Chapter

7

"We will either find a way, or make one."

Hannibal, 247-182 BC

"It is not the answer that enlightens, but the question"

Eugen Ionescu, 1909-1994

Adding Robustness to Geometrical Attacks

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together.

For uncompressed video, the geometric attacks tend to be less severe compared to those for image watermarking, mostly due to visibility considerations and due to the TV studio particularities [Cheveau et al, 2000]. On the other hand, the recovery problem is compounded for video since it must be carried out blind due to the difficulty of storing the original.

In this case, in order to re-establish synchronisation one could use sliding window cross-correlators, as described in Chapter 3 and Chapter 5. Unfortunately the search space grows very quickly, making it difficult to recover the watermark in a reasonable time (section 3.2.2 and section 5.4.2). Clearly, given that (ideally) retrieval in a video context must be done in near real time, the computational problem is very significant in the presence of attacks.

This chapter provides a solution[1] to this difficult problem by employing an additional reference watermark and image registration techniques, while maintaining in the same time the high capacity and the robustness of the Wavelet based system presented in Chapter 6.

[1] The system can cope with all geometric attacks specified in the EBU recommendations [Cheveau et al, 2000], with the exception of bending/shearing attacks which are still an open problem. In the actual configuration, this problem can be solved by employing a sliding window correlator.

7.1 Methods of Combating Geometrical Attacks

Invariant transforms

One of the answers of the research community to this difficult problem was to use transforms invariant to these attacks.

For example, embedding a watermark in the magnitude of the DFT coefficients yields a shift invariant system, therefore eliminating the need of a 2-D sliding correlator, which normally would be used to recover this type of attack. Unfortunately, as section 2.4.3 already mentioned, due to its disadvantages, a FFT-based watermarking scheme is not the best choice for digital watermarking.

The Fourier-Mellin transform (FMT) was first used in [O'Ruanaidh et al, 1998] to achieve rotation, scaling and translation invariance for image watermarking. The FMT doesn't have a fast direct computing algorithm, but can be simulated by transforming the Cartesian coordinates into log-polar coordinates and then performing a Fourier transform of this output.

Unfortunately, marking in the FMT domain has two major drawbacks: the need to compute the inverse log-polar transform which is a lossy operation that drastically reduces system performance, and the need to maintain the FFT symmetry, which halves the watermark capacity. Due to these disadvantages the technique offers only poor results, as even the author acknowledged later on. An improved technique was later proposed in [Lin et al, 2000].

A very promising technique is suggested in [Loo et al, 2000], which uses the Complex Wavelet Transform (CWT) domain for watermarking. The dual-tree CWT transform offers shift invariance, directionally selective filters and limited redundancy [Kingsbury, 1998 and 1999]. One of the most appealing features of the CWT is that the phase of the CWT coefficients can be used to infer pixel shifts quite accurately. Therefore using a registration algorithm based on motion estimation, one could combat even the non affine geometric distortions produced by StirMark [Loo et al, 2000].

Special watermark arrangements

Another possibility is to cleverly arrange the watermark so that the search space reduces considerably. Some examples could be: circularly symmetric watermarks [Solachidis et al, 1999], tiled (cyclic) watermarks [Kalker et al, 1999-1 and 1999-2] and [op de Beeck et al, 2001], self-similar watermarks [Dittmann et al, 2000]. This approach is in general more successful than the previous method, but still has a major drawback: although the geometric

attacks are more easily handled, the capacity of such a system is relatively limited, due to the special arrangement of the watermark.

Reference watermarks

A slightly different approach is to use reference watermarks. These are separate watermarks additional to the main watermark which carries the actual watermark data. The reference watermark is in fact not carrying any information at all, being used only as a countermeasure for geometric attacks. They are sometimes called templates or patterns. In this scenario, the image/video sequence will carry two distinct watermarks, on top of each other, and which are normally embedded in the same domain in order to keep the complexity of the method low. The reference watermarks can be used to identify the parameters of the (affine) geometrical attack. Once these parameters are identified it is possible to undo the geometrical attack and then recover the main watermark.

Most of the existing schemes are using a combination between these two later methods [Kalker et al, 1999-1 and 1999-2], [op de Beeck et al, 2001], [Pereira et al, 1999]. Usually the reference is embedded in a block wise manner (the size of the template is smaller than the size of the image/frame) [Kalker et al, 1999-1 and 1999-2], [Pereira et al, 1999] in different locations of the image/frame, so the peaks obtained after the cross-correlation will be distributed on a grid. Their position on the grid depends on the attack, and therefore it is possible to estimate (identify) the parameters of the affine geometrical attack using this grid [Kalker et al, 1999-1 and 1999-2], [Wolberg et al, 2000].

In order to get shift invariance, the reference (and the main watermark as well) is sometimes embedded in the Fourier domain [Pereira et al, 1999]. One could achieve robustness to scaling and rotation attacks by using log-polar transforms which translate the scaling and rotation to spatial shifts which can then be easily recovered. Robustness to aspect ratio changes can be achieved by using log-log transforms to convert the scale changes to spatial shifts. Compared with watermarking in the Fourier–Mellin domain, this approach has the advantage that doesn't require the computation of the lossy inverse Fourier-Mellin transform. In fact this technique is very close to the well known image registration and template matching problem.

[Pereira et al, 1999] partially uses this method, but instead of following the normal template matching approach they transform it in a point-matching problem over a log-polar or log-log map which involves a limited exhaustive search. The both watermarks are embedded in

the Fourier domain and therefore the scheme inherits the disadvantages of Fourier based watermarking techniques described in section 2.4.3.

Moreover embedding two watermarks in the same domain raises two problems: the possible interference between the watermarks and the efficiency of the embedding. In order to counteract the first problem, the watermarks have to be orthogonal, and therefore the PN sequence corresponding to each watermark has to be carefully selected. The second problem arises because in this case, two watermarks have to "share" the maximum value given by the visual model for modifying a certain coefficient. There can be only a limited amount of modification allowed for a certain coefficient, in order to maintain the invisibility of the watermark(s). So there is always a problem in finding the right balance between the strength of each watermark, and therefore the energy of each watermark is lower than when embedding only a single watermark.

7.2 Symmetrical Phase-Only Matched Filtering

Another way of using the advantages of the Fourier transform is to use it for implementing fast cross-correlators (section 2.4.3). The roots of FFT cross-correlation can be found in optics, where a lens basically performs a Fourier transform. Then the concept was then quickly adopted by the image processing community.

Cross-correlation or matched filtering is particularly used for pattern (template) matching, image registration and recognition and motion estimation. The aim of matching is either to determine the presence of a template/image in a noisy scene (pattern recognition), or to determine the parameters of a geometric transformation relating two images (image registration). Such a FFT based cross-correlator can easily recover 2-D shifts, saving allot of computing time compared with the classical cross-correlators [Kuglin et al, 1975], [Horner et al, 1984], [Pech-Pacheco et al, 199x], [Chen et al, 1994] and [Hill et al, 1999].

Let's consider two images $f_1(x,y)$ and $f_2(x,y)$ which differ only by a displacement (x_0,y_0). This can be expressed as $f_2(x,y)=f_1(x-x_0,y-y_0)$. The Fourier transforms of these images are $F_1(u,v)$ and respectively $F_2(u,v)$. It is well known that the output of a classical matched filter is primarily dependent on the energy of the image rather than its own spatial structures. This is why the matched filter provides a relatively poor discrimination between objects of different shapes but similar size or energy content. Furthermore, the filter

output is proportional with the image auto-correlation and the shape of the filter output around its maximum (x_0, y_0) is broad. Therefore locating this maximum in the presence of noise is relatively difficult.

The answer to this problem is the *Phase Only Matched Filter* (POMF), who's transfer function is equal with the spectral phase of the image [Chen et al, 1994]

$$H_{POMF}(u,v) = Phase\left(F_1^*(u,v)\right) = \exp\left(-j\phi_1(u,v)\right) \tag{7.1}$$

where $j^2 = -1$ and $\phi_1(u,v)$ is the spectral phase of the image $f_1(x,y)$. Since the spectral phase preserves the location of the objects but is insensitive to the image energy, the application of a POMF to a pair of identical images under the constraint of translation will result in a much sharper peak than in the case of the classical matched filter. This is very well illustrated in [Horner et al, 1984]. This explains the popularity of the POMF in the image processing community: the detection and location of the maximum is easier and therefore the POMF allows much better discrimination between different objects.

A further improvement of the POMF can be achieved by extracting and correlating the phases of both input images. In this case the filter can be defined as

$$H_{SPOMF}(u,v) = \frac{F_2(u,v)}{\left|F_2(u,v)\right|}\frac{F_1^*(u,v)}{\left|F_1^*(u,v)\right|} = \exp\left[j\left(\phi_2(u,v) - \phi_1(u,v)\right)\right] \tag{7.2}$$

where the $\phi_1(u,v)$ and $\phi_2(u,v)$ are the spectral phases of the image $f_1(x,y)$ and respectively $f_2(x,y)$. In the absence of noise, this reduces to

$$H_{SPOMF}(u,v) = \exp\left[-j2\pi\left(ux_0 + vy_0\right)\right] \tag{7.3}$$

The output of the filter is given by the inverse Fourier transform of $H_{SPOMF}(u,v)$ which is in fact a Dirac δ function, centred at the location (x_0, y_0). This filter is called a *Symmetrical Phase Only Matched Filter* (SPOMF) and can be seen as a two step process: first the extraction of the phases of the input images and then the phase only matched filtering. An assessment of the comparative performance of different matched filters can be found in [Chen et al, 1994].

7.3 Image Registration and Watermarking

By using such a Phase Only Matched Filter, the shift problem is easily solved. Unfortunately, this technique can be applied only to frame shifts.

On the other hand, it is well known in image processing that transformation of Cartesian coordinates into log-polar coordinates converts scale and rotation to spatial shifts [Casasent et al, 1976-1 and 1976-2], [Chen et al, 1994], [Reddy et al, 1996], [Cheng et al, 1998], [Wolberg, et al, 2000]. As stated before, performing the Fourier transform of this output is in fact equivalent with performing the Fourier-Mellin transform (FMT) of the original image, and the advantage of the Fourier-Mellin transform is its scaling and rotation invariance.

The idea developed by Casasent and Psaltis [Casasent et al, 1976-1 and 1976-2] was quickly adopted for image processing in the context of *image registration*. This involves two images; the original and an attacked copy, and the objective of an image registration module is to determine the parameters of the geometric distortion. The attack can then be inverted to give geometric alignment of two images.

When registering two images, typically the noise is relatively small, and so the correlator usually performs very well. However, the problem is more difficult for video watermarking since the original video frame is not available and "blind" recovery is necessary. In this case one can use spread spectrum watermarking to compensate for the unavailable original. Assuming that this (reference) watermark is a one bit watermark having the size of the (original) image, then is possible to see a correspondence with the classical image registration problem. The "original image" corresponds to the PN sequence used to embed the watermark, and the "attacked image" corresponds to the unsynchronised (attacked) watermarked image. This "attacked image", which is in fact nothing else than the watermarked image (possibly attacked) is composed by two components: the first component is the reference watermark and the second component is the image/video itself. The reference watermark represents the signal and the image/video itself can be regarded as (additive) noise. In the case of video watermarking/registration, the signal to noise ratio is therefore very low relative to that for typical image registration. By making a parallel with the watermarking, this technique can be called *"blind" registration*.

7.4 Log-Polar and Log-Log Mapping

As already mentioned, the highly desired geometric invariance can be achieved by using the FMT to convert rotation and scale to spatial shifts, which are then easily recovered with a SPOMF cross-correlator. The FMT itself does not have a fast algorithm, but as [Casasent et al, 1976-1 and 1976-2] shows, a simple variable change $x = \exp\xi$ on the input $f(x)$, followed

by a Fourier transform will yield the FMT. For the bi-dimensional case, this is equivalent to a log-log transform of the input, and permits recovery from arbitrary scale changes (aspect ratio changes).

Consider the case of arbitrary scaling with a factor *(a, b)*. If frame f_2 is the scaled replica of frame f_1 with a factor *(a, b)* then

$$f_2(x,y) = f_1(ax,by) \tag{7.4}$$

and its Fourier pair is

$$F_2(u,v) = \frac{1}{|ab|} F_1(u/a, v/b) \tag{7.5}$$

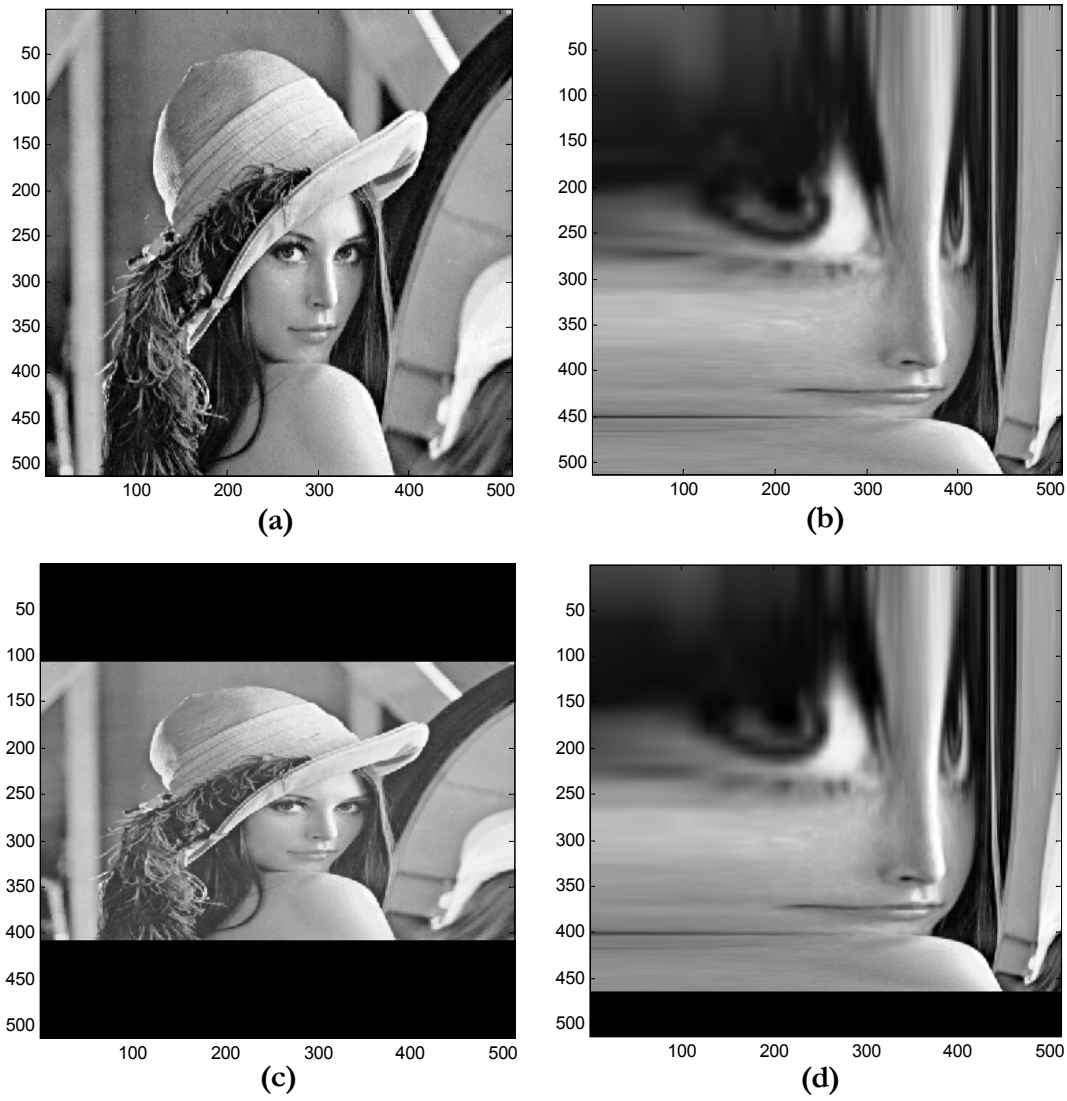


Figure 7-1 The log-log transformation and its results: **(a)** the original Lena image; **(b)** the log-log transformation of (a); **(c)** aspect ratio change attack and **(d)** the log-log transform of (c), the vertical scaling was transformed into a vertical shift in log-log coordinates.

If the Cartesian coordinates are transformed into *log-log* coordinates, and the multiplicative factor ignored, then

$$F_2\left(\log u, \log v\right) = F_1\left(\log u - \log a, \log v - \log b\right) \tag{7.6}$$

By using a SPOMF, the shifts $\left(-\log a, -\log b\right)$ can be found, and so the scale factors *(a, b)* are determined. An example illustrating this mechanism is provided in **Figure 7-1**. The scaling (aspect ratio change) is converted into spatial shifts. This can be easily observed by comparing the set of images **(a)** and **(b)** with their corresponding pair **(c)** and **(d)**.

Now consider f_2 to be a rotated replica of f_1, with angle θ_0

$$f_2\left(x, y\right) = f_1\left(x\cos\theta_0 + y\sin\theta_0, -x\sin\theta_0 + y\cos\theta_0\right) \tag{7.7}$$

$$F_2\left(u, v\right) = F_1\left(u\cos\theta_0 + v\sin\theta_0, -u\sin\theta_0 + v\cos\theta_0\right) \tag{7.8}$$

In order to convert this rotation into a shift, as in the previous case, the Cartesian coordinates are transformed into polar coordinates using

$$\begin{aligned} \rho &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}\left(y / x\right) \end{aligned} \tag{7.9}$$

This leads to

$$F_2\left(\rho, \theta\right) = F_1\left(\rho, \theta - \theta_0\right) \tag{7.10}$$

and the rotation can be easily recovered from the frequency domain shift.

Finally, when f_2 is both scaled with a factor *a* and rotated with angle θ_0, then

$$f_2\left(x, y\right) = f_1\left(a\left(x\cos\theta_0 + y\sin\theta_0\right), a\left(-x\sin\theta_0 + y\cos\theta_0\right)\right) \tag{7.11}$$

$$F_2\left(u, v\right) = \frac{1}{a^2} F_1\left(\left(u\cos\theta_0 + v\sin\theta_0\right)/a, \left(-u\sin\theta_0 + v\cos\theta_0\right)/a\right) \tag{7.12}$$

In order to convert both scaling and rotation to shifts, it is necessary to convert the Cartesian coordinates into log-polar coordinates, using the following equations

$$\begin{aligned} x &= e^{\log\rho}\cos\theta \\ y &= e^{\log\rho}\sin\theta \end{aligned} \tag{7.13}$$

The result is

$$F_2\left(\log\rho, \theta\right) = F_1\left(\log\rho - \log a, \theta - \theta_0\right) \tag{7.14}$$

where the scale and rotation factors can be retrieved by SPOMF correlation. **Figure 7-2** shows the log-polar mapping and its effects. In this case the rotation is converted into a spatial shift. Again, this can be easily observed by comparing the set of images **(a)** and **(b)** with their corresponding pair **(c)** and respectively **(d)**.

Using the shift invariance property of the Fourier transform, it is possible to recover even combined attacks like shift combined with rotation and scaling or shift combined with aspect ratio changes. But since the FMT is not shift invariant, it is necessary to apply the Fourier magnitude of the frame (rather than the frame itself) to the input of the log-polar conversion module. The Fourier magnitude is shift invariant and so the rotation and scaling parameters can be found even in the presence of shift. After undoing rotation and scaling, the shift is then recovered by performing a simple SPOMF correlation. This technique works well in the particular case of image-image registration [Reddy et al, 1996], since the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, for video watermarking

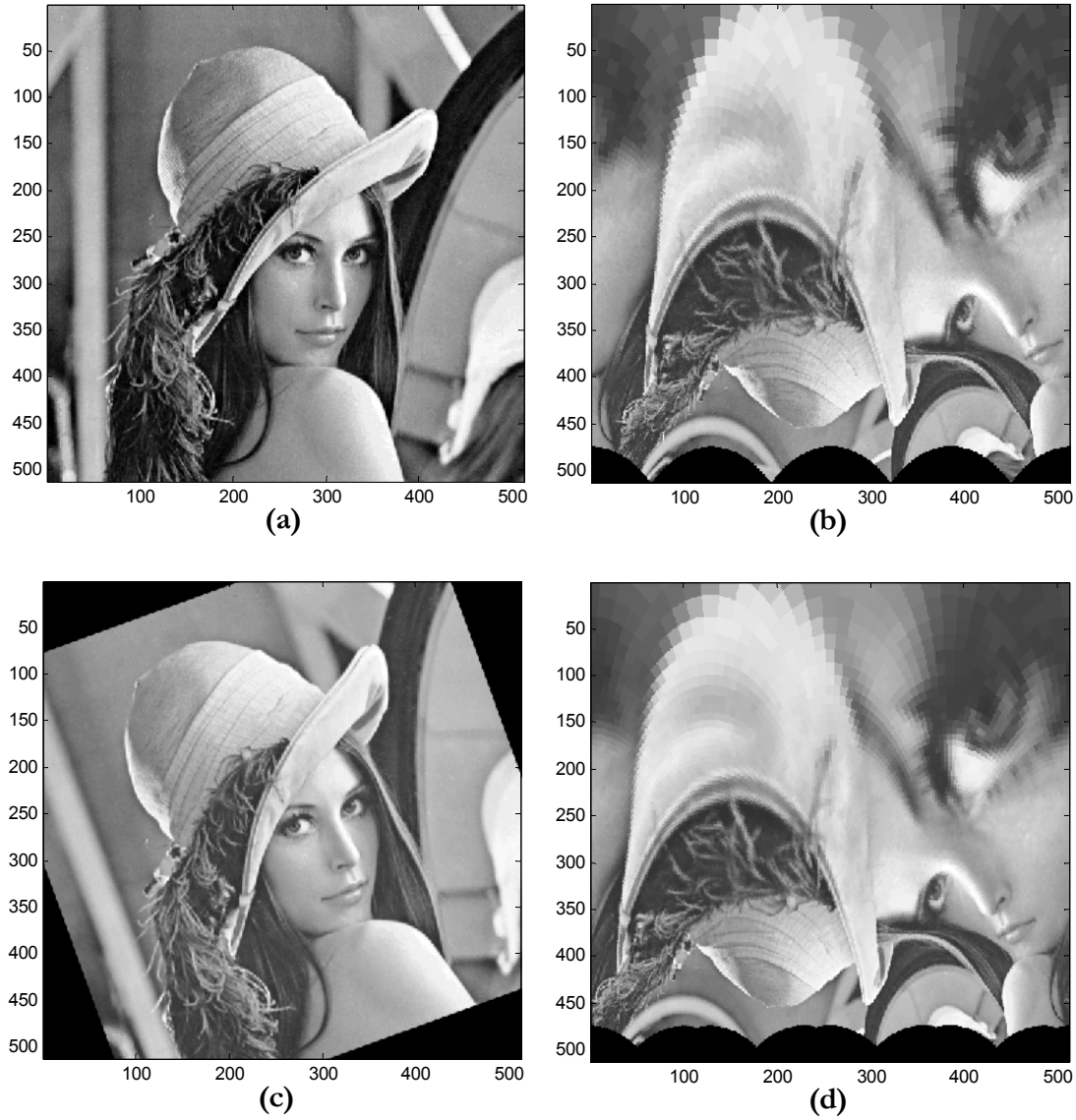


Figure 7-2 The log-polar transformation and its effects: **(a)** the original Lena image; **(b)** the log-polar representation of (a); **(c)** rotation attack and **(d)** the log-polar representation of (c), where the rotation was converted to a spatial (right) shift in the log-polar representation.

(blind registration), the loss can make cross-correlation unreliable, and this approach cannot be used for retrieval under combined attack.

A log-polar map permits recovery over a wide range of scale changes, rotation, or even combined scale-rotation attack. If a log-log map is used, then it is possible to recover arbitrary aspect ratio changes. The shifts alone are easily recovered using a SPOMF module. However, shift recovery from a combined attack (e.g. shift combined with scaling and rotation, or shift plus aspect-ratio change) requires a comprehensive search for all of the possible shifts [Wolberg et al, 2000], and is computationally intensive.

7.5 A High Capacity, Robust System

Using the method described in section 7.3, two different watermarks must be embedded. As section 7.1 mentioned, having two watermarks embedded in the same domain has some inconvenient. To overcome this problem, each watermark is embedded in a different domain. The first one, is a 1-bit watermark used exclusively for geometric reference, and for simplicity is embedded in the spatial domain. The second, multi-bit watermark is used for the data payload, and is embedded in the DWT domain. The advantage of using such an approach is obvious: the watermarks are orthogonal since they are embedded in different domains, so the cross-talk between them is minimal. Each watermark is embedded with the full strength dictated by its own visual model and overall the resulting system is relatively simple. Moreover the system combines the advantages offered by the blind video registration with the clear advantages offered by the Wavelet based embedding (Chapter 6). By combining these two

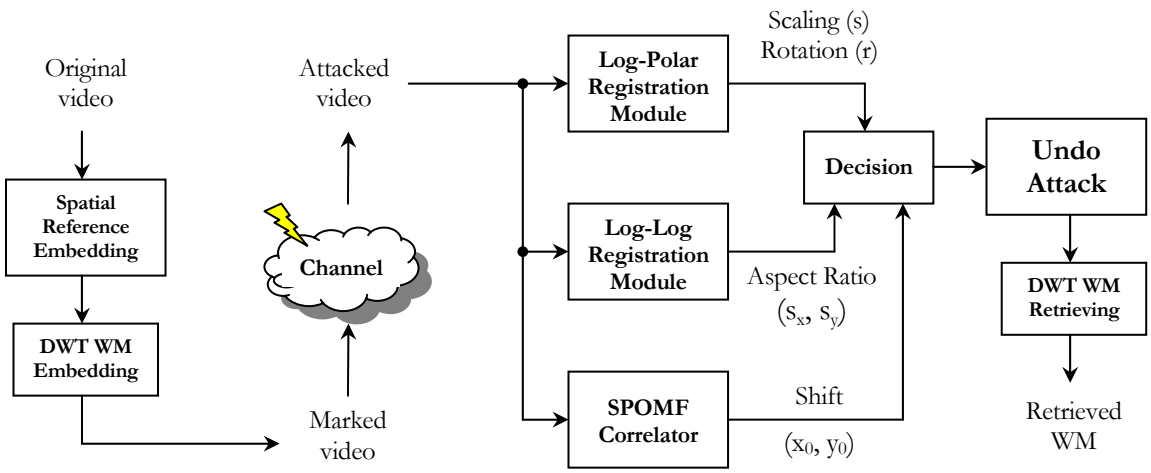


Figure 7-3 Block schematic of the geometric invariant video watermarking system.

techniques not only that the system ensures a good robustness against a wide range of attacks, but this can be achieved while maintaining the high capacity of the scheme intact.

The proposed system is presented in **Figure 7-3**. The first step is to embed the spatial reference watermark. Although the spatial domain is not the best place to cast a watermark, the reference is quite robust because effectively only one data bit is embedded in the entire frame. Since the capacity is not an issue in this case, the spatial domain can be successfully used and by doing so, one can exploit its biggest advantage: simplicity. This is embedded using the classical spread spectrum approach, according to a simple visual model that inserts a stronger watermark in those regions where it is less easily observed (at edges and in high texture regions). The same reference watermark is embedded in all the frames in order to increase the SNR at the correlator input via frame averaging. As a result, the registration takes place only once, and not for each separate frame. This is possible because attacks must be identical for each frame in order to avoid temporal artefacts.

The second step consists in embedding the high capacity watermark, according to the scheme described in Chapter 6.

At the retrieval side, the system employs three registration modules: the log-polar registration module which takes care of the rotation and scaling attacks, even when these two attacks are combined together; the log-log module which handles the aspect ratio change attacks and finally, a simple SPOMF correlator which handles attacks like spatial shifts and cropping, together with non-geometric attacks like compression or even a combination of these attacks.

The decision block determines if the reference watermark is present (to within a desired false detection probability), and if present it automatically determines the attack parameters. Once the parameters of the attack are identified, the attack is reverted back and the main watermark is then recovered.

Another advantage of using two watermarks is now apparent: if the reference cannot be found, one can assume that either the video is not marked, or that the mark is destroyed, and recovery of the main watermark payload can be abandoned (saving computation time).

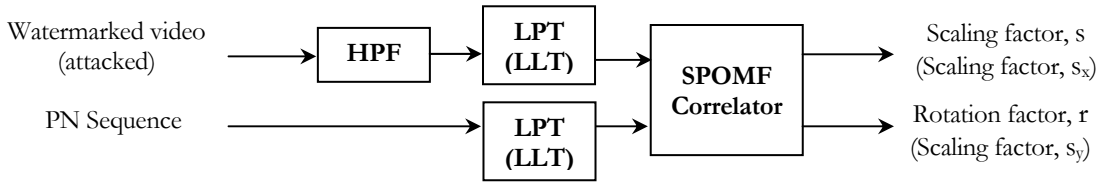


Figure 7- 4 The log-polar / log-log registration module.

Figure 7- 4 shows implementation detail of the log-polar transform (LPT) and the log-log transform (LLT) registration module. The role of the Laplacian high pass filter (HPF) is to remove low and medium frequency video components (which represent noise) and pass only the high frequency components, which contain the spread spectrum, noise-like watermark. This significantly improves the correlator performance.

7.6 Performance of the System

The registration module provides invariance to frame shift, rotation, scaling, rotation combined with scaling, and aspect ratio change. The system can also handle a range of other attacks, such as cropping, shift combined with cropping, compression, shift combined with

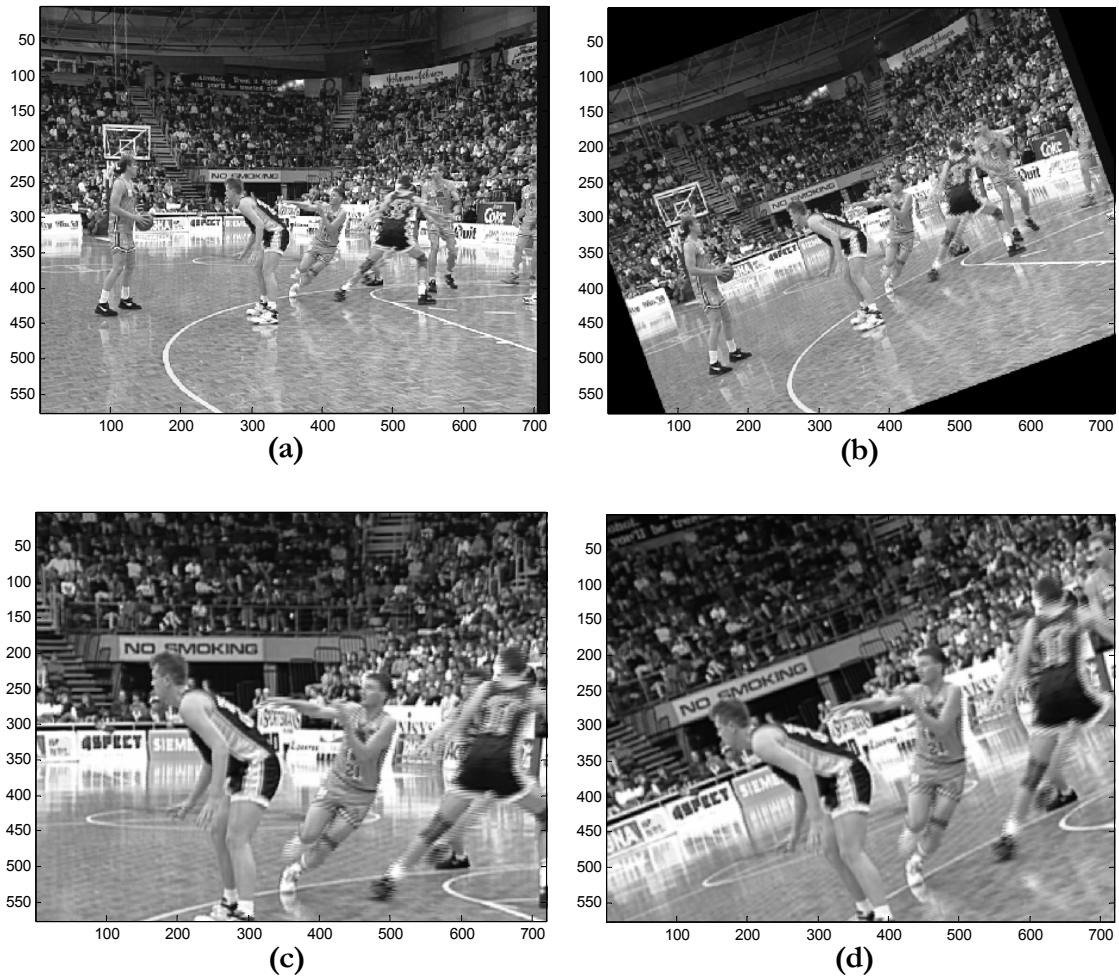


Figure 7-5 The effects of different geometrical attacks: **(a)** original basketball sequence, **(b)** 20^0 rotation, **(c)** 100% scaling and **(d)** 20^0 rotation combined with 100% scaling.

compression and shift combined with cropping and compression. The performance of the system for these attacks is presented below.

7.6.1 Scaling and Rotation Attack

These attacks are illustrated in **Figure 7-5** for "basketball" video sequence both as a discrete attack and as a combined attack. **Figure 7-7(a)** shows the performance of the system for different degrees of rotation, when n frames are averaged in order to improve the robustness of the system. Since the minimum watermarking segment is 25 frames, then $n \le 25$. Compared with the $n = 1$ case, the cross-correlation peak for $n = 25$ is about four times larger. Similar results are presented in **Figure 7-7(b)** for scaling. **Figure 7-8(a)** and **Figure 7-8(b)** illustrates the system performance for $n = 25$ and different degrees of rotation

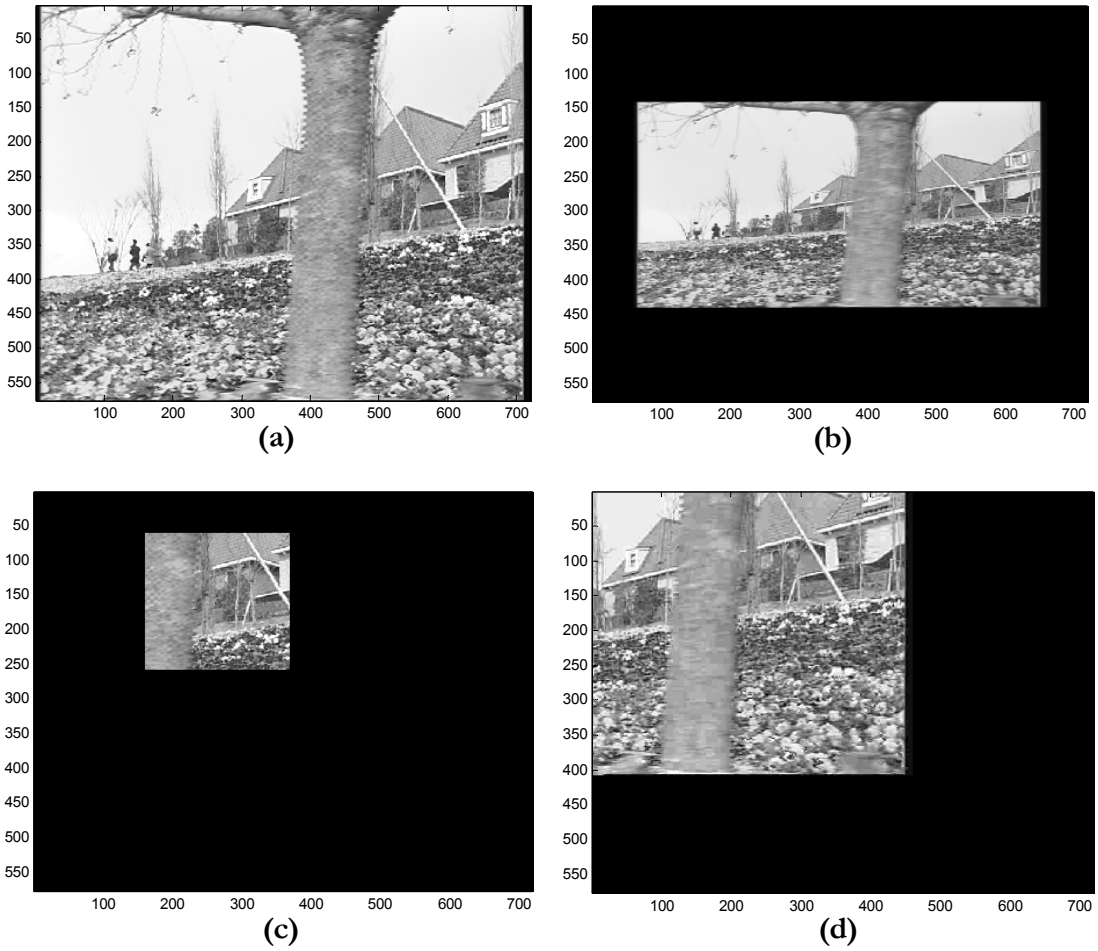


Figure 7-6 The effects of different attacks: **(a)** original flower sequence, **(b)** arbitrary scaling, from [576x720] to [300x600], **(c)** cropping [400,200,208,196] combined with shift [140,240], **(d)** shift [170,260] combined with 3Mbps MPEG2 compression.

and respectively scaling for three typical video sequences. As expected, the worse case is the "flower" sequence. Finally, the combined attack (rotation plus scaling), is shown in **Figure 7-10** for the "basketball" sequence ($n = 25$).

As **Figure 7-8(a)** and **(b)** suggests, the system is invariant to any amount of rotation smaller than 70^0, and it can handle any degree of scaling up to 180%. The system is also capable to handle scaling up to –50% (i.e. smaller frames). Therefore the EBU recommendations [Cheveau et al, 2000] are exceeded for both rotation and scaling.

When rotation is combined with scaling, up to 120% scaling and up to 20^0 rotation can

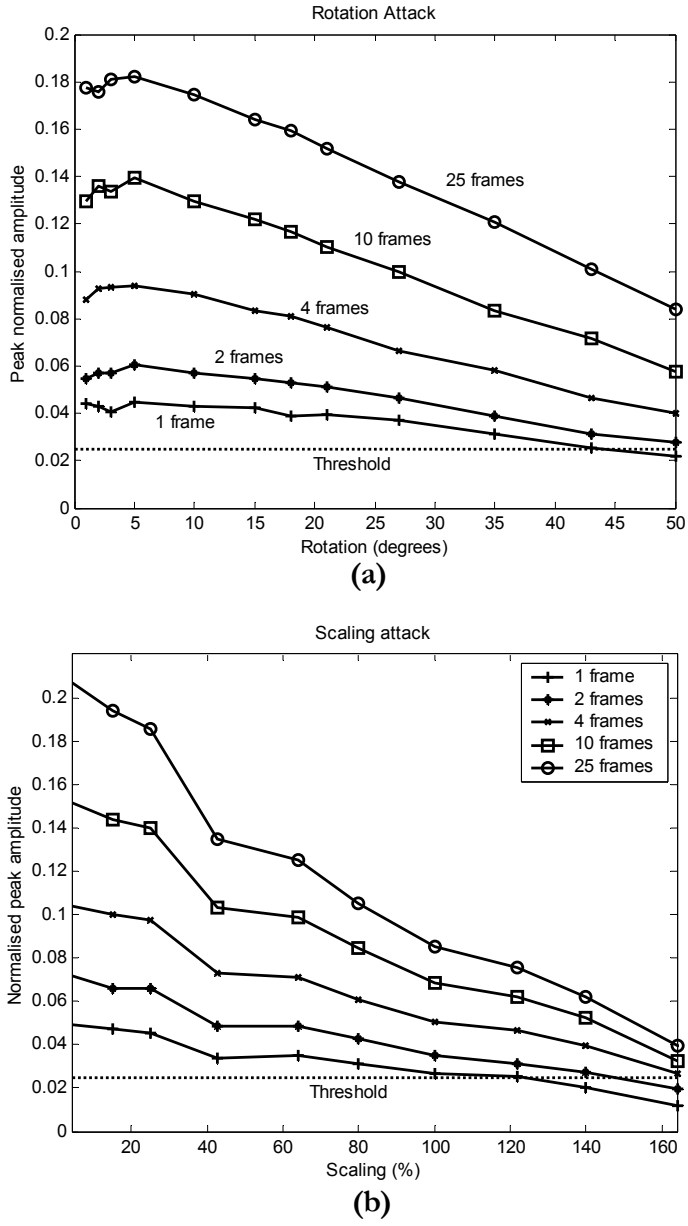
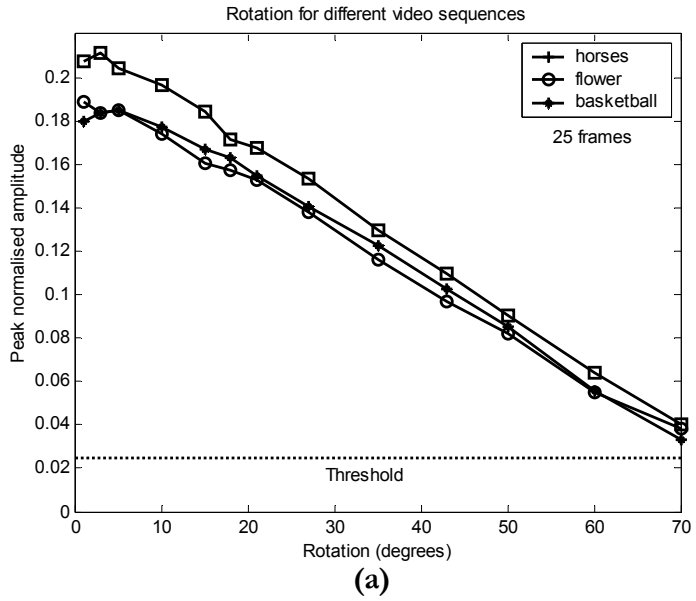
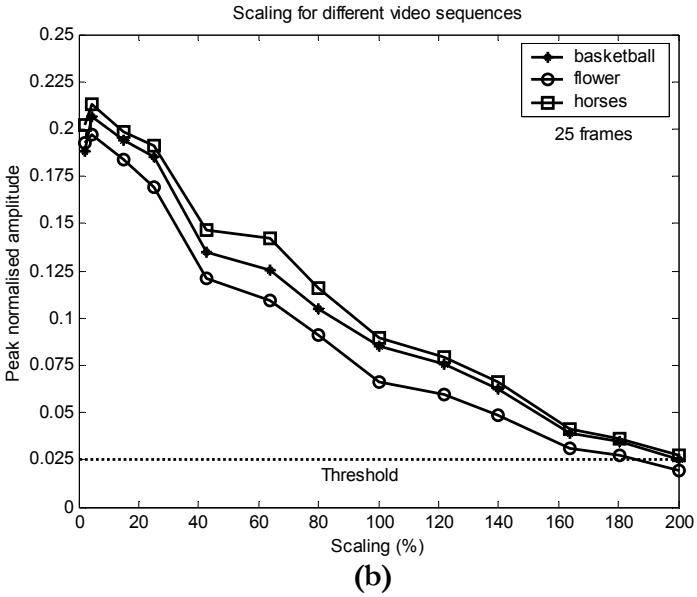


Figure 7-7 Performance of the system when averaging frames for: **(a)** rotation and **(b)** scaling.

be tolerated, even for the "flower" sequence. All simulations assume a bilinear interpolation in the log-polar module. The experiments show that bilinear interpolation leads to a substantial performance increase (almost double) compared with a simple nearest neighbour interpolation. These two cases are illustrated in **Figure 7-9**. It is obvious from **Figure 7-9(a)** that the nearest neighbour interpolation leads to a much coarser result compared with the bilinear interpolation (**Figure 7-9(b)**).



(a)



(b)

Figure 7-8 Peak normalised amplitude for different video sequences under: **(a)** rotation attack and **(b)** scaling attack, when averaging 25 frames together.

7.6.2 Aspect Ratio Changing Attack

This attack is illustrated in **Figure 7-6(b)**. The difference between scaling and aspect ratio changing is that in the latter case two different scaling factors are present: one for horizontal scaling and a different one for vertical scaling. Using a log-log registration module, the system is invariant to arbitrary aspect ratio changes in the range –100% to 200% on both axes. The experiments show that this attack is much easier to handle than rotation combined with scaling. Again, bilinear interpolation performs much better compared with nearest neighbour interpolation (slightly more than double in this case).

7.6.3 Shifting and Cropping Attack

Shifting and cropping alone, or even the combined attack are not a posing a significant threat for the system, provided that at least 30% of the frame is still intact. These attacks are handled by the SPOMF registration module.

An illustration of a shifting and cropping attack is provided in **Figure 7-6(c)**. Even for this severe case, unlikely to happen in practice, the watermark can be recovered relatively easy. Again, in this case the EBU recommendations are largely exceeded.

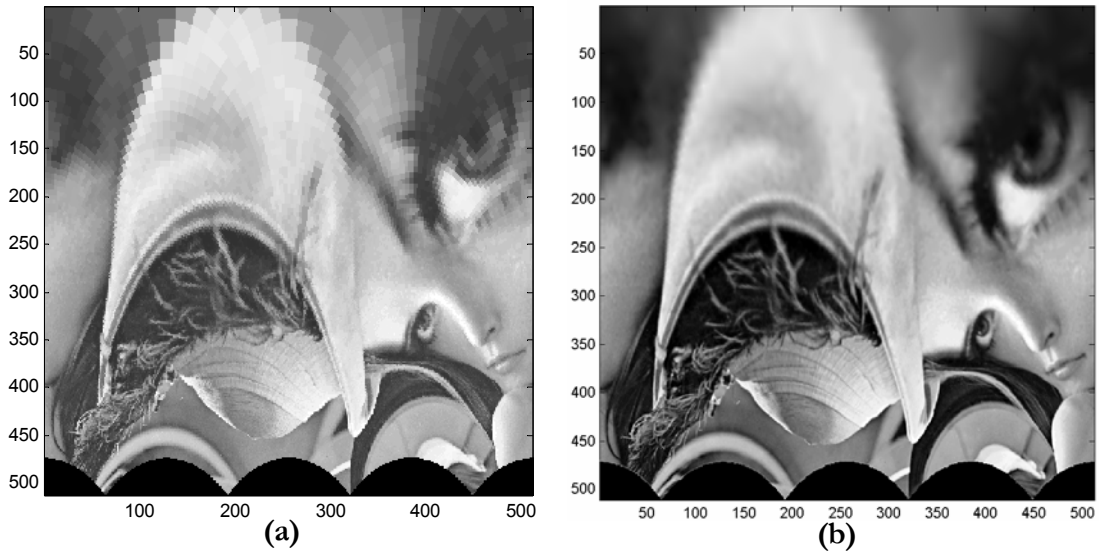


Figure 7-9 The log-polar map of image "Lena" for: **(a)** nearest neighbour interpolation and **(b)** bilinear interpolation.

7.6.4 Compression Attack

The system can cope very well with MPEG2 compression. The results under MPEG2 compression are presented in **Figure 7-11(a)**. One can see that the system can cope even with MPEG2 compression at 2Mbps, for all the test sequences involved.

Combined attacks like MPEG2 compression plus arbitrary frame shifts can be handled as long as the MPEG2 compression is at least 3-4Mbps (**Figure 7-11(b)**).

7.7 The False Detection Probability

A threshold value of 0.025 can be observed in each figure (**Figure 7-7**, **Figure 7-8**, **Figure 7-10**, **Figure 7-11**). This guarantees a false detection probability better than 10^{-8} when the correlation peak exceeds the threshold.

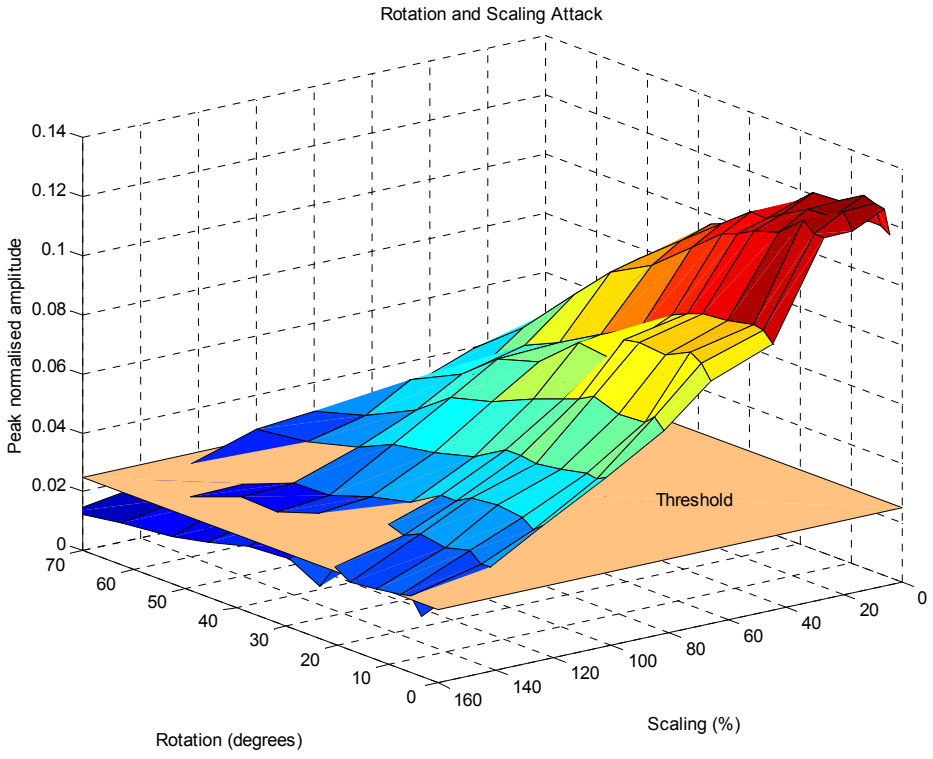
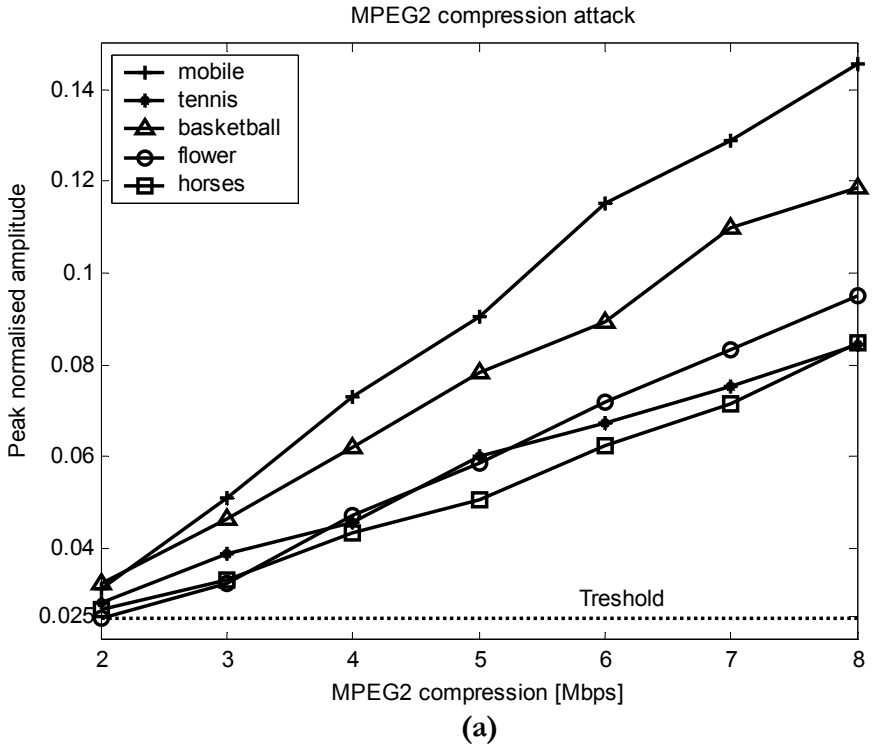
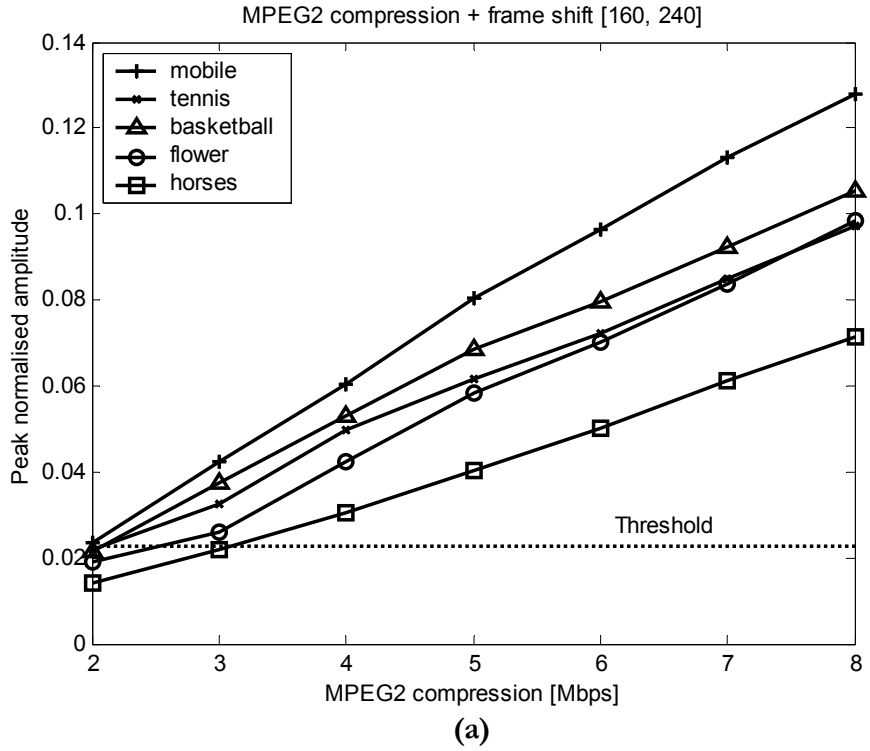


Figure 7-10 Performance of the system for rotation combined with scaling (25 frames, basketball video sequence).

The value was experimentally derived for a set of 3 test sequences and a wide range of scaling and rotation attacks: the pdf (probability distribution function) of the peaks was



(a)



(a)

Figure 7-11 Compression attack: **(a)** MPEG2 compression, for different video sequences and **(b)** MPEG2 compression combined with spatial shift [160, 240].

computed for each case and the worse-case scenario determined. The resulting pdf is not Gaussian due to the large number of very small peak values, but by fitting a zero-mean Gaussian distribution with the same standard deviation as the experimentally determined pdf, the resulting Gaussian distribution can be used to determine the optimum threshold for a given false error probability. The Gaussian distribution fits very well the worse-case scenario pdf in the zone of interest (at the extremities), and is actually chosen to be quite pessimistic.

Several hypotheses were examined: when the sequence was marked with the correct watermark (all the "parasite" peaks were taken into account), when the sequence was not marked and when the sequence was marked with a wrong reference watermark. These cases were examined for different attacks (rotation alone, scaling alone and combined attacks) and a wide range of strengths of the attacks, and finally for 3 different video sequences. The results (**Figure 7-12**) suggest that the worse-case scenario is when the sequence is marked with the correct mark, and show that the 0.025 threshold is appropriate for a false detection probability of 10^{-8}.

Again, this threshold was chosen to be rather pessimistic, and was determined for rotation and scaling attacks (and combined). Experiments show that for other attacks (aspect ratio change, shifting, cropping and compression) this threshold is even more pessimistic, because in these cases the "parasite" noise is lower. A more appropriate value in these cases is

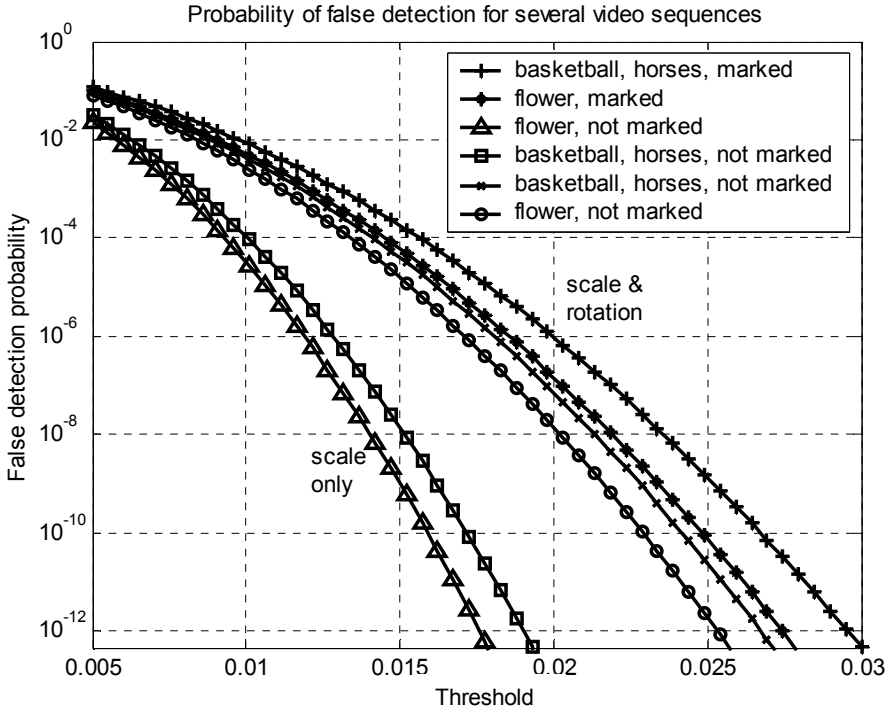


Figure 7-12 Threshold selection for a desired probability of false detection.

somewhere between 0.015-0.020.

7.8 Conclusions

Robustness to geometric attack is one of the most important requirements for a watermarking system. To satisfy this requirement an efficient approach based on the Fourier-Mellin transform and log-polar and log-log representations of the video frames has been developed. This is combined with the advantages of the DWT, HVS-based marking, and turbo coding to produce a very robust, high capacity video watermarking system.

With turbo coding, capacity can be as high as 1500 bits/frame (37.5 Kbps) even under severe cropping, and the system is invariant to a wide range of geometric attacks, such as scaling, rotation, aspect-ratio change, shifting, cropping and compression. It can also handle combined attacks, such as scaling/rotation, cropping/shifting, cropping/compression, shifting/compression, and cropping/shifting/compression. The search space is considerably reduced by using fast SPOMF-based cross-correlation.

For a false detection probability of 10^{-8}, the proposed system is invariant to scaling in the range –50% to 180%, invariant to rotation up to 70^0, and invariant to arbitrary aspect ratio changes in the range –100% to 200% on both axes. Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping. In these respects it exceeds the EBU recommendations for video watermarking. Also, very low quality JPEG compression (10%) can be handled even when combined with shifting and cropping, although this is not directly applicable to high quality uncompressed video. Capacity reduces to about 100 bits/frame (2500 bps) under 30% JPEG attack.

To achieve these results, the proposed system combines the advantages of geometric invariance of the FM transform, fast SPOMF processing, DWT and HVS-based marking, and turbo code protection.