result in the spectrum

$$g_i(\lambda, \mu) = |a(\lambda, \mu)|^2 f_i(\lambda, \mu), i = 1, 2.$$ 

(7)

Consequently, it follows from definition that

$$d_1(g_1, g_2) =$$

$$\left( \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |a(\lambda, \mu)|^4 (f_1(\lambda, \mu) - f_2(\lambda, \mu))^2 d\lambda d\mu \right)^{\frac{1}{2}}$$

(8)

$$d_2(g_1, g_2) = d_2(f_1, f_2); \quad d_3(g_1, g_2) = d_3(f_1, f_2).$$

(9)

In other words, while $d_1$ depends on the transfer function $a(\lambda, \mu)$, $d_2$ and $d_3$ are invariant with respect to linear filtering. It is therefore advantageous to use $d_3$ since it is then a filtration-invariant metric.
5 Discrete Multiwavelet Transform

Multiwavelets have some advantages in comparison to scalar ones. For example, such features as short support, orthogonality, symmetry, and higher order of vanishing moments, are known to be important in signal processing. A scalar wavelet cannot possess all these properties at the same time. Therefore, multiwavelets can give better results than the scalar wavelets in image compression and de-noising (Strela et al. (1995)). Multiwavelet basis uses translations and dilations of \( L \geq 2 \) scaling functions \( \{\phi_k(x)\}_{1 \leq k \leq L} \) and \( L \) mother wavelet functions \( \{\psi_k(x)\}_{1 \leq k \leq L} \). If we write \( \Phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_L(x))^T \) and \( \Psi(x) = (\psi_1(x), \psi_2(x), \ldots, \psi_L(x))^T \), then we have

\[
\Phi(x) = 2^{N-1} \sum_{k=0}^{2N-1} H_k \Phi(2x - k), \quad (10)
\]

and

\[
\Psi(x) = 2^{N-1} \sum_{k=0}^{2N-1} G_k \Phi(2x - k). \quad (11)
\]

where \( \{H_k\}_{0 \leq k \leq 2N-1} \) and \( \{G_k\}_{0 \leq k \leq 2N-1} \) are \( L \times L \) filter matrices.

As an example, we give the most commonly used multiwavelets developed by Geronimo, Hardin and Massopust (1994). Let

\[
H_0 = \begin{pmatrix} 3/10 & 2\sqrt{2}/5 \\ -\sqrt{2}/20 & -3/20 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 3/10 & 0 \\ 9\sqrt{2}/40 & 1/2 \end{pmatrix},
\]

\[
H_2 = \begin{pmatrix} 0 & 2\sqrt{2}/5 \\ 9\sqrt{2}/40 & -3/20 \end{pmatrix}, \quad H_3 = \begin{pmatrix} 0 & 0 \\ -\sqrt{2}/20 & 0 \end{pmatrix},
\]

and

\[
G_0 = \begin{pmatrix} -\sqrt{2}/20 & -3/20 \\ 1/20 & -3\sqrt{2}/20 \end{pmatrix}, \quad G_1 = \begin{pmatrix} 9\sqrt{2}/40 & -1/2 \\ 9/20 & 0 \end{pmatrix},
\]

\[
G_2 = \begin{pmatrix} 9\sqrt{2}/40 & -3/20 \\ -9/20 & 3\sqrt{2}/20 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 1/20 & 0 \\ 0 & 0 \end{pmatrix}.
\]

then the two functions \( \phi_1(x) \) and \( \phi_2(x) \) can be generated via \( (10) \). Similarly, the two mother wavelet functions \( \psi_1(x) \) and \( \psi_2(x) \) can be constructed by \( (11) \).

Let \( V_j \) be the closure of the linear span of \( 2^j/\phi_i(2^j t - k), l = 1, 2; k \in Z \). With the above constructions, it has been proved that \( \phi_i(t - k), l = 1, 2; k \in Z \) form an orthonormal basis for \( V_0 \), and moreover the dilations and translations \( 2^j/\psi_i(2^j - k), l = 1, 2; j, k \in Z \) form an orthonormal basis for \( L^2(R) \). In other words, the spaces \( V_j, j \in Z \), form an orthogonal multiresolution analysis of \( L^2(R) \). Let

\[
H(w) = \sum_{k=0}^{3} H_2 e^{i\omega k}, \quad (12)
\]

\[
G(w) = \sum_{k=0}^{3} G_3 e^{i\omega k}. \quad (13)
\]

From the orthogonality, we have

\[
H(w)H^T(w) + H(w + \pi)H^T(w + \pi) = I_2 \quad (14)
\]

\[
G(w)G^T(w) + G(w + \pi)G^T(w + \pi) = I_2 \quad (15)
\]

\[
H(w)G^T(w) + H(w + \pi)G^T(w + \pi) = 0_2 \quad (16)
\]

where \( T \) means the complex conjugate transpose, \( I_2 \) and \( 0_2 \) denote the \( 2 \times 2 \) identity and all zero matrix, respectively.

6 De-noising

The noisy image is transformed into multiwavelet domain by applying translation invariant multiwavelets along the rows and columns, respectively. For example, the GHM multiwavelet transform needs two rows for the input, we group two adjacent rows when we do the transform along row and two adjacent columns when transforming along columns. A soft or hard thresholding [7-9] is applied to the resulting multiwavelet coefficients. In order to get the denoised image, we perform inverse multiwavelet transform along the columns and rows, respectively.

At the first glance, this proposal seems to be impractical because a naive implementation of the method increases the complexity of wavelet shrinkage algorithm from \( O(n^2) \) to \( O(n^4) \) for \( n \times n \) images. Fortunately, the \( O(n \log n) \) fast algorithm in 1-D can easily be generalised to 2-D case so that the algorithm can be implemented with \( O(n \log n)^2 \) complexity.

The choice of the threshold is critical in wavelet shrinkage. If it is too small or too big then the wavelet shrinkage estimator will tend to overfit or underfit the data. Donoho and Johnstone(1994) proposed the universal threshold \( \sigma \sqrt{2\log n} \) where \( n \) is the number of points in the signal. Despite the simplicity of such a threshold, they showed that the resulting nonlinear wavelet estimator is spatially adaptive and is asymptotically near-minimax within the whole range of Besov spaces. In the case of image de-noising the threshold can be adjusted to be \( \sigma \sqrt{2\log n^2} = 2\sigma \sqrt{\log n} \). However, we notice that this threshold is far too big for image de-noising.

7 Experimental Results

We explore the applications of the TI multiwavelet transform in image de-noising. In order to compare the results with other papers, we use the famous image, Lenna, in our experiments. The dimension of the image is 256 x 256 pixels with each pixel having a gray level ranging from 0 to 255. The multiple wavelets used in our experiments is the GHM multiwavelets. Even though hard thresholding can be used in the
thresholding process, we only use soft thresholding. The reason for this is that hard thresholding does not perform as well as soft thresholding in our previous experiments for signal de-noising (Bui and Chen (1998)). The results of this work for image de-noising were published in a separate paper (Bui, Chen and Roy (1999)).

The techniques described in this paper have been used in various problems: The road sign recognition problem was studied in a thesis by Daqing Li ( Li (1998)). Face detection and recognition was studied by SiNguyen Vo (Vo (2001)). The problem of recognizing oriental characters that are very similar was investigated in (Anh et al. (1998)). And finally, work in progress includes recognition of hand-written numerals and characters, image understanding, and document analysis.

Acknowledgements

This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada and by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche of Quebec.

References:


