Effects of Reversible Watermarking on Iris Recognition Performance

Andrew Lock, Alastair Allen

Abstract—Fragile watermarking has been proposed as a means of adding additional security or functionality to biometric systems, particularly for authentication and tamper detection. In this paper we describe an experimental study on the effect of watermarking iris images with a particular class of fragile algorithm, reversible algorithms, and the ability to correctly perform iris recognition. We investigate two scenarios, matching watermarked images to unmodified images, and matching watermarked images to watermarked images. We show that different watermarking schemes give very different results for a given capacity, highlighting the importance of investigation. At high embedding rates most algorithms cause significant reduction in recognition performance. However, in many cases, for low embedding rates, recognition accuracy is improved by the watermarking process.

Keywords—Biometrics, iris recognition, reversible watermarking.

I. INTRODUCTION

BIOMETRICS are proving an increasingly popular alternative to traditional authentication mediums as they resolve many issues associated with token or password-based systems. However, many works identify threats to biometric systems such as spoofing, mimicry and man in the middle attacks [1], [2]. Watermarking has been suggested as one possible means to reduce the security vulnerabilities of biometrics, and in some cases to enhance their functionality [3].

All watermarking methods alter the host image to some extent, though the perceptibility of this modification varies with the watermarking scheme and the capacity – the quantity of data embedded. The acceptability of the distortion depends on the particular application. The effect of watermarking methods on biometrics, in particular iris detection, has been investigated for robust techniques [4], [5]. However, the effect of fragile algorithms has not been studied.

In this paper, we investigate the effect fragile watermarking algorithms, specifically reversible or lossless algorithms, have on iris recognition accuracy. We aim to address whether particular watermarking techniques are more suited to this application than others, and the impact that capacity has on iris recognition performance.

Section II provides an overview of watermarking in the context of biometric systems. Section III describes experiments in which we test the impact of seven watermarking algorithms on iris recognition accuracy over a range of embedding capacities. Section IV discusses the results and Section V concludes the paper.

A. Lock and A. Allen are with the Department of Engineering, University of Aberdeen, Aberdeen, UK (phone: +44 1224 272501, e-mail: r04a99@abdn.ac.uk).

II. WATERMARKING AND BIOMETRICS

A. Watermarking

Watermarking is a technique of hiding data within a host image by modifying the image such that the data can be extracted later. By definition, watermarking modifies the cover work, introducing distortion. The level to which this distortion is permissible depends on the particular application, but is often required to be imperceptible [6]. If modifications of the watermarked image prevent recovery of the embedded data, the watermark is termed fragile, otherwise it is robust.

In some fields, for example in the medical or military domain, even slight distortions of the original image are not permissible. Consequently many watermarking algorithms have been proposed that allow recovery of the original image during the extraction process. These are termed reversible or lossless algorithms. While these have the additional capabilities of allowing recovery of the original image, they retain all other characteristics of normal watermarking schemes. In this paper, we focus on this area of fragile reversible watermarking algorithms. Reversibility is not essential for most biometric applications; however its presence does not affect the conclusions that may be drawn from studying the effect of fragile watermarking on iris recognition.

B. Watermarking in Biometric Systems

In [3] multiple scenarios are suggested where watermarking can improve the security of biometrics. The first is purely steganographic, in which biometric data is embedded within an arbitrary host image, with the aim of concealing the embedded data from a potential attacker. This technique is suggested by Jain and Uludag in [7], in which facial data is hidden in an artificial fingerprint image. Hassanien describes a watermarking method using the discrete wavelet transform in [8] to embed iris images into a host image. Dong and Tan explore the effect of embedding iris templates into host images using robust watermarking in [5], and the extent to which attacks on the host image reduce the ability to extract and match the embedded iris template. Although the majority of algorithms proposed for biometric steganography are robust, it is suggested in [3] that robust watermarking should not be used for steganographic purposes, as the robustness typically implies an increase in perceptibility, the key feature in this scenario [6].

Watermarking can also be used to authenticate the sample and sensor by embedding a fragile watermark containing an identifier unique to the camera and a cryptographic hash of
the original, creating a ‘secure camera’. On verification, if the watermark cannot be extracted and verified, then the image has been tampered with and cannot be authenticated. Blythe and Fridrich use this approach in [9] to create a secure digital camera for securing iris images, while Yeung and Pankanti use a fragile watermark to secure fingerprint data in [10].

III. EXPERIMENTAL METHOD

Experiments were performed using the ICE-Right database [22] released by the National Institute of Standards and Technology (NIST) in 2005. It contains 1425 monochrome 8-bit images, from 124 subjects’ right eyes. In order to focus the watermarking effect on the iris itself, all images were cropped to 320 × 320 pixels, centred on the iris.

Seven watermarking algorithms were implemented for the study and are briefly described below. They were selected to investigate a variety of different watermarking techniques, including state-of-the art approaches.

C. Iris Detection

The majority of iris recognition algorithms follow the well-known work of Daugman [19], including the OSIRIS (Open Source for IRIS) [20] implementation used for this study.

After image acquisition, the pupil is located, and from here the iris is localized using a circular Hough transform followed by an active contour approach [21], to identify the two approximately concentric circles of the pupil and iris. The identified iris is unwrapped and normalized using polar coordinates to a rectangular texture of 512 × 64 pixels. A mask of the same size is generated that identifies regions which should be excluded from the feature detection, e.g. eyelids. Finally, the iris code is generated by applying 2D Gabor filters to the masked image at selected points and coding the output coefficients to produce a binary sequence. Matching of iris codes is performed by finding the minimum Hamming distance between two codes at multiple shifted positions.
Watermarking was performed using the seven algorithms discussed with a randomly generated payload at four embedding capacities: 0.25, 0.50, 0.75 and 1.00 bits per pixel (bpp). Where the required capacity could not be achieved with an algorithm natively, as was the case for APD at all levels, and DIFF, PDA and SP at and above 0.50 bpp, the image was recursively watermarked until the total required payload capacity was reached. Due to the nature of the algorithms, not all schemes could watermark all the images at all capacities, even using recursion. Where the number of images that could be watermarked by an algorithm at a given capacity was less than half the total number of images, the algorithm/capacity pair was excluded from the group comparison. Consequently LSB at 1.00 bpp, and both APD and DCT at 0.75 and 1.00 bpp were excluded. A total of 608 images were successfully watermarked at all of the selected capacities by all algorithms to create 23 sets of 608 watermarked images, plus the original images set.

All of the image sets were processed with OSIRIS to generate the iris codes. Two matching scenarios were investigated: matching of original images to watermarked images (Orig-Wmk), and matching of watermarked images to watermarked images (Wmk-Wmk). The former scenario describes the situation where an unwatermarked template is compared to a sensor image which has been watermarked to create a ‘secure camera’. The later represents the case where the stored template was also generated from a watermarked sensor image.

For Wmk-Wmk matching, each watermarked code in a set was matched with every other watermarked code in the set, for a total of 181,302 inter-class and 3,226 intra-class comparisons within each set. For Orig-Wmk matching, all original images were matched to all other watermarked images, for a total of 362,604 inter-class comparisons and 6,452 intra-class comparisons. Given the computational cost of performing over 500,000 comparisons in total for each set, it was not possible to investigate more than one iris recognition algorithm. Hmmerle-Uhl et al. [4] found that the specific recognition algorithm used made no difference in results for robust watermarking, suggesting the use of a single algorithm should not affect the conclusions that may be drawn. However, extension to further iris algorithms is a clearly desirable area of further work.

### IV. Experimental results

The mean quality of the watermarked images compared to the originals is shown in Table I. The PSNR ranges from 25.9 to 49.5 dB, while the SSIM [30] ranges from 0.64 to 0.99. This level of degradation is approximately comparable to the PSNR of 30 dB and 45 dB used by Hmmerle-Uhl et al. [4] when investigating robust watermarking, although in that case two embedding capacities were used, approximately 0.001 bpp and 0.010 bpp. The considerable additional payload capacity achieved thanks to the use of fragile watermarking could facilitate additional applications.

Fig. 1 shows the detection error trade-off (DET) curves for Wmk-Wmk matching. They show the trade-off between the false accept rate (FAR) and false reject rate (FRR) along with the equal error rate (EER) for each matching case. The match rate for the original images, where no watermarking has been applied, is shown for reference in each plot. The optimal DET curve would be positioned to the bottom left of each.

<table>
<thead>
<tr>
<th>Watermark &amp; bpp</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
<th>Wmk-Wmk</th>
<th>Orig-Wmk</th>
<th>Wmk-Wmk</th>
<th>Orig-Wmk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AOC p-value</td>
<td>EER (%)</td>
<td>AOC p-value</td>
<td>EER (%)</td>
<td>AOC p-value</td>
<td>EER (%)</td>
</tr>
<tr>
<td>APD 0.25</td>
<td>41.6</td>
<td>0.95</td>
<td>0.988</td>
<td>0.183</td>
<td>5.08</td>
<td>0.988</td>
</tr>
<tr>
<td>0.50</td>
<td>36.1</td>
<td>0.90</td>
<td>0.987</td>
<td>0.152</td>
<td>5.10</td>
<td>0.987</td>
</tr>
<tr>
<td>DCT 0.25</td>
<td>39.0</td>
<td>0.96</td>
<td>0.988</td>
<td>0.259</td>
<td>4.93</td>
<td>0.988</td>
</tr>
<tr>
<td>0.50</td>
<td>35.5</td>
<td>0.92</td>
<td>0.988</td>
<td>0.161</td>
<td>5.13</td>
<td>0.988</td>
</tr>
<tr>
<td>DIFF 0.25</td>
<td>40.6</td>
<td>0.97</td>
<td>0.987</td>
<td>0.842</td>
<td>4.78</td>
<td>0.989</td>
</tr>
<tr>
<td>0.50</td>
<td>33.6</td>
<td>0.88</td>
<td>0.990</td>
<td>0.544</td>
<td>4.91</td>
<td>0.991</td>
</tr>
<tr>
<td>0.75</td>
<td>31.3</td>
<td>0.84</td>
<td>0.988</td>
<td>0.425</td>
<td>5.09</td>
<td>0.989</td>
</tr>
<tr>
<td>1.00</td>
<td>25.9</td>
<td>0.64</td>
<td>0.985</td>
<td>0.000</td>
<td>5.94</td>
<td>0.982</td>
</tr>
<tr>
<td>DWT 0.25</td>
<td>41.3</td>
<td>0.97</td>
<td>0.988</td>
<td>0.918</td>
<td>5.07</td>
<td>0.989</td>
</tr>
<tr>
<td>0.50</td>
<td>36.5</td>
<td>0.93</td>
<td>0.989</td>
<td>0.694</td>
<td>4.98</td>
<td>0.989</td>
</tr>
<tr>
<td>0.75</td>
<td>31.3</td>
<td>0.84</td>
<td>0.989</td>
<td>0.264</td>
<td>4.82</td>
<td>0.990</td>
</tr>
<tr>
<td>1.00</td>
<td>27.7</td>
<td>0.73</td>
<td>0.987</td>
<td>0.034</td>
<td>5.16</td>
<td>0.988</td>
</tr>
<tr>
<td>LSB 0.25</td>
<td>49.5</td>
<td>0.99</td>
<td>0.990</td>
<td>0.048</td>
<td>4.83</td>
<td>0.991</td>
</tr>
<tr>
<td>0.50</td>
<td>46.5</td>
<td>0.98</td>
<td>0.990</td>
<td>0.039</td>
<td>4.68</td>
<td>0.991</td>
</tr>
<tr>
<td>0.75</td>
<td>35.7</td>
<td>0.85</td>
<td>0.989</td>
<td>0.240</td>
<td>4.85</td>
<td>0.990</td>
</tr>
<tr>
<td>PDA 0.25</td>
<td>41.4</td>
<td>0.97</td>
<td>0.990</td>
<td>0.000</td>
<td>4.55</td>
<td>0.993</td>
</tr>
<tr>
<td>0.50</td>
<td>34.9</td>
<td>0.91</td>
<td>0.989</td>
<td>0.187</td>
<td>4.80</td>
<td>0.990</td>
</tr>
<tr>
<td>0.75</td>
<td>31.2</td>
<td>0.83</td>
<td>0.989</td>
<td>0.397</td>
<td>5.03</td>
<td>0.990</td>
</tr>
<tr>
<td>1.00</td>
<td>26.1</td>
<td>0.65</td>
<td>0.987</td>
<td>0.027</td>
<td>5.34</td>
<td>0.985</td>
</tr>
<tr>
<td>SP 0.25</td>
<td>43.1</td>
<td>0.98</td>
<td>0.989</td>
<td>0.056</td>
<td>4.79</td>
<td>0.991</td>
</tr>
<tr>
<td>0.50</td>
<td>37.8</td>
<td>0.94</td>
<td>0.989</td>
<td>0.049</td>
<td>4.90</td>
<td>0.990</td>
</tr>
<tr>
<td>0.75</td>
<td>33.6</td>
<td>0.88</td>
<td>0.987</td>
<td>0.041</td>
<td>5.02</td>
<td>0.986</td>
</tr>
<tr>
<td>1.00</td>
<td>28.4</td>
<td>0.72</td>
<td>0.987</td>
<td>0.048</td>
<td>5.02</td>
<td>0.985</td>
</tr>
</tbody>
</table>
plot. In general, it can be seen that for lower embedding rates the DET curve for each algorithm closely follows that of the reference curve, in some cases outperforming it. At the highest embedding level, 1.00 bpp, the curve deviates significantly for SP, PDA and particularly DIFF. These large deviations are associated with an observable increase in EER. Space considerations prevent displaying DET curves for Orig-Wmk matching, however the watermarked curves broadly follow those of Wmk-Wmk matching.

Table I summarizes the DET curves for each matching scenario by providing the area over the DET curve (AOC) and the EER. The larger the AOC, the better the performance, with the optimal area being 1.0. Also shown is the p-value found when testing for significant difference in AOC between each watermarked DET curve and the reference curve, using the method of DeLong et al. [31].

The precise behaviour of the DET curves is different for each watermarking algorithm, varying with embedded capacity and matching scenario. APD shows no significant difference to the original curve in terms of AOC, however the EER is greater in all cases. As may be expected, the EER for 0.50 bpp is larger than for 0.25 bpp. Also, Wmk-Wmk matching produces
higher EERs than Orig-Wmk matching.

DCT also shows no significant difference in terms of AOC, however the EER is not larger than the originals in all cases. While the EER for 0.50 bpp is greater than for 0.25 bpp in both matching scenarios, Wmk-Wmk matching actually outperforms both Orig-Wmk and the original EER. Watermarking in this case results in an improvement in the segmentation stage of the iris recognition algorithm compared to the original images, giving a slight improvement in EER. This may be due to increased high frequency content in the watermarked image allowing easier delineation of boundaries.

This effect is also observed for DIFF watermarking at lower embedding levels, with the improvement in segmentation resulting in an AOC that is significantly improved compared to the original (p < 0.05) at 0.50 bpp for Wmk-Wmk matching. However at the highest embedding level, the AOC is significantly worse than the original curve and the EER significantly larger. This is due to the effect of watermarking on the template generation stage of the recognition algorithm.

Although watermarking can improve the segmentation, the effect of watermarking is always found to be negative on the template generation stage, with the negative effects larger for Wmk-Wmk matching. Consequently the overall DET curve at a given capacity is the interaction between the improvement in segmentation and deterioration of template generation.

This trade-off between segmentation and template generation is responsible for the trends seen in the remaining algorithms. For DWT, the improvement relative to the original images peaks at 0.75 bpp for both scenarios. However, at 1.00 bpp, the adverse effect on template generation results in a significantly poorer curve for Orig-Wmk matching.

The comparatively high image quality of LSB results in significant improvements over the original images for both scenarios at 0.25 and 0.50 bpp. The watermarking distortion allows improved segmentation but is insufficient to cause disruption to the template generation process. Improvements in EER can also be seen across all embedding capacities.

PDA shows a significant improvement in AOC at 0.25 bpp, and a significant decrease in performance of AOC at 1.00 bpp. This is again typical of the segmentation improvement and template generation detriment trade-off described previously.

SP shows slightly better performance in Wmk-Wmk matching than in Orig-Wmk matching. The AOC is significantly improved for Wmk-Wmk at 0.25 bpp but not for Orig-Wmk, while at 0.75 bpp the performance is significantly worse for Orig-Wmk but not for Wmk-Wmk. In both cases 1.00 bpp causes a significant reduction in performance.

The experiments presented demonstrate some unexpected consequences of fragile watermarking when applied to iris recognition. Firstly, the impact of watermarking is not always negative, but rather was found to improve recognition performance in several cases. This is due to an improvement in the segmentation stage of the recognition algorithm. The level of this effect varied between watermarkers.

In contrast to this improvement, the template generation stage of the recognition algorithm was always adversely affected by watermarking. The interplay of these two factors leads to the second observation, that increasing the capacity does not always result in a reduction in recognition performance. The capacity of the optimum performing DET curve varies with the particular watermarking scheme.

Finally, Orig-Wmk matching does not always outperform Wmk-Wmk matching. Again related to the described trade-off, Wmk-Wmk matching tends to outperform Orig-Wmk at lower capacities where the improvement due to segmentation is large. It is typically outperformed by Orig-Wmk at high capacities where the detriment due to template generation dominates.

The levels of embedding, and hence degradation, used in this study are significantly higher than would be expected for many scenarios. For example, iris template data typically requires only 2048 bits; in this case equivalent to 0.02 bpp of data. However, the additional capacity allows embedding of multiple biometrics, along with further security data such as image hashes etc., creating the possibility of more powerful security applications.

While reversible algorithms were used in the study, they were treated simply as fragile watermarking algorithms. Although reconstructing the original image prior to iris matching would give unaffected results, it is the effect of watermarking, without the security risks and computational overhead of original image extraction, that is of interest. Also removing the watermark would leave the image unprotected, and hence would offer little advantage over purely cryptographic methods. The real value of watermarking lies in the direct replacement of an original image with its watermarked version.

V. CONCLUSION

This study showed that fragile watermarking significantly affects iris recognition performance in many cases. The extent to which iris recognition is affected, and whether the effect is positive or negative, depends highly on the watermarking scheme, the level of embedding, and the matching scenario, i.e. whether watermarked images are matched to watermarked or original images.

No significant decrease in AOC was found for any algorithm at rates of 0.50 bpp or less, with many algorithms showing a significant improvement. This suggests that for the watermarking algorithms studied, rates of embedding less than 0.50 bpp would be acceptable.

While this study shows promising results for the reversible watermarking algorithms studied, it is by no means conclusive in determining the best algorithm for the scenario, with many different approaches proving successful at low rates. However at the highest embedding rate, the error expansion techniques of DIFF, PDA and SP show the largest errors, suggesting they are not suited to applications with large capacity requirements. A more extensive study with a wider variety of capacities, alternative iris recognition algorithms and additional iris databases may clarify this effect.
REFERENCES


