WATERMARKING USING COMPLEX WAVELETS WITH RESISTANCE TO GEOMETRIC DISTORTION

Patrick Loo *Nick Kingsbury*
Department of Engineering, University of Cambridge,
Cambridge CB2 1PZ, UNITED KINGDOM
Tel: +44 1223 332767; fax: +44 1223 332662
e-mail: {pl201,ngk}@eng.cam.ac.uk
Web: www-sigproc.eng.cam.ac.uk/∼pl201,∼ngk

ABSTRACT

We present a watermarking algorithm in the complex wavelet domain and show why complex wavelets are better than real wavelets. Being an oversampled transform, the Complex Wavelet Transform requires special precautions during watermark embedding. We then model the watermarking process as a communcation channel and our results show that the complex wavelets domain has relatively higher capacity than both the spatial and the real wavelets domains. We will also outline a motion-based algorithm for image registration, which can help recovering watermarks from images suffering from geometric distortion.

1 INTRODUCTION

In recent years, digital watermarks have emerged as a means to protect the copyright of digital images. Most existing watermarking algorithms transform the host image into a *critically sampled* domain, add a suitably scaled pseudorandom sequence to the transformed image coefficients and inverse transform the modified coefficients back to obtain the watermarked image. The Discrete *Real* Wavelet Transform (DWT), the (block-based) Discrete Cosine Transform (DCT) and the Discrete Fourier Transform (DFT) are among the most popular transform domains. In general, the DWT produces watermark images with the best visual quality due to the absence of blocking artefacts. However, it has two drawbacks. The DWT lacks shift invariance, which means small shifts in input signal can cause big changes in the energy distribution of the wavelet coefficients. Secondly, the DWT has poor directional selectivity for diagonal features, which is evident from the impulse responses of the filters of individual subbands (fig. 1a). There is only one filter for diagonal features.

A straightforward way to provide shift invariance is to use the Undecimated Discrete Wavelet Transform (UDWT), but it is computationally expensive and still has poor selectivity for diagonal features, as it uses the same filters as the DWT. The Complex Wavelet Transform (CWT), on the other hand, is more computationally efficient and has only a modest amount of redundancy; yet it provides approximate shift invariance and good directional selectivity (fig. 1b).

Complex wavelets have not been widely employed in the past due to the difficulties in designing complex wavelets with perfect reconstruction (PR) properties. A new implementation of the CWT, called the Dual-Tree Complex Wavelet Transform (DT CWT) has been developed [1], which is both efficient and satisfies PR.

This paper is organised as follows. The DT CWT and the watermarking algorithm will be described in section 2. The capacity of the watermarking channel will be quantified in section 3. In section 4 we will describe our motion-based registration algorithm for combating geometric distortion. We will conclude in section 5.

2 WATERMARKING IN THE COMPLEX WAVELET DOMAIN

In contrast to conventional implementations of the CWT, which uses a *single* tree (for 1-D signals) of filters with *complex* coefficients, the DT CWT uses *two* trees, each with *real* coefficients, to give the real and the imaginary parts of the complex coefficents separately. Therefore the redundancy at the output is still 2:1. For 2-D signals, the two trees first operate on the rows and then the columns of the data. The redundancy thus rises to 4:1. At each resolution, there are six subbands, instead of three as in the DWT case. The impulse responses of the filters for each subband are shown in fig. 1b. The CWT can distinguish between opposing diagonal features, because there are separate filters oriented at $\pm 45^\circ$. The redundancy of the CWT has implications on watermark embedding.

The essence of many watermarking algorithms is the addition of a pseudorandom sequence to the host image coefficients in some domain. However, this approach needs to be modified to work in the CWT domain. Due to the inherent redudancy, some components of an arbitrary sequency may be lost during the inverse transform. The lost information correponds to the component that lies in the null space of the inverse transform.

We can reduce this information loss if we use valid wavelet coefficients (i.e. coefficients which come from the CWT of an image) as our pseudorandom sequence. In our approach, the CWT of a pseudorandom bipolar image w is computed and the resulting coefficients W are modulated by the pay-

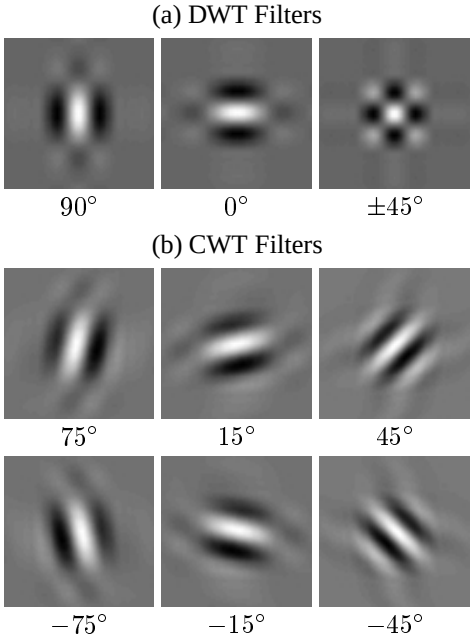

Figure 1: Filter impulse reponses in 2D (a) for the DWT and (b) for the CWT. The CWT has separate filters for $\pm 45^\circ$ and so can distinguish opposing diagonals.

load (a binary bit stream[1]) to form W', which is then scaled and added to the host image CWT coefficients X at each location u as follows:

$$Y_u = X_u + G_u \cdot W'_u \tag{1}$$

$$\text{with } G_u = \sqrt{k^2 \cdot \overline{|X|^2_U} + \gamma^2} \tag{2}$$

where Y_u is the watermarked wavelet coefficients, and $\overline{|X|^2_U}$ is the average squared magnitude in a 3×3 neighbourhood U around u. k and γ are level dependent constants which are designed to make the watermark imperceptible and yet produce significant watermark energy even in smooth regions, where the wavelet coefficients are small. Many authors (e.g. Daugman [2]) suggest that the processing of visual data within the visual cortex resembles filtering by an array of Gabor filters of different orientations and scales, which can in turn be approximated by the CWT. This means that the magnitudes of the CWT coefficients are closely related to the contrast perceived by humans. The scaling factor in (1) thus offers a simple yet effective means to adapt the watermark strength to the local image activity. This watermark embedding approach is also suitable for other oversampled filter bank transforms.

The embedding process is applied to levels 2 and 3 (level 1 is the finest level) and the watermarked CWT coefficients are inverse transformed to obtained the watermarked image y. Modifying coefficients at levels coarser than 3 tends to be relatively ineffective and to introduce visual artefacts. Figure 2 shows an example of watermarking the "Lena" image.

[1]In our implementation, a 32-bit Hadamard code (i.e. 6-bit symbol, with the MSB represented by the polarity of the codeword) is used to encode the payload prior to modulation.

We can see that the high strength areas of the watermark are aligned with the edges of the image, which is difficult to achieve with DWT watermarks, because the DWT does not have separate subbands for opposing diagonal features. Another advantage of the CWT over DWT is that the phase of the CWT coefficients allows us to infer pixel shifts quite accurately. However, no such information is available with the DWT coefficients. As we will see later, this information allows us to register a distorted image.

During watermark detection, the CWT of the received image (possibly corrupted) $\widehat{y}$ is computed and the scaling factors $\widehat{G}$ are estimated using (2), with the CWT coefficients $\widehat{Y}$ of $\widehat{y}$ replacing X. The CWT coefficents $\widehat{Y}$ are then *inversely* scaled by $\widehat{G}$, and correlated with the watermark CWT coefficients W to obtain original watermark data [2].

$$T = \sum_u \text{Re}\{(\widehat{Y_u}/\widehat{G_u}) * \text{conj}(W_u)\} \tag{3}$$

Re{.} and conj(.) denote taking the real part and conjugate respectively. "*" is pointwise multiplication and the sum is performed over all the coefficients representing a symbol.

3 CAPACITY OF THE CWT DOMAIN

The watermarking process can be analysed as a communication channel, where the watermark is the signal and the host image is the noise. Without loss of generality, we ignore additional attacks on the watermark at the moment. The simplest communication model is the Additive White Gaussian Noise (AWGN) channel, which is valid as long as all the statistics are Gaussian and the noise is simply added to the signal. The corresponding channel capacity C is then given by [3]:

$$C = \frac{1}{2}\log_2(1 + \frac{S}{N}) \tag{4}$$

where S and N are the signal and noise variance respectively.

Unfortunately, in a typical watermarking scenario, the statistics of the noise is not Gaussian and the embedding rule is not additive, because the watermark is normally scaled before being added to the host image (c.f. (1)).

Ramkumar *et al.* [4] introduced the idea of an "ideal information processor" which allows us to obtain the Gaussian equivalent variance from an arbitrary distribution. In addition, if we divide (1) by G:

$$\tilde{Y_u} = \tilde{X_u} + W'_u \tag{5}$$

where $\tilde{Y_u} = Y_u/G_u$ etc. The channel then becomes additive. It turns out that we can model the inversely scaled coefficients $\tilde{X_u}$ reasonably well with a two-Gaussian mixture with proportions p and $1-p$. We can treat this as two parallel AWGN channels and the capacity is given by:

$$C = \frac{p}{2}\log_2(1 + \frac{\sigma^2_W}{\sigma^2_{\tilde{X}1}}) + \frac{1-p}{2}\log_2(1 + \frac{\sigma^2_W}{\sigma^2_{\tilde{X}2}}) \tag{6}$$

[2]We modulate W with each codeword in the codetable in turn and then correlate with the inversely scaled coefficients. The codeword which gives the largest absolute correlation is decoded as the watermark symbol. The polarity of the correlation T gives the MSB of the symbol.

(a) (b) (c)

Figure 2: A watermark example. (a) The original image. (b) The watermarked image. (c) The enhanced watermark, showing the effect of directional filtering.

σ^2_W is the variance of W (which is the same as the variance of W'). $\sigma^2_{\bar{X}1}$ and $\sigma^2_{\bar{X}2}$ are the variances of the noise in the two-Gaussian mixtures respectively.

The data are partitioned according to p and then the variances $\sigma^2_{\bar{X}1}$ and $\sigma^2_{\bar{X}2}$ are estimated from two sets of the data respectively. At present, p is adjusted manually to obtain the best fit, although it is possible to use an iterative EM algorithm to determine the optimal value of p. An image can be regarded a set of such channels and the total capacity is given by the sum of the capacities of individual channels. The capacities of a set of test images (each of 256×256 pixels) were measured and the results are summarised in the table 1. We can see that the CWT domain has higher capacities than both the DWT and the spatial domain. The discrepancy between the theoretical and the empirical values is mainly due to the fact that (6) assumes the use of an ideal code, which is impractical.

Transform Domain	(a) Theoretical	(b) Empirical
Spatial	2.9	1.1
DWT	5.5	2.5
DT CWT	6.3	3.9

Table 1: Theoretical (a) and empirical (b) capacities in kbits of the set of 256×256 test images. For the DT CWT, the real and the imaginary parts are modelled separately as two-Gaussian mixtures and the capacities are averaged.

4 COMBATING GEOMETRIC DISTORTION

Random geometric distortion, such as that introduced by StirMark [5], is one of the most effective attacks on watermarks. Each pixel is shifted by a small amount, so that the resulting image still resembles the original. These small distortions are enough to jeopardise many watermark detectors. Insertion of a template (e.g. Pereira *et. al.* [6]) helps with the registration against a global transformation like rotation but it is not very useful against geometric distortion since the transformation is local. Another problem with the template approach is that it assumes the underlying transformation is affine, which is generally not true for geometrically distorted images. However, in a small neighbourhood, the transformation is likely to be approximately affine [7].

One of the most accurate techniques for registering images is based on surface splines, which requires the use of a reference image in addition to the distorted one. The major step in surface spline registration is the identification of matched features between the distorted and the reference images. Goshtasby [8] gives the details of surface spline registration. The drawback with this approach is that feature points typically do not span the entire image and in regions which lack features, the registration results may be inaccurate. Measurement noise (inexact locations of features) will also introduce inaccuracy into the results.

We can regard the distortion as motion between the distorted and the reference image. Our proposed registration algorithm is based on motion estimation in the CWT domain [9], and works in a hierarchical manner. The main steps are outlined below:

1. Compute the CWT of both the distorted and the watermarked reference images. Then for each resolution, starting from the coarsest one (level 4 in our case), do steps 2 to 5.

2. Compute an initial estimate of the motion field, based on the phase difference between the CWT coefficients of the distorted and the reference image, and the associated confidence ellipse [9], which is an indication of the amount of measurement noise present.

3. Detect and reject motion vectors which look "wrong". Since the distortion is small and the distorted image resembles the original, the motion field must be smooth. Wrong motion vectors therefore manifest themselves as discontinuities in the motion field. In the current implementation, we compute the median of the motion vectors at each location based on a local neighbourhood. Any vector which differs too much from the median are flagged as "wrong".

4. Interpolate the "holes" left by the rejected motion vectors using radial basis functions [10].

(a) (b) (c) (d) (e)

Figure 3: Registering "Lena". (a) Reference image. (b) Distorted image. (c) Reconstructed image. (d) Difference between (a) and (b), rms error is 32.9. (e) Difference between (a) and (c), rms error (excluding boundary regions) is reduced to 4.7.

5. Apply a relaxation algorithm on the interpolated motion field based on the confidence of each motion vector. This allows the more confident motion vectors to influence the less confident ones.

6. Repeat steps 2-5 for the next finer resolution, using the current corrected motion field as the initial input. In practice the algorithm stops at level 2 since the CWT coefficients at level 1 are too noisy and the resulting motion field is not reliable.

7. Reconstruct the distorted image with motion compensation using overlapping blocks [11]. Cubic splines are used to obtain image intensities with sub-pixel shifts.

Although we require a reference image for the registration, this does not pose a serious problem. As long as we use an undistorted copy of the watermarked image as the reference, we do not need to reveal the original, unwatermarked host image. Fig. 3 shows an example of image registration. It is not possible to reconstruct the boundary regions since they are truncated by StirMark. Nevertheless, the registration has almost completely inverted the distortion. We found that the watermark can be successfully extracted from the reconstructed image.

5 CONCLUSIONS

In this paper, a watermarking algorithm in the complex wavelet domain is proposed. We then model watermarking as a communication process and it is shown that the CWT domain has relatively higher capacity than both the spatial and the DWT domains. The Gabor-like nature of CWT filters allows us to adapt the watermark strength to the local image activity better than the DWT filters, which are real and separable. On the other hand, the approximate shift-invariant property of the CWT coefficients allows us to infer pixel shifts, which can be used to invert geometric distortion. A registration algorithm based on this idea is described in this paper. The CWT also lacks blocking artefacts, which are present in block-based transforms like the DCT. All these properties suggest the CWT would be a potentially good domain for watermarking. Current work is directed towards investigating the effect of wavelet-based denoising on watermarked images, and the possibility of blind image registration.

References

[1] N. G. Kingsbury, "Shift invariant properties of the Dual-Tree Complex Wavelet Transform", *Proc. ICASSP 99*, Phoenix, AZ, 16-19, Mar 1999.

[2] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles", *Vision Research*, 20:847-856, 1980.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[4] M. Ramkumar and A. N. Akansu, "Information theoretic bounds for data hiding in compressed images", *Proc. of IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, Dec 1998.

[5] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn, "Attacks on copyright marking systems", *Proceedings of the Second International Workshop on Information Hiding*, Portland, Oregon, Apr 1998 vol. 1525 of Lecture Notes in Computer Science, pp. 218-238, Springer-Verlag, Berlin, Germany.

[6] S. Pereira, J. J. K. Ó Ruanaidh, F. Deguillaume, G. Csurka and T. Pun, "Template based recovery of Fourier-based watermarks using log-polar and log-log maps", *Proc. IEEE International Conference on Multimedia Computing and Systems*, pp 870-874, Florence, Italy, 7-11 Jun 1999.

[7] F. Hartung, J. K. Su and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counter-Attacks", in P. W. Wong and E. J. Delp Ed., *Proc. of SPIE vol. 3657, Security and Watermarking of Multimedia Contents, Electronic Imaging*, San Jose, CA Jan 1999.

[8] A. Goshtasby, "Registration of images with geometric distortion", *IEEE Trans. on Geoscience and Remote Sensing*, 26(1):60-64, Jan 1988.

[9] J. F. A. Magarey and N. G. Kingsbury, "Motion estimation using a complex-valued wavelet transform", *IEEE Trans. on Signal Processing*, 46(4), Apr 1998.

[10] J. C. Carr, W. R. Fright and R. K. Beatson, "Surface Interpolation with Radial Basis Functions for Medical Imaging", *IEEE Trans. on Medical Imaging*, 16(1):96-107, Feb 1997.

[11] R. W. Young and N. G. Kingsbury, "Frequency domain motion estimation using a complex lapped transform", *IEEE Trans. on Image Processing*, 2(1), Jan 1993.