White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/108099/

Version: Accepted Version

Article:
Bhowmik, D., Oakes, M. and Abhayaratne, C. (2016) Visual Attention-based Image Watermarking. IEEE Access. p. 1.

https://doi.org/10.1109/ACCESS.2016.2627241

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

Visual Attention-based Image Watermarking

Deepayan Bhowmik, Matthew Oakes and Charith Abhayaratne

***Abstract*—Imperceptibility and robustness are two complementary but fundamental requirements of any watermarking algorithm. Low strength watermarking yields high imperceptibility but exhibits poor robustness. High strength watermarking schemes achieve good robustness but often infuse distortions resulting in poor visual quality in host media. If distortion due to high strength watermarking can avoid visually attentive regions, such distortions are unlikely to be noticeable to any viewer. In this paper, we exploit this concept and propose a novel visual attention-based highly robust image watermarking methodology by embedding lower and higher strength watermarks in visually salient and non-salient regions, respectively. A new low complexity wavelet domain visual attention model is proposed that allows us to design new robust watermarking algorithms. The proposed new saliency model outperforms the state-of-the-art method in joint saliency detection and low computational complexity performances. In evaluating watermarking performances, the proposed blind and non-blind algorithms exhibit increased robustness to various natural image processing and filtering attacks with minimal or no effect on image quality, as verified by both subjective and objective visual quality evaluation. Up to 25% and 40% improvement against JPEG2000 compression and common filtering attacks, respectively, are reported against the existing algorithms that do not use a visual attention model.**

***Index Terms*—Visual saliency, wavelet, watermarking, robustness, subjective test.**

I. Introduction

As digital technologies have shown a rapid growth within the last decade, content protection now plays a major role within content management systems. Of the current systems, digital watermarking provides a robust and maintainable solution to enhance media security. Evidence of popularity of watermarking is clearly visible as watermarking research has resulted in 11,833 *image watermarking* papers published in last 20 years and 1385 (11.7%) alone in 2014-15[1]. The visual quality of host media (often known as imperceptibility) and robustness are widely considered as the two main properties vital for a good digital watermarking system. They are complimentary to each other and hence challenging to attain the right balance between them. This paper proposes a new approach to achieve high robustness in watermarking while not affecting the perceived visual quality of the host media by exploiting the concepts of visual attention.

The Human Visual System (HVS) is sensitive to many salient features that lead to attention being drawn towards specific regions in a scene and it is a well studied topic in psychology and biology [1], [2]. Visual Attention (VA) is an important concept in a complex ecological system: for example, identifying potential danger, *e.g.*, prey, and predators quickly in a cluttered visual world [3] as attention to one target leaves other targets less available [4]. Recently a considerable amount of work has been reported in the literature in modelling visual attention [5]–[7] that has applications in many related domains including media quality evaluation [8] and computer vision [9]–[11]. Visual attention modelling characterises the scene (image) to segment regions of visual interest and hence a suitable concept for assessing the relevance of a region in an image for embedding watermark data without affecting the perceived visual quality. This paper proposes a new framework for highly robust and imperceptible watermarking that exploits this concept.

By employing VA concepts within the digital watermarking, an increased overall robustness against various adversary attacks can be achieved, while subjectively limiting any perceived visual distortions by the human eye. Our method proposes a new frequency domain Visual Attention Model (VAM) to find inattentive areas in an image, so that the watermarking strength in those areas can be made higher to make it more robust at the expense of the visual quality in such areas as shown in the example in Fig. 1. Fig. 1a) shows an example of a low strength watermarking that has highest imperceptibility but a very low robustness while Fig. 1c) shows an example of high strength watermarking resulting in high level of visual distortion. Fig. 1b) shows an example of the proposed concept where VAM-based watermarking is used for embedding the high strength watermarking in visually inattentive areas (mainly in the background) leading to negligible distortion.

Related work includes defining a Region of Interest (ROI) [12]–[19] and increasing the watermark strength in the ROI to address cropping attacks. However, in these works, the ROI extraction were only based on foreground-background models rather than VAM. There are major drawbacks of such solutions: *a)* increasing the watermark strength within eye catching frame regions is perceptually unpleasant as human attention will naturally be drawn towards any additional embedding artefacts, and *b)* scenes exhibiting sparse salience will potentially contain extensively fragile or no watermark data. Moreover, Sur *et al.* [20] proposed a pixel domain algorithm to improve embedding distortion using an existing visual saliency model described in [3]. However, the algorithm only discusses its limited observation on perceptual quality without considering any robustness. A zero watermark embedding scheme is proposed in [21] that also used the saliency model proposed by Itti *et al.* [3]. However, a *zero watermarking* algorithm is often considered as an image signature and does

D. Bhowmik is with the Department of Computing, Sheffield Hallam University, Sheffield, S1 1WB, U.K. (e-mail: deepayan.bhowmik@shu.ac.uk).

M. Oakes is with the University of Buckingham, Buckingham, MK18 1EG, U.K. (e-mail: matthew.oakes@buckingham.ac.uk).

C. Abhayaratne is with the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, U.K. (email: c.abhayaratne@sheffield.ac.uk).

Manuscript received XXX XX, XXXX; revised XXX XX, XXXX.

[1]Sources: www.scopus.com

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

(a) Low strength watermarking: Highest imperceptibility but lowest robustness.

(b) VAM-based watermarking: High imperceptibility and high robustness.

(c) High strength watermarking: Highest robustness but lowest imperceptibility.

Fig. 1: Example scenario of visual attention model based watermarking.

not qualify for a comparison with traditional watermarking schemes as it does not embed any watermark data.

In this paper, we propose a novel visual attention-based approach for highly robust image watermarking, while retaining the high perceived visual quality as verified by subjective testing. Firstly, we propose a bottom-up saliency model that estimates salience directly within the wavelet domain to enhance compatibility with watermarking algorithms that are based on the same wavelet decomposition schemes. Secondly, the watermark is embedded in the wavelet domain with the watermark strength controlled according to the estimated saliency level in image pixels (in wavelet domain) leading to highly robust image watermarking without degrading the media quality. Both non-blind and blind watermarking algorithms are proposed to demonstrate the capability and the effectiveness of this proposed approach. Performance of the saliency model and its application to watermarking are evaluated by comparing with existing schemes. Subjective tests for media quality assessment recommended by the International Telecommunication Union (ITU-T) [22], which are largely missing in the watermarking literature for visual quality evaluation, are also conducted to complement the objective measurements. The main contributions of this work are as follows:

- A wavelet-based visual attention model that is compatible for wavelet-based image watermarking applications.
- New blind and non-blind watermarking algorithms that result in highly imperceptible watermarking that is robust to common filtering and compression attacks.
- Watermark embedding distortion evaluation based on subjective testing that follows ITU-T recommendations.

The saliency model and the watermarking algorithms are evaluated using existing image datasets described in § V-B. The initial concept and the results were reported earlier in the form of a conference publication [23] while this paper discusses the proposed scheme in detail with exhaustive performance evaluation.

II. Background and related work

A. Visual attention models

Our eyes receive vast streams of visual information every second (108-109 bits) [24]. This input data requires significant processing, combined with various intelligent and logical mechanisms to distinguish between any relevant and insignificant redundant information. This section summarises many of the available computational methodologies to estimate the VA of an image or static scene.

Human vision behavioural studies [25] and feature integration theory [26] have prioritised the combination of three visually stimulating low level features: intensity, colour and orientation which comprise the concrete foundations for numerous image domain saliency models [3], [5], [27]–[29]. Salient objects are not size specific therefore Multi-Resolution Analysis (MRA) is adopted within many models [3], [28], [30], [31]. Classical low level bottom-up computational saliency model framework was proposed by Itti *et al.* [28] and commonly know as Itti model. In Itti model, the image is down sampled into various scales. Colour features were extracted using Gaussian Pyramids while the orientation features were extracted by Gabor pyramids. This is followed by combining features across scales using a center-surround difference and normalisation approach to determine contrasting regions of differing intensity, colour and orientation. A winner-takes-all system fuses together each of the feature maps into an output saliency estimation.

Itti model has provided the framework for various recent works [3], [32], [33]. For example, Erdem [32] adopts classical architecture, as used within the Itti model [28], to segment intensity, colour and orientation contrasts. However, a non-linear feature map combination is implemented. Firstly, the input image is decomposed into numerous non-overlapping frame regions and the visual saliency of each area is computed by examining the surrounding regions. Any regions portraying a high visual saliency exhibit high dissimilarity to their neighbouring regions in terms of their covariance representations based on intensity, colour and orientation.

The Ngau model [33] estimates visual salience by locating coefficients which diverge greatly from the local mean within the low frequency approximation wavelet subband. On contrary, in this work we propose a novel wavelet-based visual saliency framework designed to perform across both Luma and Chroma channels to provide an improved estimation of visual salience by combining colour, orientation and intensity contrasts.

Similar to Itti model, Li model [30] first down samples the

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

image into various scales. Then instead of Gabor pyramids, it uses a 1-level wavelet decomposition to extract orientation maps. To generate a specific orientation map, it replaces all the wavelet coefficients apart from those in the subband of the particular orientation with zeros and the inverse wavelet transform is performed. Three orientation maps for horizontal, vertical and diagonal subbabands are generated, followed by generating 3 saliency maps for each orientation and simply adding the directional saliency maps to generate the final saliency map. This separable treatment of orientation maps for finding the final saliency map is a major weak point of Li model.

Some studies incorporate high level features within the low level saliency design, such as, face detection [34], text detection [35] and skin detection [36]. A major advantage of these high and low level feature models is the simplicity to incorporate additional features, within the existing framework, combined with a linear feature weighting, depending on the application. These top-down models [34], [36] are dependent upon prior scene knowledge upon distinguishable features.

The main drawback in all saliency bottom-up models lies within the computational complexity as the MRA approach generates many processable feature maps for combination. Various other proposed techniques can detect attentive scene regions by histogram analysis [37], locating inconsistencies within neighbouring pixels [38], object patch detection [31], graph analysis [39], log-spectrum analysis [40] and symmetry [41]. In another example, the Rare model [42] combines both colour and orientation features, deduced from multi-resolution Gabor filtering. A rarity mechanism is implemented to estimate how likely a region is to be salient, by histogram analysis.

Our proposed saliency model (presented in details in § III) uses a multi-level wavelet decomposition for multi-resolution representation, so that the same framework can be used in wavelet-domain watermarking. It does not use down sampled images as in [3] or [30]. Moreover, it does not use Gabor pyramids as in [3] or 1-level wavelet selected subband reconstruction as in [30]. Instead, it uses all detail coefficients across all wavelet scales for center-surround differencing and normalisation. Finally it treats 3 orientation features in a non-separable manner to fuse them and obtain the saliency map.

B. Wavelet-based watermarking

Frequency-based watermarking, more precisely wavelet domain watermarking, methodologies are highly favoured in the current research era. The wavelet domain is also compliant within many image coding, *e.g.*, JPEG2000 [43] and video coding, *e.g.*, Motion JPEG2000, Motion-Compensated Embedded Zeroblock Coding (MC-EZBC) [44], schemes, leading to smooth adaptability within modern frameworks. Due to the multi-resolution decomposition and the property to retain spatial synchronisation, which are not provided by other transforms (the Discrete Cosine Transform (DCT) for example), the Discrete Wavelet Transform (DWT) provides an ideal choice for robust watermarking [45]–[61].

When designing a watermarking scheme there are numerous features to consider, including the wavelet kernel, embedding coefficients and wavelet subband selection. Each of these particular features can sufficiently impact the overall watermark characteristics [62] and is largely dependant upon the target application requirements.

a) Wavelet Kernel Selection: An appropriate choice of wavelet kernel must be determined within the watermarking framework. There have been previous studies to show that the performance of watermark robustness and imperceptibility is dependant on the wavelet kernels [47], [49], [63]. The orthogonal Daubechie wavelets are a favourable choice with many early watermarking schemes [52]–[57], although the later introduction of bi-orthogonal wavelets within the field of digital watermarking has increased their usage [58]–[61].

b) Host Coefficient Selection: Various approaches exist to choose suitable transform coefficients for embedding a watermark. In current methods, coefficient selection is determined by the threshold values based upon the coefficient magnitude [59] or a pixel masking approach based upon HVS [55] or the median of 3 coefficients in a 3×1 overlapping window [54] or simply by selecting all the coefficients [52], [53], [56].

c) Wavelet Subband Selection: The choice of subband bears a large importance when determining the balance between robustness of the watermark and imperceptibility. Embedding within the high frequency domain subbands [52], [53], [55], [56], [64] can often provide great imperceptibility but with limited watermark robustness capabilities. Contradictory schemes embed data only within the low frequency subbands [54], [57], [61] aimed towards providing a high robustness. Spread spectrum domain embedding [59], [65]–[67] modifies data across all frequency subbands, ensuring a balance of both low and high frequency watermarking characteristics. The number of decomposition levels is also an important factor. Previous studies have researched watermarking schemes using two [52], [53], [64], three [60], [61], [68] and four or more [55]–[57] wavelet decomposition levels.

With the motivation to propose a highly imperceptible as well as robust watermarking algorithm, our proposed approach requires an efficient wavelet-based saliency model for directly integrating within the wavelet-based watermarking framework. Previous wavelet domain saliency models, either provide insufficient model performance as they are based on coefficient average variance [33] or require multiple frame resizing prior to saliency estimation [30] resulting in spawning multiple instances of wavelet transforms. Estimating salience directly from within the wavelet domain enhances compatibility with the wavelet-based watermarking framework as described in § IV.

III. The visual attention model

In this section, a novel model to detect saliency regions within an image is proposed. The proposed model, as shown in Fig. 2, employs the multi-level 2D wavelet decomposition combined with HVS modelling to capture the orientation features on luminance and chrominance channels leading to overall saliency information. Physiological and psychophysical evidence demonstrate that visually stimulating regions occur at different scales within the visual content [69]. Consequently,

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

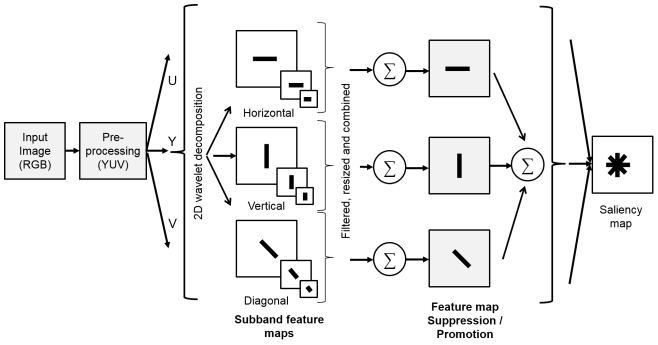


Fig. 2: Overview of the proposed saliency model.

the model proposed in this work exploits the multi-resolution property of the wavelet transform. The image saliency model is presented in the following subsections. Firstly, § III-A analyses the spatial scale implemented within the design and § III-B describes the saliency algorithm. Finally, § V-B shows the model performance.

A. Scale Feature Map Generation

As the starting point in generating the saliency map from a colour image, RGB colour space is converted to YUV colour spectral space as the latter exhibits prominent intensity variations through its luminance channel Y. Firstly, the 2D forward DWT (FDWT) is applied on each Y, U and V channel to decompose them L levels. The wavelet kernel used is the same as that used for watermarking. At this juncture, we define wavelet related acronyms used later in describing the proposed model. The 2D FDWT decomposes an image in frequency domain expressing coarse grain approximation of the original signal along with three fine grain orientated edge information at multiple resolutions. As shown in Fig. 3, LH_i, HL_i and HH_i subbands in the decomposition level $i \in \mathbb{N}_1$ emphasise horizontal, vertical and diagonal contrasts within an image, respectively, portraying prominent edges in various orientations. These notations are used herein to refer respective subbands. The absolute magnitude of wavelet coefficients is considered in the subsequent analysis in order to prevent negative salient regions as contrasting signs can potentially nullify salient regions when combined. The absolute values of the coefficients are then normalised within the range $[0, R]$, where R is the upper limit of the normalised range. This normalises the overall saliency contributions from each subband and prevents biassing towards the finer scale subbands.

To provide full resolution output maps, each of the high frequency subbands is consequently interpolated up to full frame resolution. Eq. (1) depicts this process showing how the absolute full resolution subband feature maps lh_i, hl_i and hh_i are generated from the LH_i, HL_i and HH_i subbands in the wavelet decomposition level, i, in a given channel in the colour space, respectively:

$$\begin{aligned} lh_i &= (|LH_i|^{\uparrow 2^i}), \\ hl_i &= (|HL_i|^{\uparrow 2^i}), \\ hh_i &= (|HH_i|^{\uparrow 2^i}), \end{aligned} \tag{1}$$

where $\uparrow 2^i$ is the bilinear up-sampling operation by a factor 2^i for the wavelet decomposition level i. Fusion of lh_i, hl_i and hh_i for all wavelet decomposition levels, L, provides a feature map for each subband in the given colour channel.

The total number of wavelet decomposition levels used in the proposed VAM depends on the resolution of the image. Due to dyadic nature of the multi-resolution wavelet transform, the image resolutions are decreased after each wavelet decomposition level. This is useful in capturing both small and large structural information at different scales. However, too many levels of decomposition may distort the spatial synchronisation of objects within the image, limiting the useful contribution of coefficients towards the overall saliency map at very coarse resolutions. An example of such distortion is shown in Fig. 4 visualising the successive coefficient magnitude of each of the subbands, lh_i for the luminance channel of an image (of resolution 414 x 288). In this example, after five levels of decomposition, the threshold to retain coefficient spatial synchronisation has been surpassed. Consequently, a highly distorted profile is obtained for the interpolated higher successive decompositions containing limited meaningful information available for saliency computation.

B. Saliency Map Generation

The interpolated subband feature maps, lh_i, hl_i and hh_i, for all L levels are combined by a weighted linear summation as illustrated in Eq. (2):

$$\begin{aligned} lh_{1\cdots L_X} &= \sum_{i=1}^{L} lh_i * \tau_i, \\ hl_{1\cdots L_X} &= \sum_{i=1}^{L} hl_i * \tau_i, \\ hh_{1\cdots L_X} &= \sum_{i=1}^{L} hh_i * \tau_i, \end{aligned} \tag{2}$$

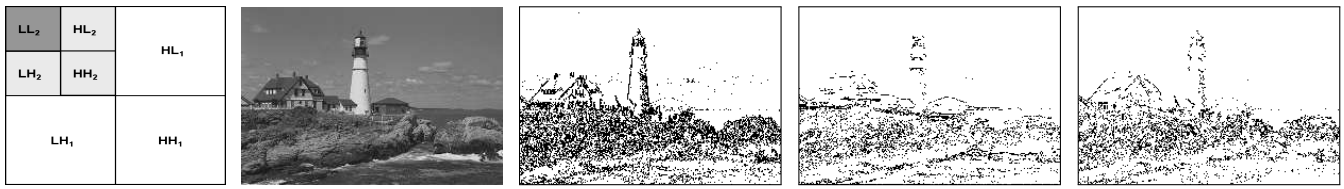
where τ_i is the subband weighting parameter and $lh_{1\cdots L_X}$, $hl_{1\cdots L_X}$ and $hh_{1\cdots L_X}$ are the subband feature maps for a given spectral channel X, where $X \in \{Y, U, V\}$.

Coarse scale subbands mainly portray edges and other tiny contrasts which can be hard to see. The finely decomposed subband levels only illustrate large objects, neglecting any smaller conspicuous regions. For most scenarios, the middle scale feature maps can express a high saliency correlation although this is largely dependable upon the resolution of the prominent scene objects. To fine tune the algorithm, it is logical to apply a slight bias towards the middle scale subband maps, *i.e.*, making $(\tau_1, \tau_L) < (\tau_2, \tau_{L-1}) < (\tau_3, \tau_{L-2}) < \cdots < \tau_c$, where c is the centre scale. However, in practice this provides a minimal algorithm performance improvement over an equal subband weighting ratio due to the fact that salience is not specific towards a definite resolution [70].

Research suggests promoting feature maps which exhibit a low quantity of strong activity peaks [28], while suppressing maps flaunting an abundance of peaks possessing similar amplitude. Similar neighbouring features inhibit visual attentive selectivity, whereas, a single peak surrounded by

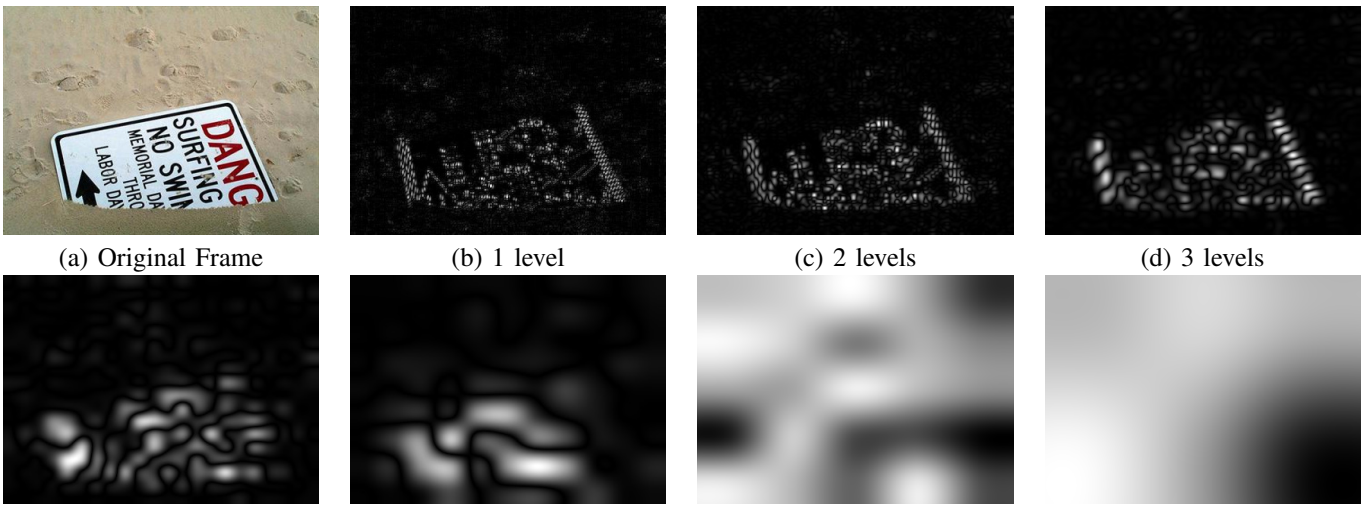
This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access



(a) DWT illustration (b) LL_1 (c) LH_1 (d) HL_1 (e) HH_1

Fig. 3: An example of multiresolution wavelet decomposition. (a) Illustration of 2-level DWT. (b)-(e) One level 2-D decomposition of an example image. (b), (c), (d) and (e) represent approximation (LL_1), vertical (LH_1), horizontal (HL_1) and diagonal (HH_1) subbands, respectively. Wavelet coefficients only with absolute values above the 0.9 quantile (largest 10%) are shown (as inverted image) for high frequency subbands ((c)-(e)) highlighting directional sensitivity.



(a) Original Frame (b) 1 level (c) 2 levels (d) 3 levels

(e) 4 levels (f) 5 levels (g) 6 levels (h) 7 levels

Fig. 4: An example of interpolated LH subbands from 7-level decomposition for each successive wavelet decomposition level.

boundless low activity facilitates visual stimuli. If $\overline{m}$ is the average of local maxima present within the feature map and M is the global maximum, the promotion and suppression normalisation is achieved by Eq. (3):

$$\begin{aligned} \overline{lh}_X &= lh_{1\cdots L_X} * (M - \overline{m})^2, \\ \overline{hl}_X &= hl_{1\cdots L_X} * (M - \overline{m})^2, \\ \overline{hh}_X &= hh_{1\cdots L_X} * (M - \overline{m})^2, \end{aligned} \tag{3}$$

where $\overline{lh}_X$, $\overline{hl}_X$ and $\overline{hh}_X$ are the normalised set of subband feature maps.

Finally, the overall saliency map, S, is generated by

$$S = \sum_{\forall X \in \{Y,U,V\}} w_X * S_X, \tag{4}$$

where w_X is the weight given to each spectral component and S_X is the saliency map for each spectral channel (Y, U, V), which is computed as follows:

$$S_X = \overline{lh}_X + \overline{hl}_X + \overline{hh}_X. \tag{5}$$

An overview of the proposed saliency map S_X generation for a colour channel is shown in Fig. 2. If U or V channels portray sparse meaningful saliency information, only a minimal effect will occur from incorporating these features within the final map, as the structural details are captured in the luminance saliency map, S_Y. However, S_U and S_V are useful for capturing saliency due to change in colour.

IV. Visual Attention-based Watermarking

A visual attention-based ROI dictates the visually most important pixels within an image. Therefore, any distortion in such a region will be highly noticeable to any viewer. In this section, a novel image watermarking scheme is presented using the VAM, where the visual saliency map is computed within the wavelet domain as described in § III. By embedding greater watermark strength (leading to higher distortions and robustness) within the less visually appealing regions, in the host media, a highly robust scheme is attained without compromising the visual quality of the data. A low watermark strength is chosen for the highly visually attentive areas, leading to less distortion. Thus, the perception of watermark embedding distortion can be greatly reduced if any artefacts occur within inattentive regions. By incorporating VA-based characteristics within the watermarking framework, algorithms can retain the perceived visual quality while increasing the overall watermark robustness, compared with non-VA methodologies. Since the VAM proposed in this work in § III-A provides an efficient wavelet domain saliency map generation for images, this can be easily incorporated into wavelet-based watermarking schemes.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

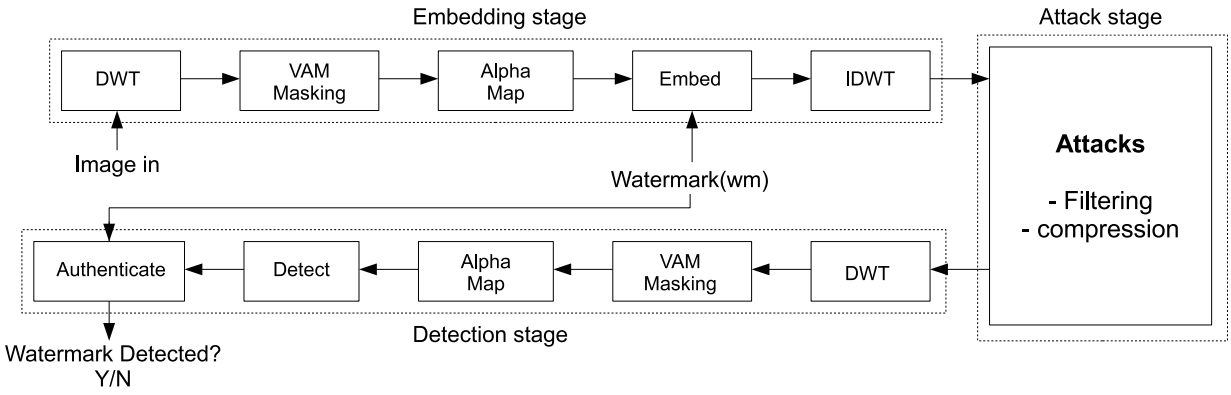


Fig. 5: Visual Attention-based Watermarking Scheme.

This section proposes VA-based watermarking for both blind and non-blind watermarking scenarios. The overview of the VA-based watermarking can be seen in Fig. 5. In both scenarios, a content dependent saliency map is generated which is used to calculate the region adaptive watermarking strength parameter alpha, $\alpha \in [0, 1]$. A lower and higher value of α in salient regions and non-salient regions, respectively, ensures higher imperceptibility of the watermarked image distortions while keeping greater robustness.

A. The watermarking schemes

At this point, we describe the classical wavelet-based watermarking schemes without considering the VAM and subsequently propose the new approach that incorporates the saliency model described in § III. The FDWT is applied on the host image before watermark data is embedded within the selected subband coefficients. The Inverse Discrete Wavelet Transform (IDWT) reconstructs the watermarked image. The extraction operation is performed after the FDWT. The extracted watermark data is compared to the original embedded data sequence before an authentication decision verifies the watermark presence. A wide variety of potential adversary attacks, including compression and filtering, can occur in an attempt to distort or remove any embedded watermark data.

1) Non-blind Watermarking: Magnitude-based multiplicative watermarking [23], [53], [55], [59], [71], [72] is a popular choice when using a non-blind watermarking system, due to its simplicity. Wavelet coefficients are modified based on the watermark strength parameter, α, the magnitude of the original coefficient, $C(m, n)$ and the watermark information, $W(m, n)$. The watermarked coefficients, $C'(m, n)$, are obtained as follows:

$$C'(m, n) = C(m, n) + \alpha W(m, n)C(m, n). \tag{6}$$

$W(m, n)$ is derived from a pseudo-random binary sequence, b, using weighting parameters, W_1 and W_2 (where $W_2 > W_1$), which are assigned as follows:

$$W(m, n) = \begin{cases} W_2 & \text{if } b = 1 \\ W_1 & \text{if } b = 0. \end{cases} \tag{7}$$

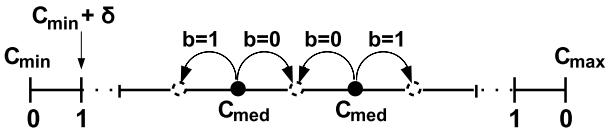


Fig. 6: Blind quantisation-based coefficient embedding.

To obtain the extracted watermark, $W'(m, n)$, Eq. (6) is rearranged as:

$$W'(m, n) = \frac{C'(m, n) - C(m, n)}{\alpha C(m, n)}. \tag{8}$$

Since the non-watermarked coefficients, $C(m, n)$, are needed for comparison, this results in non-blind extraction. A threshold limit of $T_w = \dfrac{W_1 + W_2}{2}$ is used to determine the extracted binary watermark b' as follows:

$$b' = \begin{cases} 1 & \text{if } W'(m, n) \geq T_w \\ 0 & \text{if } W'(m, n) < T_w. \end{cases} \tag{9}$$

2) Blind Watermarking: Quantization-based watermarking [54], [64], [73]–[76] is a blind scheme which relies on modifying various coefficients towards a specific quantization step. As proposed in [54], the algorithm is based on modifying the median coefficient towards the step size, δ, by using a running non-overlapping 3×1 window. The altered coefficient must retain the median value of the three coefficients within the window, after the modification. The equation calculating δ is described as follows:

$$\delta = \alpha \frac{(C_{min}) + (C_{max})}{2}, \tag{10}$$

where C_{min} and C_{max} are the minimum and maximum coefficients, respectively. The median coefficient, C_{med}, is quantised towards the nearest step, depending on the binary watermark, b. Quantisation-based watermark embedding is shown in Fig. 6. The extracted watermark, b', for a given window position, is extracted by

$$b' = \left[\frac{C_{max} - C_{med}}{\delta} \right] \%2, \tag{11}$$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

where % denotes the modulo operator to detect an odd or even number and C_{med} is the median coefficient value within the 3×1 window.

3) Authentication of extracted watermarks: Authentication is performed by comparison of the extracted watermark with the original watermark information and computing closeness between the two in a vector space. Common authentication methods are defined by calculating the similarity correlation or Hamming distance, H, between the original embedded and extracted watermark as follows:

$$H(b, b') = \frac{1}{N} \sum b \oplus b', \tag{12}$$

where N represents the length of the watermark sequence and $\oplus$ is the XOR logical operation between the respective bits.

B. Saliency map segmentation with thresholds

This subsection presents the threshold-based saliency map segmentation which is used for adapting the watermarking algorithms described in § IV-A in order to change the watermark strength according to the underlying visual attention properties. Fig. 7(a) and Fig. 7(b) show an original host image and its corresponding saliency map, respectively, generated from the proposed methodology in § III. In Fig. 7(b), the light and dark regions, within the saliency map, represent the visually attentive and non-attentive areas, respectively. At this point, we employ thresholding to quantise the saliency map into coarse saliency levels as fine granular saliency levels are not important in the proposed application. In addition, that may also lead to reducing errors in saliency map regeneration during watermark extraction as follows. Recalling blind and non-blind watermarking schemes, in § IV-A, the host media source is only available within non-blind algorithms. However in blind algorithms, identical saliency reconstruction might not be possible within the watermark extraction process due to the coefficient values changed by watermark embedding as well as potential attacks. Thus, the saliency map is quantised using thresholds leading to regions of similar visual attentiveness. The employment of a threshold reduces saliency map reconstruction errors, which may occur as a result of any watermark embedding distortion, as justified further in § IV-D.

The thresholding strategy relies upon a histogram analysis approach. Histogram analysis depicts automatic segmentation of the saliency map into two independent levels by employing the saliency threshold, T_s, where $s \in S$ represents the saliency values in the saliency map, S. In order to segment highly conspicuous locations within a scene, firstly, the cumulative frequency function, f, of the ordered saliency values, s, (from 0 to the maximum saliency value, s_{max}) is considered. Then, T_s is chosen as

$$T_s = f^{-1}(p * f_{max}), \tag{13}$$

where p corresponds to the percentage of the pixels that can be set as the least attentive pixels and $f_{max} = f(s_{max})$ corresponds to the cumulative frequency corresponding to the maximum saliency value, s_{max}. An example of a cumulative frequency plot of a saliency map and finding T_s for $p = 0.75$ is shown in Fig. 7(c).

Saliency-based thresholding enables determining the coefficients' eligibility for a low or high strength watermarking. To ensure VA-based embedding, the watermark weighting parameter strength, α, in Eq. (6) and Eq. (10) is made variable $\alpha(j,k)$, dependant upon T_s, as follows:

$$\alpha(j,k) = \begin{cases} \alpha_{max} & \text{if } s(j,k) < T_s, \\ \alpha_{min} & \text{if } s(j,k) \geq T_s, \end{cases} \tag{14}$$

where $\alpha(j,k)$ is the adaptive watermark strength map giving the α value for a the corresponding saliency at a given pixel coordinate (j,k). The watermark weighting parameters, α_{min} and α_{max} correspond to the high and low strength, values respectively and their typical values are determined from the analysis within § IV-C. As shown in Fig. 7(d), the most and the least salient regions are given watermark weighting parameters of α_{min} and α_{max}, respectively. An example of the final VA-based alpha watermarking strength map is shown in Fig. 7(e), where a brighter intensity represents an increase in α. Further test images, with corresponding alpha maps are shown in Fig. 8.

C. Watermark Embedding Strength Calculation

The watermark weighting parameter strengths, α_{max} and α_{min} can be calculated from the visible artifact PSNR limitations within the image. Visual distortion becomes noticeable as the overall Peak Signal to Noise Ratio (PSNR) drops below 40dB [77], so minimum and maximum PSNR requirements are set to approximate 35dB and 40dB, respectively, for both the blind and non-blind watermarking schemes. These PSNR limits ensure maximum amount of data can be embedded into any host image to enhance watermark robustness without substantially distorting the media quality. Therefore it is sensible to incorporate PSNR in determining the watermark strength parameter α.

Recalling PSNR, which measures the error between two images with dimensions $X \times Y$ is expressed on pixel domain as follows:

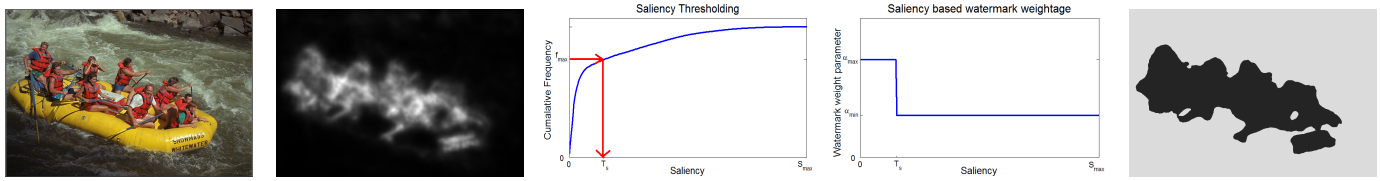
$$\text{PSNR}(I, I') = 10 \log \left(\frac{M^2}{\frac{1}{XY} \sum_{j=1}^{X} \sum_{k=1}^{Y} (I'(j,k) - I(j,k))^2} \right), \tag{15}$$

where M is the maximum coefficient value of the data, $I(j,k)$ and $I'(j,k)$ is the original and watermarked image pixel values at (j,k) indices, respectively. Considering the use of orthogonal wavelet kernels and the Parseval's theorem, the mean square error in the wavelet domain, due to watermarking, is equal to the mean square error in the spatial domain [47]. Therefore, Eq. (15) can be redefined on transform domain for non-blind magnitude based multiplicative watermarking, shown in Eq. (6), as follows:

$$\text{PSNR}(I, I') = 10 \log \left(\frac{M^2}{\frac{1}{XY} \sum_{m=1}^{X} \sum_{n=1}^{Y} (\alpha W(m,n) C(m,n))^2} \right). \tag{16}$$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access



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

Fig. 7: (a) Host image (b) VAM saliency map (saliency is proportional to the grey scale) (c) Cumulative saliency histogram (d) α step graph (e) α strength map (dark corresponds to low strength).

Fig. 8: α strength map examples: *Row 1:* Original Image & *Row 2:* Corresponding α strength map.

By rearranging for α, an expression determining the watermark weighting parameter, depending on the desired PSNR value is derived for non-blind watermarking in Eq. (17) as follows:

$$\alpha = \frac{M}{\sqrt{\frac{10^{(\text{PSNR}(I,I')/10)}}{XY}\sum_{m=1}^{X}\sum_{n=1}^{Y}(W(m,n)C(m,n))^2}}. \quad (17)$$

Similarly for the blind watermarking scheme described in § IV-A2, PSNR in transform domain can be estimated by substituting the median and modified median coefficients, $C_{(med)}$ and $C'_{(med)}$, respectively, in Eq. (15). Then subsequent rearranging results in an expression for the total error in median values, in terms of the desired PSNR as follows:

$$\sum_{m=1}^{X}\sum_{n=1}^{Y}(C'_{(med)} - C_{(med)})^2 = XY\frac{M^2}{10^{(\text{PSNR}/10)}}. \quad (18)$$

Eq. (18) determines the total coefficient modification for a given PSNR requirement, hence is used to α in Eq. (10).

D. Saliency Map Reconstruction

For non-blind watermarking, the host data is available during watermark extraction so an identical saliency map can be generated. However, a blind watermarking scheme requires the saliency map to be reconstructed based upon the watermarked media, which may have got pixel values slightly different to the original host media. Thresholding the saliency map into 2 levels, as described in § IV-B, ensures high accuracy within the saliency model reconstruction for blind watermarking. Fig. 9 demonstrates the saliency map reconstruction after blind watermark embedding compared with the original. A watermark strength of $\alpha_{max} = 0.2$ is embedded within the LL subband after 3 successive levels of wavelet decomposition, giving a PSNR of 34.97dB, using the blind watermarking scheme described in § IV-A2. Fig. 9 shows how applying thresholds to the saliency map can limit any potential reconstruction errors due to embedding artifacts distorting the VAM. The left and right columns show the thresholded original frame and watermarked frame, respectively. By visual inspection Fig. 9(c) and Fig. 9(d) appear indistinguishable, although objective analysis determines only 55.6% of coefficients are identical, leading to difference in computed saliency values. In Fig. 9(e) and Fig. 9(f) 99.4% of saliency coefficients match, hence reconstruction errors are greatly reduced due to thresholding.

V. Performance Evaluation

The performance of the proposed visual attention based watermarking is reported and discussed in this section. The aim of the proposed work is to exploit the visual saliency concepts to embed high strength watermarking leading to high robustness without affecting the perceived visual quality due to embedding distortion. Therefore, the proposed method is evaluated for both visual quality (in § V-C1) as well as robustness (in § V-C2). The visual quality is evaluated using subjective evaluation methods (in § V-A2) as well as traditional objective metrics (in § V-A1). As an intermediate evaluation step, the suitability of the proposed visual attention model also evaluated and compared with the state-of-the-art algorithms in terms of accuracy of estimation and computational complexity in § V-B.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

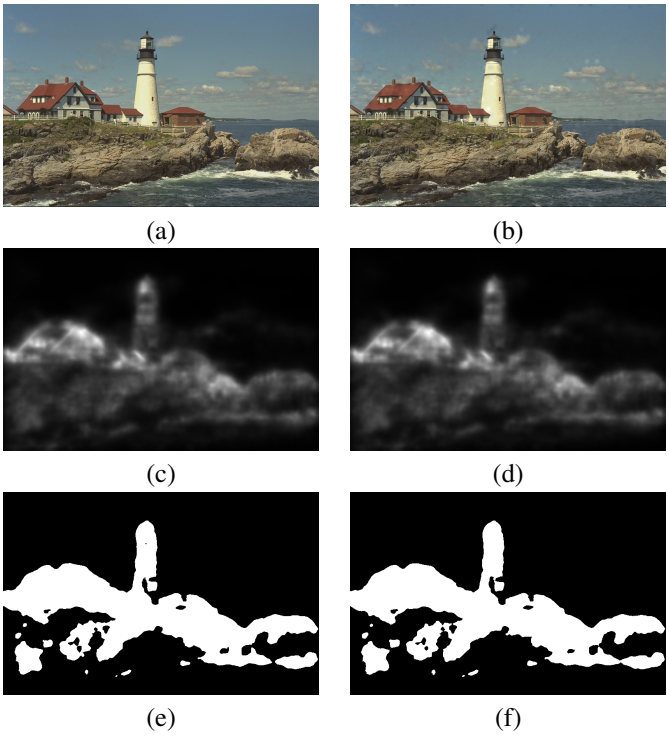


Fig. 9: Saliency map reconstruction - (a) Original host image, (b) Watermarked image embedded using a constant α_{max}, (c) Host image saliency map, (d) Saliency map of watermarked image, (e) Original thresholded saliency map and (f) Reconstructed saliency map thresholded after blind watermark embedding.

A. Visual Quality Evaluation Tools

Visual quality due to embedding distortion in watermarking work is often evaluated using the objective metrics, like PSNR, in the watermarking literature. While objective quality metrics are based on mathematical models, they do not represent the accurate perceived quality. Although, some objective metrics are designed using the HVS model concepts and easy to compute, subjective evaluation allows the accurate measurement of viewers' Quality of Experience (QoE). The subjective evaluations are vital in this work in order to measure the effectiveness of the proposed saliency model for maintaining the imperceptibility in the proposed VA-based watermarking.

1) Objective Evaluation Tools: Objective metrics define a precise value, dependant upon mathematical modelling, to determine visual quality. Such metrics include PSNR, Structural Similarity Index Measure (SSIM) [78] and Just Noticeable Difference (JND) [79]. One of the most commonly used metric, PSNR, stated in Eq. (15), calculates the average error between two images. SSIM focuses on quality assessment based on the degradation of structural information. It assumes that the HVS is highly adapted for extracting structural information from a scene. By using local luminance and contrast rather than average luminance and contrast, the structural information in the scene is calculated.

2) Subjective Evaluation Techniques: Subjective evaluation measures the visual quality by recording the opinion of human

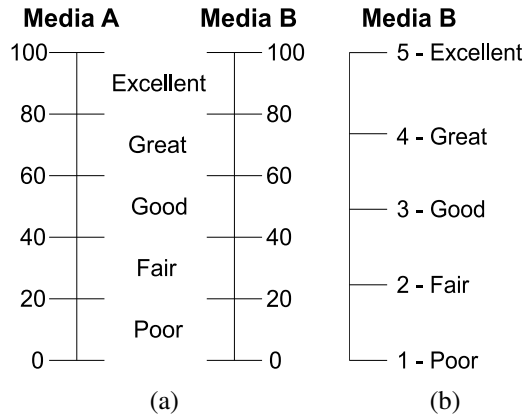


Fig. 10: Subjective testing visual quality measurement scales (a) DCR continuous measurement scale (b) ACR ITU 5-point discrete quality scale.

subjects on the perceived visual quality. In this work, the testing standard specification, defined within the International Telecommunication Union (ITU-T) [22] was followed. This work employs two subjective evaluation metrics, that are computed based on the subjective viewing scores, as follows:

DSCQT: Double Stimulus Continuous Quality Test (DSCQT) subjectively evaluates any media distortion by using a continuous scale. The original and watermarked media is shown to the viewer in a randomised order, who must provide a rating for the media quality of the original and watermarked images individually using a continuous scaling, as shown in Fig. 10(a). Then the Degradation Category Rating (DCR) value is calculated by the absolute difference between the subjective rating for the two test images.

DSIST: Double Stimulus Impairment Scale Test (DSIST) determines the perceived visual degradation between two media sources, A and B, by implementing a discrete scale. A viewer must compare the quality of B with respect to A, on a 5-point discrete Absolute Category Rating (ACR) scale, as shown in Fig. 10(b).

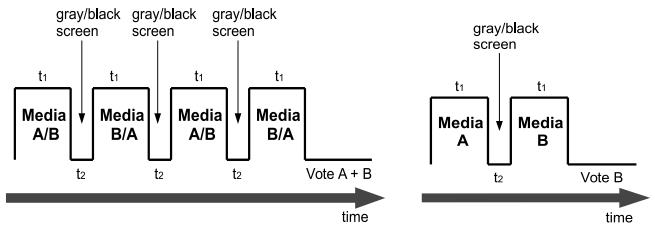
In a subjective evaluation session, firstly, training images are shown to acclimatize viewers to both ACR and DCR scoring systems. In either of the two subjective tests, a higher value in DCR or ACR scales represents a greater perceived visual quality. Fig. 11 illustrates an overall timing diagram for each subjective testing procedure, showing the sequence of tests image display for scoring by the viewers. Note that the media display time, t_1, and blank screen time, t_2, before the change of images, should satisfy the following condition: $t_1 > t_2$.

B. Saliency Model Evaluation

For saliency model evaluation, the Microsoft Research Asia (MSRA) saliency dataset (by Liu *et al.* [80]), popularly used in state-of-the art visual saliency estimation research is used in this work. MSRA saliency datadase provides thousands of publicly available images, from which 1000 are selected to form the MSRA-1000. Subsequent ground truth ROI frames,

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access



(a) (b)

Fig. 11: Stimulus timing diagram for (a) DCR method (b) ACR method.

governed by the outcome of subjective testing, have been manually created as part of the same database. The data test set has been manually labelled by 3 users. The dataset has been narrowed down to 5,000 frames by selecting the most consistent data. Salient portions within each of the 5,000 frames are labelled by 9 users into a binary ground truth map, segmenting the ROI, and the most consistent 1,000 frames make up the MSRA-1000 database, which was used in evaluating the proposed saliency model against the state-of-the-art methodologies. Four state-of-the-art methods representing different approaches are selected in these evaluations. The orthogonal Daubechies-4 (D4) wavelet with 5-level decomposition was chosen for the proposed model's experimental set up.

Fig. 12 shows the saliency model performance, comparing the proposed method against four differing state-of-the-art techniques. Four exemplar original images are shown in *column 1*. *Column 2* demonstrates the performance of the Itti model [28], which portrays moderate saliency estimation, when subjectively compared to the ground truth frames in *column 8*. A drawback to this model is the added computational cost, persisting approximately twice the proposed algorithm. The Rare algorithm [42] is a highly computationally exhaustive procedure to cover both high and low level saliency features by searching for patterns within a frame. A good approximation can be seen from *column 3*, but processing large batches of data would be irrational due to the iterative nature of the algorithm, taking 45 times the proposed model computation time. The Ngau wavelet-based model [33] is shown in *column 4*, but delivers a poor approximation highlighting attentive regions. This model is highly dependant on a plain background with salient regions to remain the same colour or intensity. For images containing a wide variety of intensities and colour, the model breaks down as shown in *row 2*, in Fig. 12, where the white portion within the sea is visually misclassified as an interesting region. *Column 5* and *column 6* show the generated saliency maps from the Erdem model [32] and Li model [30], respectively. The proposed model is shown in *column 7* and identifies any salient activity within in each of the four frames, by locating the presence of intensity and colour contrasts. For example, the proposed method clearly highlights the orange, bird, strawberries and players.

Visual inspection of the saliency model alone does not provide an adequate algorithm evaluation. The Receiver operating characteristics (ROC) considering various threshold values for segmenting the saliency maps are computed for the MSRA-1000 database with respected to the ground truth maps. The ROC plots for the proposed method and the state-of-the-art methods are shown in Fig. 13. The Area Under Curves (AUC) represent the efficient performance of the models. Higher AUC corresponds to better performance. TABLE I *row 1* reports the AUC values for the corresponding ROC plots in Fig. 13. The mean computational time for MSRA-1000 data set images for each of the methods is shown in *row 2* of TABLE I. For a fair comparison of the computational complexity, all algorithms were implemented in MATLAB by the authors and the experiments were performed on the same computer.

TABLE I: AUC and Computational time comparing state-of-the-art image domain saliency models.

Algorithm ⇒	Itti [28]	Rare [42]	Ngau [33]	Erdem [32]	Li [30]	Proposed
ROC AUC for 1000 images	0.875	0.906	0.856	0.878	0.708	0.887
Mean computing time / image (s)	0.281	6.374	0.092	16.540	0.257	0.142

The comparison with the state-of-the-art methods, in terms of the AUC of ROC plots and computational time is shown in Fig. 14. According to the figure, the proposed method is in the top left quadrant of the scatter plot showing the best joint AUC and computational time performance. The proposed saliency models shows superior performance compared with the algorithms proposed by Itti, Ngau, Erdem and Li having an ROC AUC values 1.4%, 3.6%, 1.03% and 25.3% higher than these models, respectively. The Rare model has an ROC AUC 2.1% higher than the proposed, but this is acceptable considering the computational complexity of the context aware algorithm (a 45× speed up in run time is achieved in the proposed model). In fact other than Ngau method, the proposed method achieves significant speed up: 1.98× against Itti method, 116.48× against Erdem method and 1.94× against Li. Additionally, these algorithms are often proposed as stand alone model while the proposed one is regarded as saliency model that is incorporated into a watermarking framework, thus the low computational complexity in the saliency model is very essential.

C. VA-based Watermarking Evaluation

The proposed VA-based watermarking is agnostic to the watermark embedding methodology. Thus, it can be used on any existing watermarking algorithm. In our experiments, we use the non-blind embedding proposed by Xia *et al.* [53] and the blind algorithm proposed by Xie and Arce [54] as our reference algorithms.

The experimental set up for evaluating the proposed watermarking scheme as follows: The MSRA-1000 database used in evaluating the saliency model contains small size images with maximum dimension of 400×400 and is often limited to one close-up salient object. This is not a suitable choice to evaluate any watermarking algorithm. Therefore, the Kodak image test

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

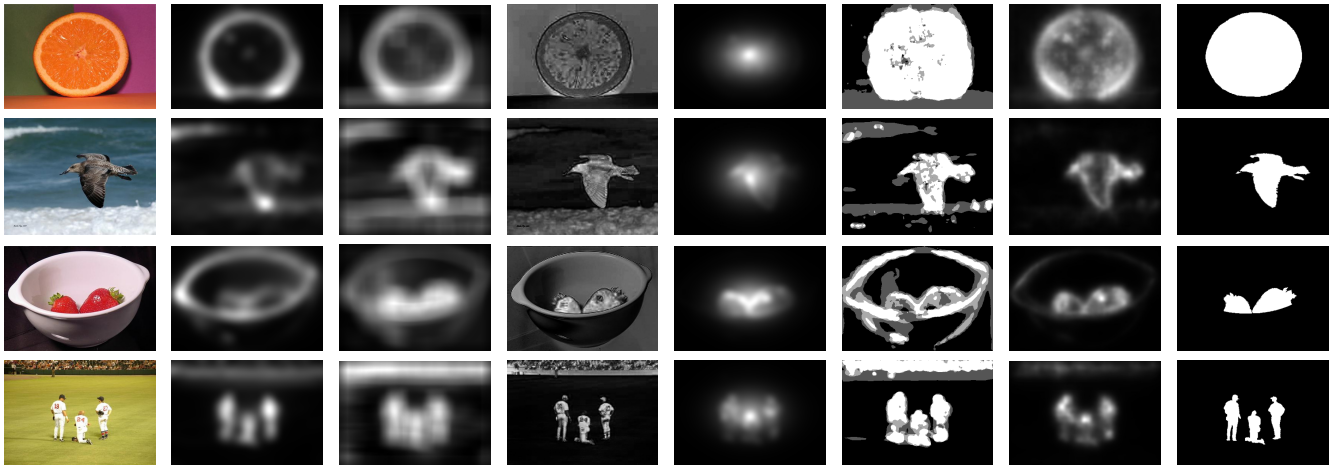


Fig. 12: Image Saliency model state-of-the-art comparison: *Column 1:* Original image from MSRA database, *Column 2:* Itti model [28], *Column 3:* Rare model [42], *Column 4:* Ngau model [33], *Column 5:* Erdem model [32], *Column 6:* Li model [30], *Column 7:* Proposed Method and *Column 8:* Ground Truth.

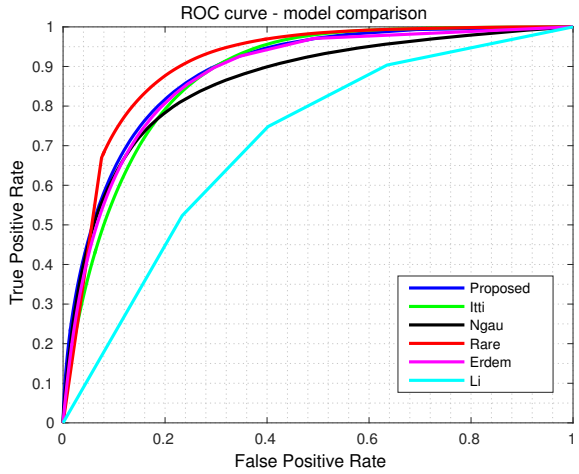


Fig. 13: ROC curve comparing the proposed model with state-of-the-art image domain saliency algorithms: Itti [28], Ngau [33], Rare [42], Erdem [32] and Li [30].

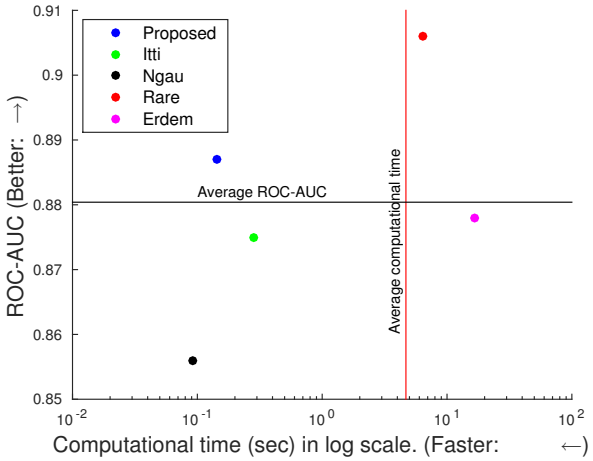


Fig. 14: AUC and Computational time comparing state-of-the-art image domain saliency models: Itti [28], Ngau [33], Rare [42] and Erdem [32].

set[2] containing 24 colour scenes is used for watermarking evaluation in this work.

For the evaluations, we choose all coefficients in a subband and embed the watermarking bit by tuning the strength parameter based on the proposed visual attention model. Therefore our aim is to extract the same bit and hence we use hamming distance metric to evaluate the robustness. For all experimental simulations, common test set parameters for watermark embedding include orthogonal Daubechies-4 (D4) wavelet kernel, embedding at all four subbands at 3rd decomposition level, a binary watermark sequence and $p = 0.75$ as the cumulative frequency threshold for segmenting the saliency maps.

The saliency adaptive strength parameters, α_{min} and α_{max}, are computed using minimum and maximum PSNR values of 35dB and 40dB, respectively as proposed in Eq. (17) and Eq. (18) in § IV-C for non-blind and blind watermarking, respectively.

Throughout this section, four different scenarios are evaluated, with α varying in each instance. The four watermarking scenarios consist of:

1) a uniform α_{min} for the entire image (**Low** strength);
2) the proposed watermarking scheme which implements an adaptive VA-based α (**VAM**);
3) a uniform average watermark strength, α_{ave}, chosen as $\alpha_{ave} = (\alpha_{min} + \alpha_{min})/2$ for the entire image (**Average** strength); and
4) a uniform α_{max} for the entire image (**High** strength).

The experimental evaluation results are consequently shown in the following two sections: embedding distortion (visual quality) and robustness. The imperceptibility of the watermarking schemes are determined by measuring any embedding distortion due to embedding using subjective evaluation as

[2]Available from http://r0k.us/graphics/kodak/

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

well as objective metrics. The former involved 30 human subjects marking their opinions in subjective evaluation test as described in § V-A2. Robustness is evaluated against natural image processing and filtering attacks as implemented by Checkmark [81], and scalable content adaptation by Watermarking Evaluation Bench for Content Adaptation Modes (WEBCAM) [82].

1) Embedding Distortion: Two images accommodating indistinguishable objective metrics, such as, PSNR and SSIM, do not necessarily radiate identical perceived visual quality. To provide a realistic visual quality eavluation, subjective testing is used to analyze the impact of the proposed watermarking scheme on the overall perceived human viewing experience. Subjective evaluation performed in this work comprises of DSCQT and DSIST and the results are shown in Fig. 15, for both blind and non-blind watermarking schemes. The top and bottom rows in Fig. 15 show subjective evaluation results for the blind and non-blind watermarking cases, respectively, whereas the left and right columns in Fig. 15 show the results using DSCQT and DSIST evaluation tools. Consistent results are portrayed for both the blind and non-blind scenarios.

For the DSCQT, the lower the DCR, the better the visual quality, *i.e.*, less embedding distortions. In the shown results, when comparing the proposed and low strength embedding methodologies, the DCR value only deviate by approximately 1 unit in the rating scale suggesting a subjectively similar visual quality. The high strength watermarking scheme shows a high DCR value indicating significantly higher subjective visual quality degradation compared with the VA-based methodology. Similar outcomes are evident from the DSIST plots, where the higher mean opinion score (MOS) on ACR corresponds to better visual quality, *i.e.*, less embedding visual distortions. DSIST plots for low-strength and VA-based schemes show a similar ACR MOS in the range 3-4, whereas the high strength watermark yields an ACR of less than 1. Compared with an average watermark strength, the proposed watermarking scheme shows an improved subjective image quality in all 4 graphs by around 0.5-1 units. As more data is embedded within the visually salient regions, the subjective visual quality of constant average strength watermarked images is worse than the proposed methodology.

For visual inspection, an example of watermark embedding distortion is shown in Fig. 16. The original, the low strength watermarked, VAM-based watermarked and the high strength watermarked images are shown in Fig. 16(a), Fig. 16(b), Fig. 16(c) and Fig. 16(d), respectively, where the distortions around the aircraft propeller and the wing are distinctively visible in high strength watermarking (Fig. 16(d)).

For completion, the objective metrics for embedding distortion evaluation are shown in TABLE II, which display PSNR and SSIM measures for both non-blind and blind watermarking cases, respectively. In both metrics, higher values signify better imperceptibility. From the tables, PSNR improvements of approximately 2dB are achieved when comparing the proposed and constant high strength models. The SSIM measures remain consistent for each scenario, with decrease of 1% for the high strength watermarking model in most cases. The proposed VA-based method successfully exploits visually uninteresting areas to mask extra embedded watermark information, in comparison to the other schemes. From both objective and subject analysis, the proposed VA-based based watermarking has visual quality comparable to low-strength watermarking as only minimal added visual distortion is perceived with respect to that low-strength watermarking. The following section reports the robustness against attacks for the same schemes.

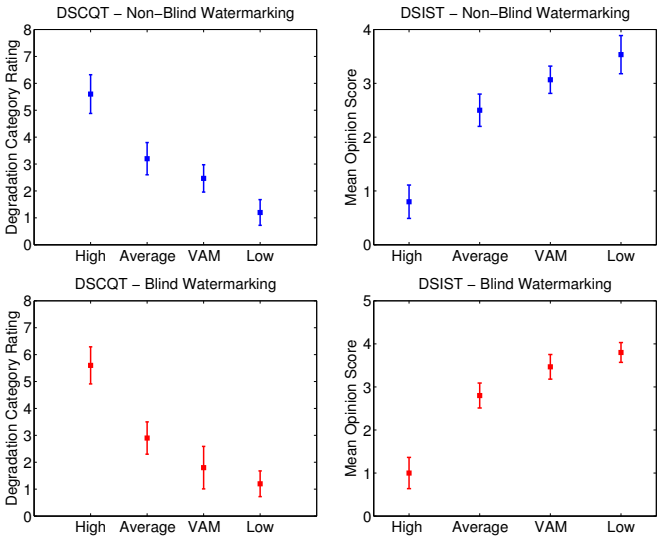


Fig. 15: Subjective Image Watermarking Imperceptibility Testing for the 4 scenarios: uniform Low, Average and High watermarking strengths and the proposed VAM-based adaptive strengths (VAM) for non-blind watermarking Xia *et al.* [53] (top row) blind watermarking Xie and Arce [54] (bottom row).

2) Robustness: The ability of the watermark to withstand intentional and non-intentional adversary attacks are tested and

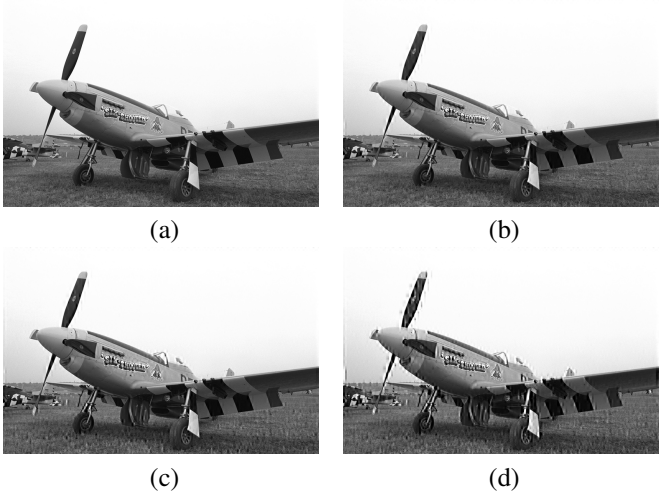


Fig. 16: HL subband watermarking - (a) original image, (b) uniform low strength watermarked image, (c) VAM-based adaptive strength watermarked image, (d) uniform high strength watermarked image.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

TABLE II: PSNR and SSIM values - non-blind and blind watermarking.

	Non-blind Watermarking - Xia *et al.* [53]				**Blind Watermarking** - Xie and Arce [54]			
	Low Strength	Proposed VAM-based	Average Strength	High Strength	Low Strength	Proposed VAM-based	Average Strength	High Strength
Embedding in LL Subband								
PSNR	39.91 ± 0.06	36.07 ± 0.24	37.37 ± 0.07	34.92 ± 0.04	39.93 ± 0.08	37.17 ± 0.26	37.44 ± 0.08	34.94 ± 0.06
SSIM	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
Embedding in HL Subband								
PSNR	39.92 ± 0.07	36.42 ± 0.26	37.28 ± 0.08	34.95 ± 0.06	39.92 ± 0.08	37.21 ± 0.29	37.38 ± 0.09	34.96 ± 0.08
SSIM	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
Embedding in LH Subband								
PSNR	39.90 ± 0.05	36.18 ± 0.28	37.39 ± 0.09	34.94 ± 0.05	39.95 ± 0.07	36.98 ± 0.29	37.35 ± 0.08	34.96 ± 0.08
SSIM	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
Embedding in HH Subband								
PSNR	39.94 ± 0.06	36.42 ± 0.29	37.45 ± 0.08	34.97 ± 0.06	39.96 ± 0.08	37.08 ± 0.31	37.46 ± 0.09	34.96 ± 0.08
SSIM	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00

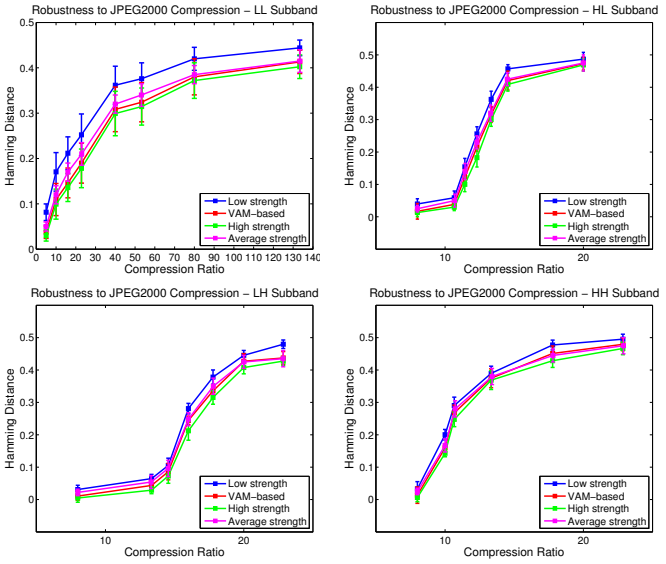


Fig. 17: Robustness to JPEG2000 Compression for the 4 scenarios: uniform Low, Average and High watermarking strengths and the proposed VAM-based adaptive strengths (VAM) for non-blind watermarking Xia *et al.* [53] embedding in LL, LH, HL and HH subbands.

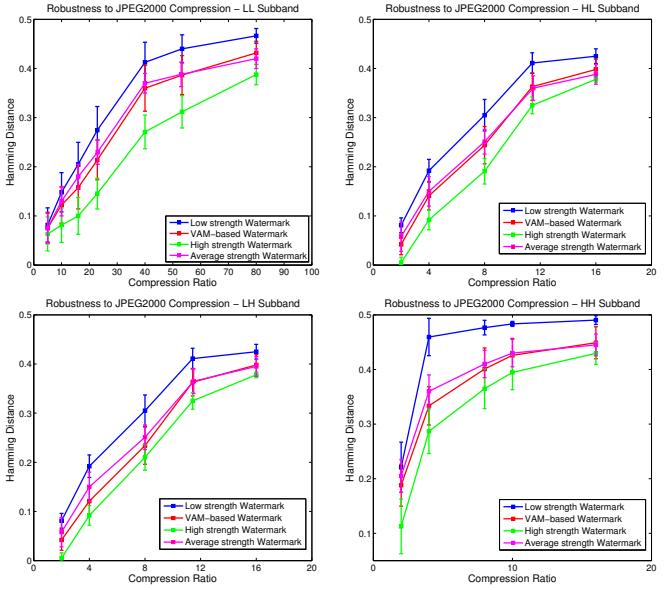


Fig. 18: Robustness to JPEG2000 Compression for the 4 scenarios: uniform Low, Average and High watermarking strengths and the proposed VAM-based adaptive strengths (VAM) for blind watermarking Xie and Arce [54] embedding in LL, LH, HL and HH subbands.

reported here. Robustness against JPEG2000 compression is shown in Fig. 17 and Fig. 18 for the non-blind and blind watermarking schemes, respectively, by plotting Hamming distance (Eq. (12)) of the recovered watermark against the JPEG2000 compression ratio. A smaller value of Hamming distance represents greater robustness. For embedding within each of the LL, HL, LH and HH subbands, up to a 25% improvement in Hamming distance is attainable by implementation of the proposed VA-based watermarking scheme, when compared with the low strength watermark.

Adversary filtering attacks, for each of the three scenarios, are simulated by convoluting the watermarked images with a filtering kernel, to distort any embedded information. TABLE III shows the watermark robustness against various low pass kernel types, namely: a 3×3 and a 5×5 mean filter, a 3×3 and a 5×5 median filter and a 5×5 Gaussian kernel. An increase in watermark robustness, ranging between 10% and 40%, is evident for the proposed method compared to the low strength watermarking, for the various types of kernel. For both filtering attacks and JPEG2000 compression, a maintained or an improvement within watermark robustness is seen in Fig. 17, Fig. 18 and TABLE III for the proposed VA-based technique when compared using an average watermark strength.

As can be seen from the results, the high strength watermark embedding results in high robustness at the expense of low visual quality. However, the proposed VA-based watermark

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

TABLE III: Watermarking robustness against image filtering .

	Non-blind Watermarking - Xia *et al.* [53]				**Blind Watermarking** - Xie and Arce [54]			
Filtering ⇓ Attacks	Low Strength	Proposed VAM-based	Average Strength	High Strength	Low Strength	Proposed VAM-based	Average Strength	High Strength
Embedding in LL Subband								
Gaussian	0.17 ± 0.03	0.13 ± 0.02	0.13 ± 0.02	0.06 ± 0.01	0.21 ± 0.02	0.18 ± 0.02	0.18 ± 0.02	0.15 ± 0.01
3x3 median	0.12 ± 0.03	0.09 ± 0.03	0.09 ± 0.02	0.06 ± 0.02	0.16 ± 0.02	0.11 ± 0.02	0.11 ± 0.02	0.09 ± 0.01
5x5 median	0.22 ± 0.02	0.16 ± 0.02	0.17 ± 0.02	0.07 ± 0.02	0.25 ± 0.04	0.18 ± 0.04	0.19 ± 0.03	0.11 ± 0.03
3x3 mean	0.06 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.03 ± 0.00	0.10 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.04 ± 0.00
5x5 mean	0.18 ± 0.02	0.13 ± 0.02	0.14 ± 0.02	0.06 ± 0.01	0.23 ± 0.02	0.19 ± 0.02	0.19 ± 0.02	0.17 ± 0.01
Embedding in HL Subband								
Gaussian	0.28 ± 0.03	0.24 ± 0.02	0.24 ± 0.02	0.19 ± 0.01	0.28 ± 0.03	0.24 ± 0.02	0.24 ± 0.03	0.19 ± 0.01
3x3 median	0.24 ± 0.02	0.19 ± 0.02	0.20 ± 0.02	0.15 ± 0.01	0.24 ± 0.02	0.19 ± 0.02	0.20 ± 0.02	0.15 ± 0.01
5x5 median	0.29 ± 0.02	0.21 ± 0.02	0.23 ± 0.03	0.17 ± 0.02	0.29 ± 0.02	0.21 ± 0.02	0.21 ± 0.02	0.17 ± 0.02
3x3 mean	0.21 ± 0.01	0.17 ± 0.01	0.17 ± 0.01	0.14 ± 0.01	0.21 ± 0.01	0.17 ± 0.01	0.18 ± 0.01	0.14 ± 0.01
5x5 mean	0.27 ± 0.02	0.22 ± 0.02	0.21 ± 0.02	0.18 ± 0.02	0.27 ± 0.02	0.22 ± 0.02	0.23 ± 0.03	0.18 ± 0.02
Embedding in LH Subband								
Gaussian	0.29 ± 0.02	0.23 ± 0.02	0.23 ± 0.03	0.19 ± 0.01	0.40 ± 0.02	0.35 ± 0.03	0.36 ± 0.02	0.31 ± 0.02
3x3 median	0.24 ± 0.02	0.20 ± 0.02	0.20 ± 0.02	0.14 ± 0.01	0.34 ± 0.01	0.29 ± 0.02	0.30 ± 0.03	0.24 ± 0.01
5x5 median	0.28 ± 0.02	0.21 ± 0.02	0.22 ± 0.02	0.17 ± 0.01	0.38 ± 0.02	0.33 ± 0.03	0.33 ± 0.02	0.29 ± 0.02
3x3 mean	0.22 ± 0.01	0.18 ± 0.01	0.18 ± 0.01	0.14 ± 0.01	0.34 ± 0.01	0.29 ± 0.01	0.29 ± 0.01	0.23 ± 0.01
5x5 mean	0.28 ± 0.02	0.22 ± 0.02	0.23 ± 0.03	0.18 ± 0.02	0.39 ± 0.02	0.36 ± 0.02	0.36 ± 0.02	0.31 ± 0.02
Embedding in HH Subband								
Gaussian	0.38 ± 0.03	0.35 ± 0.03	0.35 ± 0.02	0.32 ± 0.02	0.43 ± 0.02	0.40 ± 0.02	0.40 ± 0.02	0.38 ± 0.01
3x3 median	0.23 ± 0.02	0.22 ± 0.03	0.22 ± 0.01	0.20 ± 0.02	0.28 ± 0.02	0.27 ± 0.03	0.27 ± 0.02	0.25 ± 0.02
5x5 median	0.36 ± 0.03	0.34 ± 0.03	0.34 ± 0.02	0.33 ± 0.02	0.41 ± 0.03	0.39 ± 0.02	0.40 ± 0.02	0.38 ± 0.02
3x3 mean	0.23 ± 0.02	0.21 ± 0.03	0.22 ± 0.01	0.20 ± 0.01	0.30 ± 0.02	0.28 ± 0.03	0.28 ± 0.02	0.26 ± 0.01
5x5 mean	0.38 ± 0.02	0.35 ± 0.04	0.36 ± 0.02	0.34 ± 0.02	0.42 ± 0.03	0.40 ± 0.03	0.40 ± 0.03	0.39 ± 0.02

embedding results in a robustness close to the high strength watermarking scheme, while showing low distortions, as in the low strength watermarking approach. The incurred increase in robustness coupled with high imperceptibility, verified by subjective and objective metrics in § V-C1 and § V-C2, deem the VA-based methodology highly suitable towards providing an efficient watermarking scheme.

VI. Conclusions

In this paper, we have presented a novel wavelet domain visual attention-based framework for robust image watermarking that has minimal or no effect on visual quality due to watermarking. In the proposed scheme, a two-level watermarking weighting parameter map is generated from the VAM saliency map using the proposed saliency model and data is embedded into the host image according to the visual attentiveness of each region. By avoiding higher strength watermarking in visually attentive region, the resulted watermarked image achieved high perceived visual quality while preserving high robustness.

The proposed VAM outperforms all but one existing VA estimation methods by up to 3.6% ROC AUC. However, in terms of run time the proposed model achieved a 45× speed-up compared to the method with the best ROC AUC, thus confirming the suitability for using in the proposed watermarking framework. The proposed low complexity saliency model was extended to propose both blind and non-blind watermarking schemes. ITU-T recommended subjective evaluation was employed to verify the superiority of the proposed VA-based watermarking with respect to high or average strength watermarking and comparability with the low-strength watermarking. For the same embedding distortion, *e.g.*, by fixing PSNR in a narrow window, the proposed VA-based watermarking achieved up to 25% and 40% improvement against JPEG2000 compression and common filtering attacks, respectively, against the existing methodology that does not use the visual attention model. Finally, the proposed VA-based watermarking has resulted in visual quality similar to that of low-strength watermarking and robustness similar to those of high-strength watermarking.

Acknowledgment

We acknowledge the support of the UK Engineering and Physical Research Council (EPSRC), through a Dorothy Hodgkin Postgraduate Award, a Doctoral Training Award at the University of Sheffield, a grant on programmable embedded platform for remote and compute intensive image processing applications (EP/K009931/1) and a Digital Catapult researchers in residence fellowship.

References

[1] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97 – 136, 1980.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

[2] O. Hikosaka, S. Miyauchi, and S. Shimojo, "Orienting of spatial attention-its reflexive, compensatory, and voluntary mechanisms," *Cognitive Brain Research*, vol. 5, no. 1-2, pp. 1–9, 1996.

[3] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar 2001.

[4] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual review of neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.

[5] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analalysis Machine Intelligence*, vol. 35, no. 1, pp. 185–207, Jan. 2013.

[6] M. Carrasco, "Visual attention: The past 25 years," *Vision Research*, vol. 51, no. 13, pp. 1484 – 1525, 2011, vision Research 50th Anniversary Issue: Part 2. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0042698911001544

[7] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Trans. Appl. Percept.*, vol. 7, no. 1, pp. 6:1–6:39, Jan. 2010. [Online]. Available: http://doi.acm.org/10.1145/1658349.1658355

[8] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, July 2011.

[9] S. Frintrop, "General object tracking with a component-based target descriptor," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, May 2010, pp. 4531–4536.

[10] A. Mishra, Y. Aloimonos, and C. Fermuller, "Active segmentation for robotics," in *IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, (IROS 2009)*, Oct 2009, pp. 3133–3139.

[11] N. Jacobson, Y.-L. Lee, V. Mahadevan, N. Vasconcelos, and T. Nguyen, "A novel approach to fruc using discriminant saliency and frame segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2924–2934, Nov 2010.

[12] D. Que, L. Zhang, L. Lu, and L. Shi, "A ROI image watermarking algorithm based on lifting wavelet transform," in *Proc. International Conference on Signal Processing*, vol. 4, 2006, pp. 16–20.

[13] R. Ni and Q. Ruan, "Region of interest watermarking based on fractal dimension," in *Proc. International Conference on Pattern Recognition*, 2006, pp. 934–937.

[14] R. Wang, Q. Cheng, and T. Huang, "Identify regions of interest (ROI) for video watermark embedment with principle component analysis," in *Proc. ACM International Conference on Multimedia*, 2000, pp. 459–461.

[15] C. Yiping, Z. Yin, Z. Sanyuan, and Y. Xiuzi, "Region of interest fragile watermarking for image authentication," in *International Multi-Symposiums on Computer and Computational Sciences (IMSCCS)*, vol. 1, 2006, pp. 726–731.

[16] L. Tian, N. Zheng, J. Xue, C. Li, and X. Wang, "An integrated visual saliency-based watermarking approach for synchronous image authentication and copyright protection," *Image Communication*, vol. 26, no. 8-9, pp. 427–437, Oct. 2011.

[17] H. K. Lee, H. J. Kim, S. G. Kwon, and J. K. Lee, "ROI medical image watermarking using dwt and bit-plane," in *Proc. Asia-Pacific Conference on Communications*, 2005, pp. 512–515.

[18] A. Wakatani, "Digital watermarking for ROI medical images by using compressed signature image," in *Proc. International Conference on System Sciences (2002)*, 2002, pp. 2043–2048.

[19] B. Ma, C. L. Li, Y. H. Wang, and X. Bai, "Salient region detection for biometric watermarking," *Computer Vision for Multimedia Applications: Methods and Solutions*, p. 218, 2011.

[20] A. Sur, S. Sagar, R. Pal, P. Mitra, and J. Mukhopadhyay, "A new image watermarking scheme using saliency based visual attention model," in *India Conference (INDICON), 2009 Annual IEEE*, Dec 2009, pp. 1–4.

[21] J. Shi, Q. Yan, H. Shi, and Y. Wang, "Visual attention based image zero watermark scheme with ensemble similarity," in *Wireless Communications Signal Processing (WCSP), 2013 International Conference on*, Oct 2013, pp. 1–5.

[22] H. G. Koumaras, "Subjective video quality assessment methods for multimedia applications," Geneva, Switzerland, Tech. Rep. ITU-R BT.500-11, April 2008.

[23] M. Oakes, D. Bhowmik, and C. Abhayaratne, "Visual attention-based watermarking," in *IEEE International Symposium on Circuits and Systems (ISCAS),*, 2011, pp. 2653–2656.

[24] K. Koch, J. McLean, R. Segev, M. A. Freed, M. J. Berry, V. Balasubramanian, and P. Sterling, "How much the eye tells the brain," *Current Biology*, vol. 16, no. 14, pp. 1428–1434, 2006.

[25] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Natural Review Neuroscience*, vol. 5, no. 1, pp. 1–7, 2004.

[26] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97 – 136, 1980.

[27] Y. Sun and R. Fisher, "Object-based visual attention for computer vision," *Artificial Intelligence*, vol. 146, no. 1, pp. 77–123, 2003.

[28] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254 –1259, Nov. 1998.

[29] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned Salient Region Detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1597 –1604.

[30] Z. Q. Li, T. Fang, and H. Huo, "A saliency model based on wavelet transform and visual attention," *Science China Information Sciences*, vol. 53, no. 4, pp. 738–751, 2010.

[31] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 2376–2383.

[32] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of Vision*, vol. 13, no. 4, pp. 1–20, 2013.

[33] C. Ngau, L. Ang, and K. Seng, "Bottom-up visual saliency map using wavelet transform domain," in *Proc. IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, vol. 1, July 2010, pp. 692–695.

[34] M. Cerf, J. Harel, W. Einhuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection." in *Advances in Neural Information Processing Systems*, vol. 20, 2007, pp. 241–248.

[35] L. Chen, X. Xie, X. Fan, W. Ma, H. Zhang, and H. Zhou, "A visual attention model for adapting images on small displays," *Multimedia Systems*, vol. 9, no. 4, pp. 353–364, Oct. 2003.

[36] W. J. Won, M. Lee, and J. Son, "Skin color saliency map model," in *Proc. International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, vol. 2, 2009, pp. 1050–1053.

[37] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM International Conference on Multimedia*, 2006, pp. 815–824.

[38] F. Stentiford, "An estimator for visual attention through competitive novelty with application to image compression," in *Proc. Picture Coding Symposium*, Arpil 2001, pp. 101–104.

[39] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, 2007, pp. 545–552.

[40] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[41] G. Kootstra, A. Nederveen, and B. D. Boer, "Paying attention to symmetry," in *Proc. British Machine Vision Conference (BMVC)*, 2008, pp. 1115–1125.

[42] N. Riche, M. Mancas, B. Gosselin, and T. Dutoit, "Rare: a new bottom-up saliency model," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 1–4.

[43] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. USA: Springer, 2002.

[44] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 10, pp. 1183–1194, 2004.

[45] G. Bhatnagar and Q.M. J. Wuand B. Raman, "Robust gray-scale logo watermarking in wavelet domain," *Computers & Electrical Engineering*, 2012.

[46] A. Piper, R. Safavi-Naini, and A. Mertins, "Resolution and quality scalable spread spectrum image watermarking," in *Proc. 7th workshop on Multimedia and Security: MM&Sec'05*, 2005, pp. 79–90.

[47] D. Bhowmik and C. Abhayaratne, "A generalised model for distortion performance analysis of wavelet based watermarking," in *Proc. Int'l Workshop on Digital Watermarking (IWDW '08), Lect. Notes in Comp. Sci. (LNCS)*, vol. 5450, 2008, pp. 363–378.

[48] M. R. Soheili, "Blind Wavelet Based Logo Watermarking Resisting to Cropping," in *Proc. 20th International Conference on Pattern Recognition*, 2010, pp. 1449–1452.

[49] D. Bhowmik and C. Abhayaratne, "Morphological wavelet domain image watermarking," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2007, pp. 2539–2543.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

[50] C. Abhayaratne and D. Bhowmik, "Scalable watermark extraction for real-time authentication of JPEG2000 images," *Journal of Real-Time Image Processing*, vol. 6, no. 4, p. 19 pages, 2011.

[51] D. Bhowmik and C. Abhayaratne, "Quality scalability aware watermarking for visual content," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5158–5172, 2016.

[52] X. C. Feng and Y. Yang, "A new watermarking method based on DWT," in *Proc. Int'l Conf. on Computational Intelligence and Security, Lect. Notes in Comp. Sci. (LNCS)*, vol. 3802, 2005, pp. 1122–1126.

[53] X. Xia, C. G. Boncelet, and G. R. Arce, "Wavelet transform based watermark for digital images," *Optic Express*, vol. 3, no. 12, pp. 497–511, Dec. 1998.

[54] L. Xie and G. R. Arce, "Joint wavelet compression and authentication watermarking," in *Proc. IEEE ICIP*, vol. 2, 1998, pp. 427–431.

[55] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 783–791, May 2001.

[56] D. Kundur and D. Hatzinakos, "Toward robust logo watermarking using multiresolution image fusion principles," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 185–198, Feb. 2004.

[57] C. Jin and J. Peng, "A robust wavelet-based blind digital watermarking algorithm," *International Journal of Information Technology*, vol. 5, no. 2, pp. 358–363, 2006.

[58] R. S. Shekhawat, V. S. Rao, and V. K. Srivastava, "A biorthogonal wavelet transform based robust watermarking scheme," in *Proc. IEEE Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2012, pp. 1–4.

[59] J. R. Kim and Y. S. Moon, "A robust wavelet-based digital watermarking using level-adaptive thresholding," in *Proc. IEEE ICIP*, vol. 2, 1999, pp. 226–230.

[60] S. Marusic, D. B. H. Tay, G. Deng, and P. Marimuthu, "A study of biorthogonal wavelets in digital watermarking," in *Proc. IEEE ICIP*, vol. 3, Sept. 2003, pp. II–463–6.

[61] Z. Zhang and Y. L. Mo, "Embedding strategy of image watermarking in wavelet transform domain," in *Proc. SPIE Image Compression and Encryption Tech.*, vol. 4551-1, 2001, pp. 127–131.

[62] D. Bhowmik and C. Abhayaratne, "On Robustness Against JPEG2000: A Performance Evaluation of Wavelet-Based Watermarking Techniques," *Multimedia Syst.*, vol. 20, no. 2, pp. 239–252, 2014.

[63] ——, "Embedding distortion modeling for non-orthonormal wavelet based watermarking schemes," in *Proc. SPIE Wavelet App. in Industrial Processing VI*, vol. 7248, 2009, p. 72480K (12 Pages).

[64] F. Huo and X. Gao, "A wavelet based image watermarking scheme," in *Proc. IEEE ICIP*, 2006, pp. 2573–2576.

[65] N. Dey, M. Pal, and A. Das, "A session based blind watermarking technique within the nroi of retinal fundus images for authentication using dwt, spread spectrum and harris corner detection," *International Journal of Modern Engineering Research*, vol. 2, pp. 749–757, 2012.

[66] H. A. Abdallah, M. M. Hadhoud, and A. A. Shaalan, "A blind spread spectrum wavelet based image watermarking algorithm," in *Proc. International Conference on Computer Engineering Systems*, 2009, pp. 251–256.

[67] T.-S. Chen, J. Chen, and J.-G. Chen, "A simple and efficient watermarking technique based on JPEG2000 codec," in *Proc. Int'l Symp. on Multimedia Software Eng.*, 2003, pp. 80–87.

[68] H.A.Abdallah, M. M. Hadhoud, A. A. Shaalan, and F. E. A. El-samie, "Blind wavelet-based image watermarking," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 1, pp. 358–363, March 2011.

[69] H. Wilson, "Psychophysical models of spatial vision and hyper-acuity," *Spatial Vision*, vol. 10, pp. 64–81, 1991.

[70] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit, "A study of parameters affecting visual saliency assessment," *Computing Research Repository*, vol. 1307, 2013.

[71] Q. Gong and H. Shen, "Toward blind logo watermarking in JPEG-compressed images," in *Proc. Int'l Conf. on Parallel and Distributed Comp., Appl. and Tech., (PDCAT)*, 2005, pp. 1058–1062.

[72] V. Saxena, M. Gupta, and D. T. Gupta, "A wavelet-based watermarking scheme for color images," *The IUP Journal of Telecommunications*, vol. 5, no. 2, pp. 56–66, Oct. 2013.

[73] C. Jin and J. Peng, "A robust wavelet-based blind digital watermarking algorithm," *Information Technology Journal*, vol. 5, no. 2, pp. 358–363, 2006.

[74] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Proc. IEEE ICASSP*, vol. 5, 1998, pp. 2969–2972.

[75] V. S. Verma and J. R. Kumar, "Improved watermarking technique based on significant difference of lifting wavelet coefficients," *Signal, Image and Video Processing*, pp. 1–8, 2014.

[76] P. Meerwald, "Quantization watermarking in the JPEG2000 coding pipeline," in *Proc. Int'l Working Conf. on Comms. and Multimedia Security*, 2001, pp. 69–79.

[77] D. Aggarwal and K. S. Dhindsa, "Effect of embedding watermark on compression of the digital images," *Computing Research Repository*, vol. 1002, 2010.

[78] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

[79] A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," in *American Institute of Aeronautics and Astronautics (AIAA)*, vol. 9, 1993, pp. 286–291.

[80] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[81] S. Pereira, S. Voloshynovskiy, M. Madueno, S. M.-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Proc. Int'l. Information Hiding Workshop, Lect. Notes in Comp. Sci. (LNCS)*, vol. 2137, 2001, pp. 340–353.

[82] D. Bhowmik and C. Abhayaratne, "A framework for evaluating wavelet based watermarking for scalable coded digital item adaptation attacks," in *SPIE Wavelet Applications in Industrial Processing VI*, vol. 7248, 2009, pp. 1–10.

Deepayan Bhowmik received the B.E. in Electronics Engineering from Visvesvaraya National Institute of Technology, Nagpur, India in 2003, the MSc. in Electronics and Information Technology from Sheffield Hallam University, UK in 2006 and the Ph.D. from the University of Sheffield, UK in 2011. Previously he worked as a research associate at Heriot-Watt University, Edinburgh & the University of Sheffield, UK and a system engineer in ABB Ltd., India. Currently he is working as a Lecturer in Computing at Sheffield Hallam University, Sheffield, UK. Dr. Bhowmik received UK Engineering and Physical Sciences Research Council (EPSRC) Dorothy Hodgkin Postgraduate Award (DHPA) in 2006-2010 and a Digital Catapult-EPSRC fellowship in 2016-2017 to conduct research on multimedia security. His research interests include multimedia forensics and security, computer vision and embedded image processing hardware.

Matthew Oakes graduated with a Masters in Electronic and Electrical Engineering from the University of Sheffield in 2009. In 2014 he obtained a Ph.D. in Visual Information Engineering also at the University of Sheffield. Matthew is currently working at the University of Buckingham as a Knowledge Transfer Partnership Associate in joint project between the Applied Computing Department at Buckingham and Deepnet Security Ltd in Bletchley Park. His main expertise lies in image / video processing, compression, digital watermarking and visual saliency estimation. Matthewss current research includes biometric security systems and machine learning.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2627241, IEEE Access

Charith Abhayaratne received the B.E. degree in electrical and electronic engineering from the University of Adelaide, Australia, in 1998, and the Ph.D. degree in electronic and electrical engineering from the Universityof Bath, U.K., in 2002. He is currently a Lecturer with the Department of Electronic and Electrical Engineering, The University of Sheffield, United Kingdom. His research interests include video and image compression, watermarking, image and video analysis, multidimensional signal processing, graph spectral analysis and computer vision. He was a recipient of the European Research Consortium for Informatics and Mathematics (ERCIM) Post-Doctoral Fellowship (2002-2004) to carry out research at the Centre of Mathematics and Computer Science (CWI), The Netherlands, and the National Research Institute for Computer Science and Control (INRIA), Sophia Antipolis, France. He has contributed to ISO/IEC Moving Picture Expert Group (MPEG) on scalable video coding standardization and wavelet video exploration (2004-2006). He has been a Reviewer for leading IEEE/IET/ACM/SPIE journals and conferences, U.K. Engineering and Physical Science Research Council (EPSRC), European Commission Framework Programme, Hong Kong Research Council, and book publishers. He has been involved in technical program committees for several leading international conferences and served as a Session Chair and a Guest Editor in journal special issues. He is the United Kingdoms Liaison Officer for the European Association for Signal Processing (EURASIP).

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.