

The  
University  
Of  
Sheffield.

**ATTENTION DRIVEN  
SOLUTIONS FOR ROBUST  
DIGITAL WATERMARKING  
WITHIN MEDIA**

submitted by

**Matthew Oakes**

for the degree of

Doctor of Philosophy

of the

Department of Electronic and Electrical Engineering  
The University of Sheffield

September, 2014





## COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author.

This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signature of Author .....

Matthew Oakes



# ABSTRACT

As digital technologies have dramatically expanded within the last decade, content recognition now plays a major role within the control of media. Of the current recent systems available, digital watermarking provides a robust maintainable solution to enhance media security. The two main properties of digital watermarking, imperceptibility and robustness, are complimentary to each other but by employing visual attention based mechanisms within the watermarking framework, highly robust watermarking solutions are obtainable while also maintaining high media quality. This thesis firstly provides suitable bottom-up saliency models for raw image and video. The image and video saliency algorithms are estimated directly from within the wavelet domain for enhanced compatibility with the watermarking framework. By combining colour, orientation and intensity contrasts for the image model and globally compensated object motion in the video model, novel wavelet-based visual saliency algorithms are provided. The work extends these saliency models into a unique visual attention-based watermarking scheme by increasing the watermark weighting parameter within visually uninteresting regions. An increased watermark robustness, up to 40%, against various filtering attacks, JPEG2000 and H.264/AVC compression is obtained while maintaining the media quality, verified by various objective and subjective evaluation tools. As most video sequences are stored in an encoded format, this thesis studies watermarking schemes within the compressed domain. Firstly, the work provides a compressed domain saliency model formulated directly within the HEVC codec, utilizing various coding decisions such as block partition size, residual magnitude, intra frame angular prediction mode and motion vector difference magnitude. Large computational savings, of 50% or greater, are obtained compared with existing methodologies, as the saliency maps are generated from partially decoded bitstreams. Finally, the saliency maps formulated within the compressed HEVC domain are studied within the watermarking framework. A joint encoder and a frame domain watermarking scheme are both proposed by embedding data into the quantised transform residual data or wavelet coefficients, respectively, which exhibit low visual salience.



*Dedicated to my Grandmother Alice  
Grayson, Grandfather Alfred Oakes  
and Aunty Marion Schindler who all  
sadly passed away during this research  
study. You will always be remembered,  
RIP....*



## ACKNOWLEDGEMENTS

First and foremost i would like to express my sincere gratitude to my supervisor, Dr. Charith Abhayaratne, as without his positive surveillance and guidance this PhD would not have been possible. His influence and hard-working ethics has helped provoke many novel research ideas, consequently i feel extremely fortunate to have been his student. My gratitude also goes to the EPSRC Doctoral Training Award, which has helped fund this work. My special thanks are extended to everyone in the VIE lab and my good friend Dr. Deepayan Bhowmik for his support. I am deeply grateful to my loving girlfriend for her support and patience. Finally i would like to thank my parents for their loving support and care as without them this thesis would not have been possible.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Aims and Objectives . . . . .	2
1.2	Contribution . . . . .	3
1.3	Thesis Outline . . . . .	4
<b>2</b>	<b>Literature Survey</b>	<b>7</b>
2.1	Image Watermarking . . . . .	7
2.1.1	Parameter Selection . . . . .	7
2.2	Video Watermarking . . . . .	9
2.2.1	Frame by Frame Domain Algorithms . . . . .	9
2.2.2	3D Frequency Domain Algorithms . . . . .	9
2.2.3	Compressed Video Codec Domain Algorithms . . . . .	10
2.2.3.1	Option 1) RAW Frame Watermarking . . . . .	10
2.2.3.2	Option 2) Residual Watermarking . . . . .	11

2.2.3.3	Option 3) Motion Vector-based Watermarking . . .	11
2.2.3.4	Option 4) Bitstream domain Watermarking . . .	12
2.3	Image Saliency Models . . . . .	12
2.4	Video Saliency Models . . . . .	14
2.5	Visual Attention and Watermarking . . . . .	15
2.5.1	Saliency-based Watermark Coefficient Selection . . . . .	16
2.5.2	State-of-the-Art Visual Attention-based Watermarking . . .	16
2.6	Conclusions . . . . .	18
<b>3</b>	<b>Background Overview</b>	<b>19</b>
3.1	Digital watermarking . . . . .	19
3.1.1	Definition, Properties and Applications . . . . .	19
3.1.2	Watermarking Process . . . . .	20
3.1.2.1	Watermark Embedding . . . . .	20
3.1.2.2	Extraction and Authentication . . . . .	22
3.1.3	Wavelet-based Algorithms . . . . .	23
3.1.3.1	Non-blind Watermarking . . . . .	24
3.1.3.2	Blind Watermarking . . . . .	24
3.1.4	Watermark Attacks . . . . .	25

3.2	The Visual Attention Model . . . . .	27
3.2.1	Approach to Visual Saliency . . . . .	27
3.2.2	Visual Saliency Model Evaluation . . . . .	28
3.2.3	Applications of Visual Saliency . . . . .	29
3.3	HEVC Codec . . . . .	31
3.3.1	Coding Structure . . . . .	32
3.3.2	Block Structure . . . . .	33
3.3.3	Intra Coding . . . . .	35
3.3.4	Inter Coding . . . . .	37
	3.3.4.1 Motion Compensation . . . . .	37
	3.3.4.2 Advanced Motion Vector Prediction (AMVP) . . . . .	38
3.4	Evaluation Tools and Datasets . . . . .	39
3.4.1	Subjective and Objective Evaluation Tools for Visual Quality . . . . .	39
3.4.2	Experimental Datasets . . . . .	42
	3.4.2.1 Image Datasets . . . . .	42
	3.4.2.2 Video Datasets . . . . .	43
3.5	Conclusions . . . . .	44
<b>4</b>	<b>Visual Saliency Estimation</b>	<b>45</b>

4.1	Introduction . . . . .	45
4.2	Image Domain Saliency . . . . .	46
4.2.1	Generating Scale Feature Maps . . . . .	46
4.2.2	Saliency Map Generation . . . . .	47
4.2.3	Results . . . . .	49
4.3	Video Domain Saliency . . . . .	56
4.3.1	2D+t Wavelet Domain . . . . .	56
4.3.2	Temporal Saliency Feature Map . . . . .	57
4.3.3	Global Motion Compensated Frame Difference . . . . .	59
4.3.4	Spatial-Temporal Saliency Map Combination . . . . .	60
4.3.5	Results . . . . .	61
4.4	Conclusions . . . . .	64
<b>5</b>	<b>Visual Attention-based Watermarking</b>	<b>69</b>
5.1	VA-based Image Watermarking . . . . .	69
5.1.1	Saliency Map Thresholding . . . . .	70
5.1.2	Watermark Embedding Strength Calculation . . . . .	73
5.1.3	Saliency Map Reconstruction . . . . .	74
5.1.4	System Architecture . . . . .	75

5.1.5	Experiments, Results and Analysis . . . . .	76
5.1.5.1	VA-based Experimental Image Data Sets . . . . .	77
5.1.5.2	Embedding Distortion . . . . .	77
5.1.5.3	Robustness . . . . .	81
5.2	VA-based Video Watermarking . . . . .	83
5.2.1	Experimental Results . . . . .	85
5.2.1.1	Imperceptibility . . . . .	87
5.2.1.2	Robustness . . . . .	89
5.3	Conclusions . . . . .	91
<b>6</b>	<b>HEVC Domain Saliency Estimation</b>	<b>93</b>
6.1	Introduction . . . . .	93
6.2	Intra-Frame Saliency Estimation . . . . .	95
6.2.1	Block Size . . . . .	95
6.2.2	Intra mode differences . . . . .	96
6.2.3	Residual Data . . . . .	99
6.2.4	Intra Saliency Map Generation . . . . .	99
6.3	Inter-Frame Saliency Estimation . . . . .	104
6.3.1	Block Size . . . . .	105

6.3.2	Motion Residual Data . . . . .	107
6.3.3	Motion Vector Difference Magnitude . . . . .	108
6.3.4	Inter Saliency Map Generation . . . . .	112
6.4	Combined HEVC Saliency Model . . . . .	113
6.5	Encoder Setup . . . . .	113
6.6	Experimental Results . . . . .	116
6.6.1	Intra-Frame Saliency . . . . .	116
6.6.2	Inter-Frame Saliency . . . . .	119
6.7	Conclusions . . . . .	122
<b>7</b>	<b>HEVC domain Watermarking</b>	<b>125</b>
7.1	HEVC Watermarking Approaches . . . . .	126
7.1.1	Frame Domain Watermarking . . . . .	126
7.1.2	Compressed Domain Watermarking . . . . .	127
7.1.3	Joint Encoder Watermark . . . . .	128
7.2	Transform Coefficient Watermarking Criteria . . . . .	128
7.3	Proposed Watermarking Scheme . . . . .	130
7.3.1	Watermark Embedding . . . . .	130
7.3.2	Watermark Detection . . . . .	133

7.4	Experimental Results . . . . .	133
7.4.1	Frame Domain Watermarking Experimental Results . . . .	134
7.4.1.1	Imperceptibility . . . . .	134
7.4.1.2	Robustness . . . . .	135
7.4.2	Joint Encoder Watermarking Experimental Results . . . .	135
7.4.2.1	Imperceptibility . . . . .	136
7.4.2.2	Robustness . . . . .	138
7.5	Conclusions . . . . .	139
<b>8</b>	<b>Conclusions and Future Work</b>	<b>143</b>
8.1	Conclusions . . . . .	143
8.2	Key Contributions . . . . .	145
8.3	Future Work . . . . .	145
<b>9</b>	<b>Appendix - Additional Results</b>	<b>149</b>
	<b>List of Figures</b>	<b>157</b>
	<b>List of Tables</b>	<b>167</b>
	<b>List of Symbols and Acronyms</b>	<b>169</b>
	<b>References</b>	<b>173</b>





# Chapter 1

## Introduction

With high fluctuations in multimedia usage over the last decade, a greater focus is given towards watermarking as a solution towards digital content protection. Consequently, the demand to efficiently compress data also peaked as recent years has seen the emergence of numerous video coding standards such as MPEG-2 [1], H.264 Advanced Video Coding (AVC) [2] and more recently High Efficiency Video Coding (HEVC) [3]. To combat digital security, watermarking trends need to be constantly upgraded to stay on top of the current coding standards and data formats.

Areas of visual interest stimulate neural nerve cells, creating human gaze to fixate towards a particular scene area. The Visual Attention Model (VAM) highlights these visually sensitive regions, which stimulate a neural response within the primary visual cortex. Whether that neural vitalization be from contrast in intensity, a distinctive face, unorthodox motion or a dominant colour, these stimulative regions diverge human attention providing highly useful saliency maps within the media processing domain.

This thesis focuses on incorporating the VAM within watermarking methodology to provide a robust system. Visual saliency models and watermarking schemes are provided within the image, raw video and HEVC domain.

## 1.1 Aims and Objectives

The main aim of this thesis is to provide a watermarking solution for visual media protection while satisfying joint robustness and imperceptibility requirements. By incorporating human visual mechanics within the watermarking framework, we can potentially increase overall limitations, such as watermark capacity, restricting the original watermarking methodology.

In this thesis, Visual Attention (VA)-based watermarking is proposed as a solution toward joint robust and imperceptible media protection. Due to increased complexity, most available saliency models are applicable only within the image domain, compared with their video domain saliency model counterparts. Therefore, our first objective was to formulate a suitable video saliency model for future use within the video watermarking framework. The suggested watermarking scheme, which schematically incorporates visual saliency mechanisms, can be summarised by the following set of objectives:

1. **Wavelet-based saliency estimation** - To provide both an image and video saliency model, consisting of spatial and temporal salient features, within the wavelet domain. The algorithms should be highly suitable for watermarking, but not limited solely toward this application.
2. **VA-based watermarking** - To devise a novel VA based watermarking scheme in both image and video domains. Evaluation should include comparison towards traditional watermarking approaches in terms of overall robustness and imperceptibility.

2010-2013 has been a key stepping stone within multimedia advancement due to development of the HEVC standard, with a scalable [4] and 3D codec extension [5] imminently expected to accommodate the high multi-platform media demand. This innovative state-of-the-art standard is an essential component to support upcoming immersive viewing experience services such as: Ultra-High Definition (HD) broadcasting (4K, 8K), HD-3DTV, mobile technology and digital cinema.

HEVC provides computationally intensive encoding to attain extremely high com-

pression rates. Within the scope of this thesis, we aim to exploit key new features within the HEVC codec design to efficiently create a unique visual saliency and watermarking scheme. The following HEVC domain objectives are set:

3. **HEVC domain saliency** - To provide a saliency model within the HEVC domain exploiting new features of the video codec.
4. **HEVC domain watermarking** - To evaluate traditional watermarking approaches within HEVC and provide a unique VA-based watermarking scheme, incorporating the previous HEVC domain saliency maps.

## 1.2 Contribution

The novel research contributions from distinct stages of this thesis have been published within various respected conference proceedings. Demonstrations are also provided from the following website link to facilitate understanding of the work.

### Software and demo's

**D1)** A demonstration of video saliency estimation can be found at:

*<http://svc.group.shef.ac.uk/va-video.html>*

**D2)** A demonstration of VA-based watermarking can be found at:

*<http://svc.group.shef.ac.uk/va-video-wm.html>*

**D3)** A demonstration of compressed domain saliency estimation can be found at:

*<http://svc.group.shef.ac.uk/hevc-va.html>*

**D4)** A demonstration of compressed domain VA-based watermarking can be found at:

*<http://svc.group.shef.ac.uk/hevc-va-wm.html>*

## Author Publications

**C4)** **M. Oakes** and C. Abhayaratne, A New Saliency Model Using Intra Coded High Efficiency Video Coding (HEVC) Frames, in Proc. 20th International Conference on Multimedia Modelling (MMM 2014), 8-10 January 2014, Dublin, Ireland. (13-pages long paper)

**C3)** **M. Oakes** and C. Abhayaratne, Visual Saliency Estimation for Video, in Proc. 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2012), 23-25 May 2011, Dublin, Ireland, pp. 1-4

**C2)** **M. Oakes**, D. Bhowmik and C. Abhayaratne, Visual Attention-Based Watermarking, in Proc. IEEE International Symposium on Circuits and Systems (ISCAS 2012), 15-18 May 2011, Rio de Janeiro, Brazil, pp. 2653-2656

**C1)** D. Bhowmik, C. Abhayaratne and **M. Oakes**, Robustness Analysis of Blind Watermarking for Quality Scalable Image Compression, in Proc. European Signal Processing Conference (EUSIPCO 2010), 23-27 August 2010, Aalborg, Denmark, pp. 810-814

## 1.3 Thesis Outline

The remainder of the thesis is structured into seven different chapters, the contents of which are summarised as follows:

**Chapter 2** reviews state-of-the-art literature analysis for digital watermarking and visual saliency models. This includes wavelet and non-wavelet based image watermarking techniques, frame domain, compressed domain and uncompressed domain video watermarking methodologies and both image and video based VA models. Finally, current proposals towards visually attentive watermarking are illustrated.

**Chapter 3** provides a brief overview into digital watermarking, the VAM and the HEVC codec. An insight into the properties, application and process of

digital watermarking is described as well as an outline of watermark attacks and evaluation performance. A background overview of the VAM depicts how conspicuous regions arise, the saliency detection process and the applications of VA. The HEVC codec is introduced and the improvements from the H.264 predecessor are characterised. A brief codec overview is introduced with a detailed description of intra and inter coding.

**Chapter 4** provides a new saliency estimation algorithm for both image and video which highlights visually attentive regions within a sequence. Static and Temporal models are combined by exploiting contrasts within each of the Y, U and V channels in the wavelet domain with local object motion determined from motion compensated temporal differences.

**Chapter 5** proposes a novel image and video watermarking scheme based upon the VAM. The new saliency algorithm, proposed in the previous chapter, is extended into the watermarking domain to help determine coefficient selection and watermark strength. The objective of the new scheme is to maintain subjective visual media quality but increase the overall watermark robustness.

**Chapter 6** proposes HEVC domain saliency estimation by exploiting key features within the codec. Saliency models for both intra-only coded frames and sequences containing P and B frames are provided. Attributes considered within the model include variable block size, intra angular prediction and advanced motion vector prediction.

**Chapter 7** proposes 2 new HEVC VA-based watermarking schemes, partly constituting from the saliency model derived in the previous chapter. The first is a frame domain wavelet based watermarking scheme, utilising the compressed domain saliency estimation. The second is a novel joint encoder technique, embedding data within the quantised transform coefficients. The objective of both proposals is to achieve a high watermark robustness while maintaining the media visual quality.

**Chapter 8** provides a summary of conclusions outlining the novel work contribution within thesis, i.e., VA-based watermarking, HEVC domain saliency estimation, etc. Possible areas for future study within this field are suggested.



# Chapter 2

## Literature Survey

This chapter comprises an analysis of the state-of-the-art, confronting existing approaches. The study is divided into 3 notable sections for both image and video domains: digital watermarking, VA estimation and VA-based watermarking.

### 2.1 Image Watermarking

Frequency-based watermarking, more precisely wavelet domain watermarking methodologies are highly favoured in the current research era. Firstly, the wavelet domain is compliant within many image and video coding 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 (DCT for example), the Discrete Wavelet Transform (DWT) provides an ideal choice for robust watermarking.

#### 2.1.1 Parameter Selection

When designing a novel watermarking scheme there are numerous features to consider: wavelet kernel, embedding coefficients and wavelet subband selection.

Each of these particular features can sufficiently impact the overall watermark characteristics and are usually largely dependant upon the target application requirements.

### **Wavelet Kernel Selection**

An appropriate choice of wavelet kernel must be determined within the watermarking framework. Studies have been performed to show the performance of watermark robustness and imperceptibility, dependant upon wavelet transform [6–8]. The orthogonal Daubechie wavelets are a favourable choice with many early watermarking schemes [9–14], although the later introduction of bi-orthogonal wavelets, within the field of digital watermarking, has increased in popularity [15–18].

### **Host Coefficient Selection**

Suitable transform coefficients need to be chosen for embedding a watermark for which various approaches exist. Coefficient selection can be determined by thresholding values based upon the coefficient magnitude [16], a pixel masking approach based upon the HVS [12], the median of 3 coefficients in a 3x1 overlapping window [11] or simply by selecting all the coefficients [9, 10, 13].

### **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 [9, 10, 12, 13, 19] can often provide great imperceptibility but with limited watermark robustness capabilities. Contradictory schemes embed data only within the low frequency subbands [11, 14, 18] aimed towards providing a high robustness. Spread spectrum domain embedding [16, 20–22] 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, contributing to the system output. Previous studies have researched watermarking schemes using only two [9, 10, 19], three [17, 18, 23] and four or more [12–14] wavelet decomposition levels.



## 2.2 Video Watermarking

In Section 2.1, state-of-the-art image domain watermarking models are discussed. Video domain watermarking is not just a simple expansion of image domain schemes, as temporal contributions from the video sequence must be considered within the watermarking framework. Video watermarking algorithms can be categorised into frame-by-frame, 3D domain and compressed video codec domain algorithms.

### 2.2.1 Frame by Frame Domain Algorithms

Frame by frame domain video watermarking is a simple extension of the schemes in Section 2.1. The image domain watermarking schemes are applied directly into each frame within the video sequence. Various literature is accessible for frame-by-frame video watermarking schemes [24–27]. A clear disadvantage is that no temporal consideration is accountable. More efficient methods are available as presented in Section 2.2.2 and Section 2.2.3. Frame-by-frame watermarking algorithms are highly fragile to temporal attacks such as collusion and frame dropping as described in Section 3.1.4.

### 2.2.2 3D Frequency Domain Algorithms

A 1D temporal transform is incorporated into the 2D spatial transform design to provide a platform for 3D watermarking. The additional 1D transform can capture any temporal redundancies from motion within the video and is highly advantageous within the field of video watermarking. Popular watermarking schemes have been researched within the 3D DWT [28–31], although literature also suggests watermarking within the 3D Discrete Fourier Transform (DFT) [32, 33] and 3D DCT [28, 34] domain. The 3D domain surmounts many of the temporal robustness issues related with frame-by-frame video watermarking in Section 2.2.1, however it is not without limitations. 3D transform domain watermarking algorithms can suffer from temporal artifact flicker and can be fragile to video com-

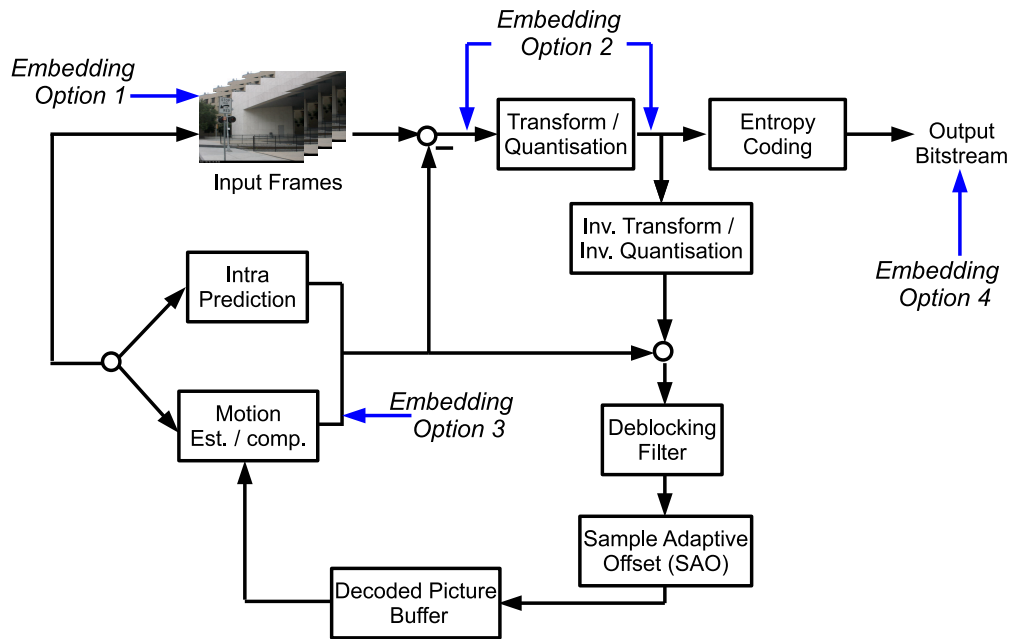


Figure 2.1: Video codec watermark possibilities.

pression, especially codec quantisation heavily dependant upon the motion. Literature suggests incorporating Motion Compensated Temporal Filtering (MCTF) within the 3D watermarking schemes [35] to reduce any potential motion flicker.

### 2.2.3 Compressed Video Codec Domain Algorithms

Video watermarking schemes can be applied directly within a video codec. There are numerous possibilities to embed data within the coding process, each with their own distinct advantages. Figure 2.1 shows an HEVC encoder with 4 possible watermark embedding options in the: 1) raw frame, 2) residual data, 3) MVs and 4) bitstream.

#### 2.2.3.1 Option 1) RAW Frame Watermarking

Raw frame domain algorithms embed watermark data into the input video sequence prior to video encoding. Typical methodologies are entailed in Section 2.2.1 and Section 2.2.2. The main limitations are the watermarking scheme

has to survive full video coding compression providing robustness problems. Further, it is more computationally efficient to provide a watermarking scheme directly from within the compressed domain by options 2) - 4), rather than as a preprocess to video encoding.

### **2.2.3.2 Option 2) Residual Watermarking**

Residual watermarking is achieved by embedding data within the prediction errors. Embedding can be performed within the transformed quantized coefficients or prior to residual transformation, in the exact prediction block residuals. Naturally researchers prefer residual transform domain embedding [36–40] to ensure watermarked data does not undergo lossy quantisation within the encoding, which could distort or remove embedded data. Residual watermarking can provide a highly imperceptible watermarking scheme, however with limited robustness capabilities. Due to advancements within codec prediction scheme accuracy [3], less margin for residual watermarking becomes available. Both overall robustness and imperceptibility can be increased by selecting particular blocks with non zero DC coefficients [36, 40] or embedding only within specific AC transform residual coefficients [36].

### **2.2.3.3 Option 3) Motion Vector-based Watermarking**

Inter predicted blocks can be watermarked by modifying the MVs. Literature is available describing MV watermarking algorithms [41–44]. MVs are usually encoded with high priority so most the majority of MV-based watermarking schemes are highly robust. However, modifying the motion within the frame is highly perceptible so schemes must be extremely cautious not to distort the overall video sequence. Significant improvements in visual quality are attainable by limiting which MVs to modify, based on their magnitude [42], angular phase [43] or texture present [44] within the frame. A major limitation of MV based watermarking is the high fragility to video reformatting.

#### 2.2.3.4 Option 4) Bitstream domain Watermarking

The watermark can be embedded directly into the bitstream or partially decoded bitstream data. It is highly complex and there is very limited room for embedding directly in the bitstream so entropy decoding is usually performed to allow residual or MV watermarking within the partially decoded bitstreams. Bitstream domain algorithms [45–48] are usually computationally economical and suitable for real-time application. The main associated problem with bitstream domain algorithms is watermark drift. Any embedded information in the bitstream is seen as an error which propagates throughout the decoder and causes distortion when predicting the intra and inter blocks. Bitstream domain watermarking schemes must consider drift compensation within their design [46].

## 2.3 Image Saliency Models

Our eyes receive vast streams of visual information every second (108-109 bits) [49]. 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 [50] and feature integration theory [51] 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 [52–56]. Salient objects are not size specific therefore Multi-Resolution Analysis (MRA) is adopted within many models [52,54,57,58]. Classical low level computational saliency model framework [54] is shown in Figure 2.2. The Itti framework [54] adopts gabor filters and an RGBY colour space. This is combined with a center-surround 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.

Later studies incorporate high level features within the low level saliency design

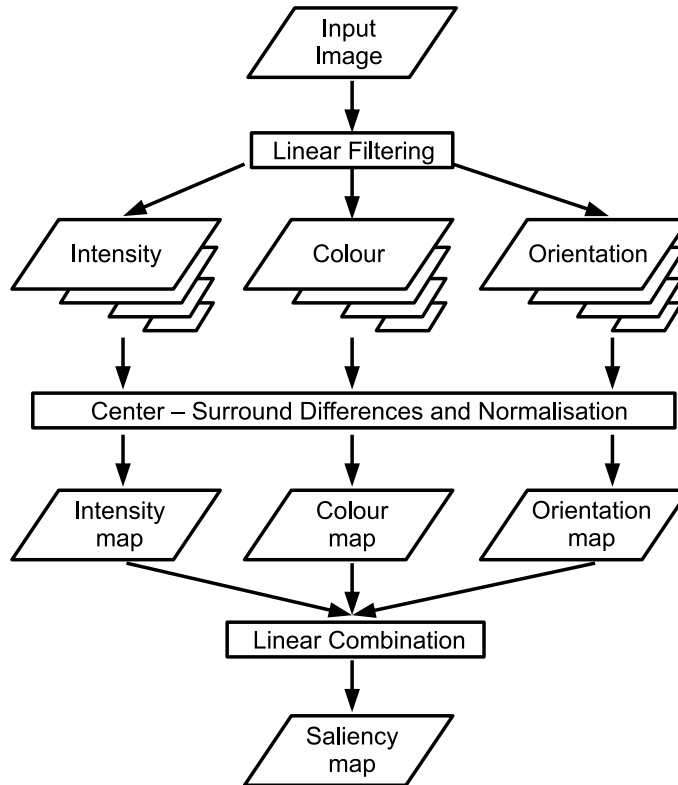


Figure 2.2: Classical feature based visual attention model structure for imagery or static scenery.

such as face detection [59], text detection [60] and skin detection [61]. A major advantage of these high and low level feature models is the simplicity to incorporate additional features, within the existing framework in Figure 2.2, combined with a linear feature weightage, dependant upon the application. However, the main drawback lies within the computational complexity as MRA the approach generates many processable feature maps for combination. Various other proposed techniques can detect attentive scene regions by histogram analysis [62], locating inconsistencies within neighboring pixels [63], object patch detection [58], graph analysis [64], log-spectrum analysis [65] and symmetry [66].

Available models can be broadly categorised into a bottom-up or top-down approach, many of which are documented within the survey paper from Borji [56]. Image driven bottom-up models [54, 57, 67, 68] are dependant solely upon information within the scene where as knowledge driven top-down models [59, 61]

depend upon prior scene knowledge upon distinguishable features. Most existing approaches adopt a bottom-up architecture. The RARE model [67] combines both colour and orientation features, deduced from multiresolution gabor filtering. A rarity mechanism is implemented to estimate how likely a region is to be salient, by histogram analysis. Erdem [68] adopts classical architecture, as used within the Itti model [54], to segment intensity, colour and orientation contrasts. However, nonlinear 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 salience exhibit high dissimilarity to their neighboring regions in terms of their covariance representations based on intensity, colour and orientation.

This research requires an efficient saliency model for direct integrability within the watermarking framework. Previous wavelet domain approaches either provide insufficient model performance as are based on coefficient average variance [69] or require multiple frame resizing prior to saliency estimation [57]. The Ngau model [69] estimates visual salience by locating coefficients which diverge greatly from the local mean within the LL wavelet subband. This procedure is performed across both Luma and chroma channels to provide a fast wavelet based estimation of visual salience.

## 2.4 Video Saliency Models

Seldom work has been directed towards video saliency estimation, in comparison to the image domain counterpart, as temporal feature consideration dramatically increases the overall VA framework complexity. Most typical video saliency estimation methodologies [52, 70, 71] [72–76] exist as a supplementary extension from their image domain algorithms, as shown in Figure 2.3. As with the image domain feature integration models, video saliency algorithms must decipher through large amounts of redundant information, across each feature map, in order attain beneficial data.

Research estimating VA within video can also be derived from exploiting spa-

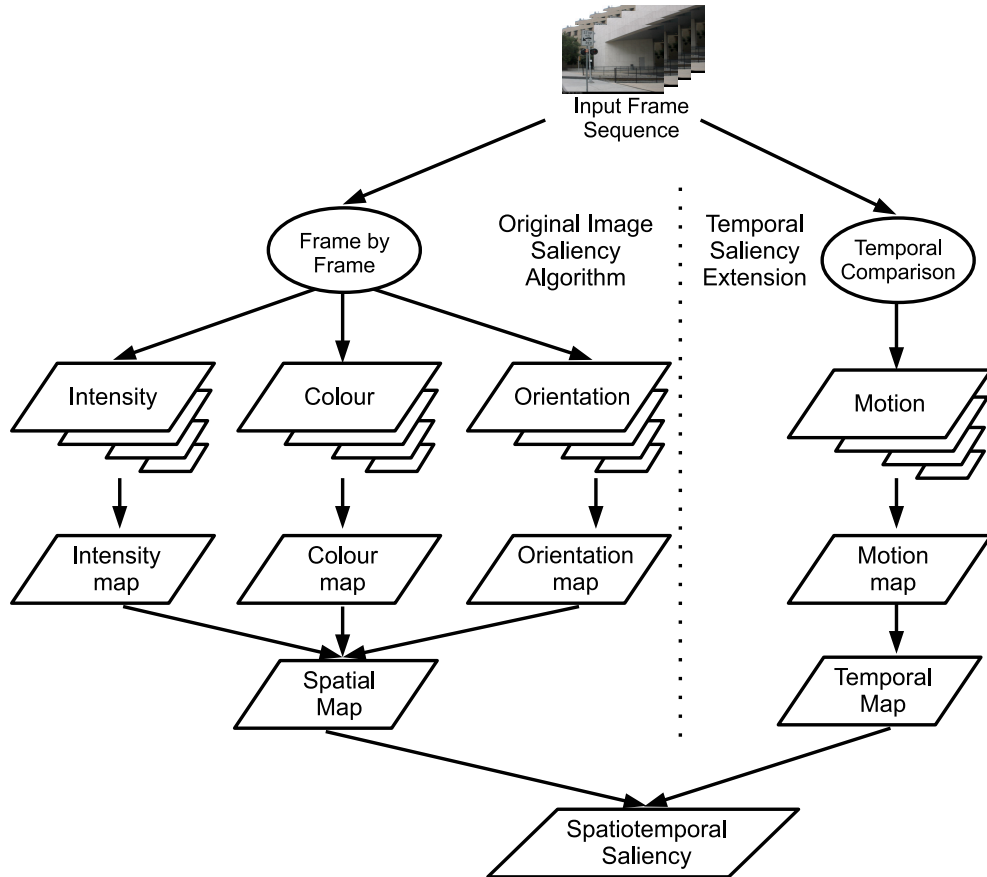


Figure 2.3: Classical feature-based visual attention model structure for video.

tiotemporal cues [77], structural tensors [78] and optical flow [79]. As with the image domain saliency model analysis in Section 2.3, this thesis requires an efficient wavelet-based saliency model which is suitable for smooth incorporation within the video watermarking framework.

## 2.5 Visual Attention and Watermarking

This section describes the link between visual saliency and the watermarking domain. Current VA-based watermarking methodologies are also described, highlighting close related works to content within this thesis.

### 2.5.1 Saliency-based Watermark Coefficient Selection

Digital watermarking is a classical tradeoff between watermark robustness and scene imperceptibility where as the VAM highlights conspicuous regions viewer attention is drawn towards. By employing VA mechanics within the digital watermarking framework, an increased overall robustness against various adversary attacks is possible, while subjectively limiting any perceived visible artifacts by the human eye. This thesis provides solutions towards VA-based watermarking where inattentive scene areas determine the most appropriate coefficients to embed within a media source.

### 2.5.2 State-of-the-Art Visual Attention-based Watermarking

Research incorporating VA mechanisms within the watermarking framework documents increased embedding strength within the scene ROI [80–87]. The motivation is to protect the key identifiable frame features contained within a scene, providing a high robustness toward cropping. The main drawbacks to this VA-based solution are:

- Increasing the watermark strength within eye catching frame regions is perceptually dangerous as human attention will naturally be drawn towards any additional embedding artifacts.
- In a video sequence the ROI can shift throughout the entire sequence, hence efficient video cropping is problematic and very uncommon. This undermines the motivation of the proposed methodology.
- Scenes exhibiting sparse saliency will potentially contain extensively fragile or no watermark data.

The closest related work to this thesis is described within the literature in [81]. A wavelet based watermark is embedded directly into the most salient frame portions, while the rest of the frame is ignored. The watermarking scheme is



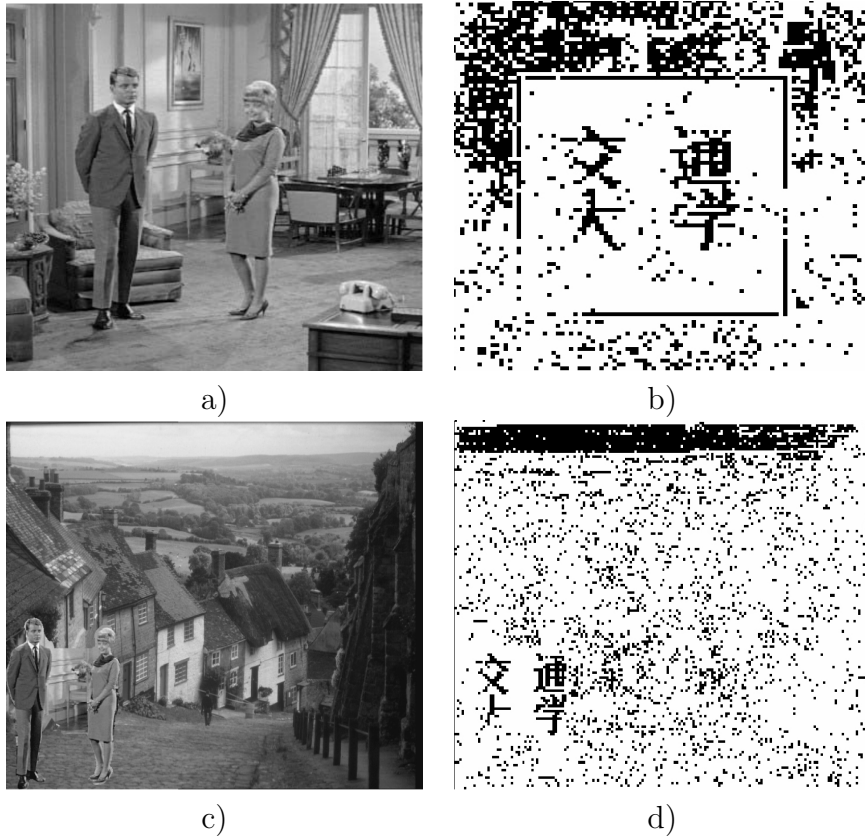


Figure 2.4: Demonstrating the effect of VA within the watermarking framework. a) Original ROI-based watermarked frame b) Extracted watermark c) Patched frame d) Extracted watermark from patched frame.

designed to provide a high robustness against cropping and patching attacks. However, only small portions of the frame are digitally protected, limiting the overall usefulness of the watermarking scheme. Figure 2.4 shows the results of an extracted watermark in the original and patched frame. The top row shows the original frame whereas the bottom row shows a different frame with the patched ROI from the top row.

Other current VA-based watermarking methodologies do provide limited usefulness dependant upon the application. Medical imagery requires a high resolution ROI enclosed by a homogeneous surrounding background, for which literature provides saliency-based watermarking schemes [85,86]. ROI-based watermarking is also useful to provide robustness against patching attacks [81], as described in Section 3.1.4.

## 2.6 Conclusions

In this chapter, existing state-of-the-art watermarking and saliency models were discussed for both image and video. Proposals documenting wavelet-based, compressed domain and uncompressed domain watermarking algorithms are debated. Further examined are notable image and video domain VA-based research advancements. The literature into VA-based watermarking methodology is discussed incorporating a high strength embedding strength with the scene ROI, without consideration toward subjective imperceptibility. Within the scope of this thesis, the aim is to provide robust VA-based digital watermarking solutions for image, uncompressed video and compressed domain video sequences. By incorporating VA-driven coefficient selection within the watermarking framework, there is substantial potential to increase the overall robustness of existing watermarking systems without compromising the subjective media visual quality.

# Chapter 3

## Background Overview

Digital watermarking, the VAM and HEVC codec are three main components of this thesis. This chapter presents an overview of digital watermarking and VA, describing possible real life application scenarios. An insight into the domain of the HEVC codec, is also provided, due to its relevance within this thesis.

### 3.1 Digital watermarking

#### 3.1.1 Definition, Properties and Applications

Digital watermarking is defined as the copyright or author identification information which is embedded directly into digital media ensuring imperceptibility, robustness and security. As digital technology has dramatically expanded within the last decade, copyright protection is now an essential part of many multimedia based companies to establish ownership identification on digital products. When designing a watermarking scheme, numerous properties must be considered which are usually defined by the application requirements. Table 3.1 lists and describes various common watermark properties.

Watermark embedding can be performed within both the pixel or transform do-

Table 3.1: General watermark properties requirements.

<b>Property</b>	<b>Brief Description</b>
Imperceptible	No visible traces of watermark artifacts should be present which cause distortion to the media.
Robust	The watermark must be reliably detectable after an adversaries intentional (or non-intentional) attacks.
Payload Capacity	The watermark payload capacity is the amount of information present within the watermarked media. Embedded systems should consider and limit potential data-rate increases.
Efficient Performance	In many applications, real-time processing is a key requirement [88]. Highly efficient algorithms are required, especially when combatting media sources streamed live.

main. Embedding data implementing frequency decomposition is a highly popular choice, as this characterises human eye perception of the media [89]. Frequency domain watermarking can potentially provide a greater imperceptibility and robustness compared with spatial domain approaches. Watermarking algorithms are highly advantageous towards various applications and range across a variety of disciplines, which are summarised within Table 3.2.

### 3.1.2 Watermarking Process

Digital watermarking consists of 2 processes: 1) Watermark Embedding and 2) Extraction and Authentication. Each of these techniques are described in further detail.

#### 3.1.2.1 Watermark Embedding

The embedding procedure inserts watermark information, within the host media, modifying all or chosen pixels (spatial domain embedding); or coefficients (frequency domain embedding), to ensure a joint robust and imperceptible tradeoff.

Table 3.2: Applications of digital watermarking.

Application	Brief Description
Access control	Determine the viewing control for applications such as pay television.
Broadcast monitoring	Tracking whenever a specific video is being broadcasted.
Copy control	Using watermarking to disable illegal copying of the media content.
Media transaction tracking	Record the receipt of a media transaction so if the content is illegally distributed it can be traced back to the buyer.
Medical	Authenticate private digital medical documents such as MRI scans, blood test and urine test results.
Owner identification	An owner can determine a legitimate claim to the media in question.
Personal documentation authentication	Authentication of financial documents such as banking, insurance etc
Video server host authentication	Control illegally uploaded content available from video hosting websites.

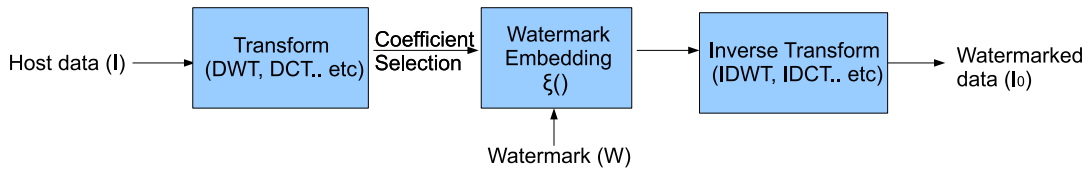


Figure 3.1: Watermark embedding procedure.

This can be expressed in elementary form as:

$$I_0 = \xi(I, W), \quad (3.1)$$

where  $I$  and  $I_0$  are the host and watermarked media source, respectively,  $W$  is the watermark information and  $\xi()$  is the embedding function. The embedding function can further be categorized into sub-processes: 1) forward transform (for frequency domain), 2) pixel / coefficient selection, 3) embedding method (additive, multiplicative, quantization etc.) and 4) inverse transform, as portrayed in Figure 3.1. Naturally, stages 2) and 4) are omitted within spatial domain watermarking.

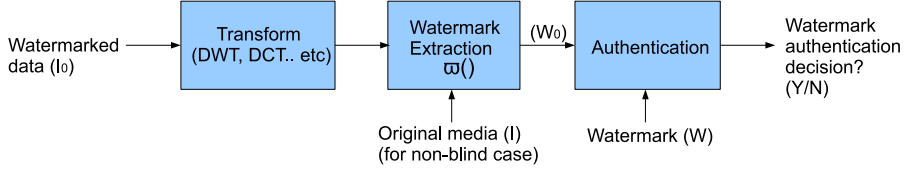


Figure 3.2: Watermark extraction and authentication procedure.

### 3.1.2.2 Extraction and Authentication

Suggested within the process name, this consists of two subprocesses, namely: 1) watermark extraction and 2) authentication of the extracted watermark. By using a similar input parameter set, the watermark extraction procedure follows a reverse of the embedding algorithm. Based upon watermark extraction criteria, any watermarking method can be categorized within either: 1) non-blind type or 2) blind type. For the first category, a copy of the original raw un-watermarked media is required for extraction, whereas blind watermark extraction transpires directly from the test image itself. The watermark extraction process can be written, in a simplified form as:

$$W_0 = \varpi(I_0, I), \quad (3.2)$$

where  $W_0$  is the extracted watermark and  $\varpi()$  is the extraction function. 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 using the following equation:

$$H(W, W_0) = \frac{1}{L} \sum_{i=0}^{L-1} W \oplus W_0, \quad (3.3)$$

where  $L$  represents the length of the watermark sequence and  $\oplus$  is the *XOR* logical operation between the respective bits. A complete overall system diagram of extraction and authentication process is shown in Figure 3.2. As with the watermark embedding procedure, shown in Figure 3.1, the transform stage is omitted within spatial domain watermarking.

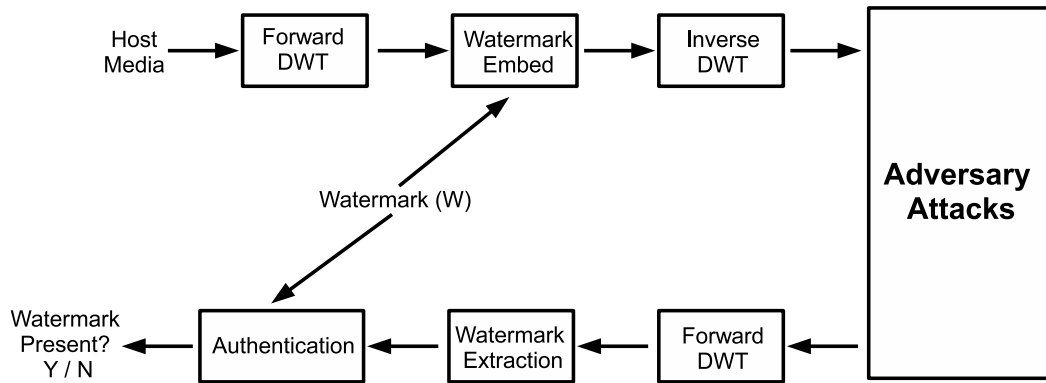


Figure 3.3: General wavelet-based watermarking framework.

### 3.1.3 Wavelet-based Algorithms

Among the available watermarking schemes, spread spectrum domain watermarking, especially the DWT domain algorithms, is a popular choice due to its joint spatial and frequency decomposition characteristics. Many wavelet domain watermarking algorithms are available [9, 11–13, 16, 19, 90–93] which offer robust performance against various attacks including filtering and natural image processing. Every Watermarking scheme falls into one of two classes namely: blind or non-blind, dependant on whether the original work is available at the decoder for watermark extraction. Both watermarking cases are described in more detail in Section 3.1.3.1 and Section 3.1.3.2.

Figure 3.3 shows the basic framework for a wavelet-based watermarking scheme. A forward DWT is applied toward the host media before watermark data is embedded within the selected subband coefficients. An Inverse Discrete Wavelet Transform (IDWT) concludes the watermark embedding procedure. A wide variety of potential adversary attacks, as described in Section 3.1.4, can occur in an attempt to distort or remove any embedded watermark information. The extraction operation is performed after a forward DWT. The extracted watermark is compared to the original embedded sequence before an authentication decision verifies the watermark presence.

### 3.1.3.1 Non-blind Watermarking

Magnitude-based additive watermarking [10, 12, 16, 93–96] is a popular choice for many people when using a non-blind watermarking system, due to its simplicity. Wavelet coefficients are modified dependant upon the watermark weighting parameter,  $\alpha$ , the magnitude of the original coefficient,  $C_{(m,n)}$  and the binary watermark information,  $W_{(m,n)}$ . The watermarked coefficients,  $C'_{(m,n)}$ , are a function of  $C_{(m,n)}$  described in Equation (3.4):

$$C'_{(m,n)} = C_{(m,n)} + \alpha\omega_{(m,n)}C_{(m,n)}. \quad (3.4)$$

$W_{(m,n)}$  is derived from a pseudorandom binary sequence,  $b \in (1, 0)$  using weighting parameters  $W_1$  and  $W_2$  (where  $W_2 > W_1$ ) which are assigned in Equation (3.5) as follows:

$$W_{(m,n)} = \begin{cases} W_2 & \text{if } b \in (1, 0) = 1 \\ W_1 & \text{if } b \in (1, 0) = 0. \end{cases} \quad (3.5)$$

To obtain the extracted watermark,  $w'_{(m,n)}$ , Equation (3.4) is rearranged in Equation (3.6):

$$w'_{(m,n)} = \frac{C'_{(m,n)} - C_{(m,n)}}{\alpha C_{(m,n)}}. \quad (3.6)$$

Since the non-watermarked coefficients  $C_{(m,n)}$  are needed for comparison, this results in non-blind extraction. A threshold limit of  $T_w = \frac{w_1 + w_2}{2}$  is used to determine whether a 1 or 0 was originally embedded as described by Equation (3.7):

$$b \in (1, 0) = \begin{cases} 1 & \text{if } W_{(m,n)} > T_w \\ 0 & \text{if } W_{(m,n)} < T_w. \end{cases} \quad (3.7)$$

### 3.1.3.2 Blind Watermarking

Quantization-based watermarking, [11, 19, 90, 95, 97], is a blind scheme which relies on modifying various coefficients towards a specific quantization step. As proposed in [11], the algorithm is based on modifying the median coefficient towards the step size,  $\delta$ , by using a running non-overlapping 3x1 window. The altered coefficient must remain the median value, of the three coefficients within the window, after the modification. The equation calculating  $\delta$  is described in



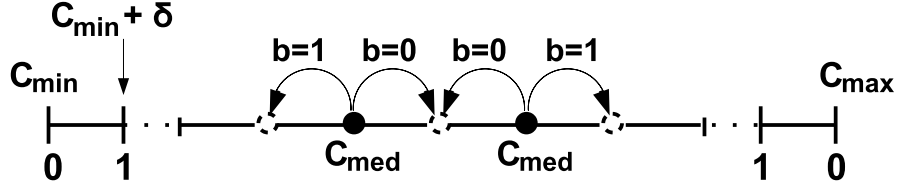


Figure 3.4: Blind quantisation-based coefficient embedding.

Equation (3.8):

$$\delta = \alpha \frac{(C_{min}) + (C_{max})}{2}, \quad (3.8)$$

where  $C_{min}$  and  $C_{max}$  are the minimum and maximum coefficients, respectively. The median coefficient,  $C_{med}$ , is quantised toward the nearest step, dependant upon  $b \in (1, 0)$ . Quantisation-based watermark embedding is shown in Figure 3.4. The watermarked bit,  $W_{ext}$ , for a particular window position, is extracted by the condition in Equation (3.9):

$$W_{ext} \in (0, 1) = \left[ \frac{C_{max} - C_{med}}{\delta} \right] \% 2 \quad (3.9)$$

where  $\%$  denotes the modulo operator to detect an odd or even number and  $C_{med}$  is the median coefficient value within the 3x1 window.

### 3.1.4 Watermark Attacks

Numerous conceivable adversaries attacks are possible, which attempt to distort, tamper or remove the watermark from copyrighted material. Any process which can distort the media watermark is classified as a watermark attack. We have categorised potential attacks under the following headings: additive, filtering, geometric, compression, editing, temporal and other. A full table detailing feasible watermark attacks is shown in Table 3.3 and Table 3.4.

Table 3.3: Digital watermark attacks 1.

Category	Method	Brief Description	Ref
Additive Attack	<ul style="list-style-type: none"> <li>- random noise addition</li> <li>- multiple watermark</li> </ul>	<p>Distort the media by modifying the magnitude of random coefficients. Usually achieved by Gaussian or salt and pepper noise.</p> <p>Add own watermark ontop of the existing watermark to claim ownership of the media.</p>	<ul style="list-style-type: none"> <li>- [98] [99] [100] [101]</li> <li>- [102] [100]</li> </ul>
Filtering Attack	<ul style="list-style-type: none"> <li>- mean</li> <li>- median</li> <li>- gaussian</li> <li>- denoising</li> <li>- deblurring</li> </ul>	<p>Apply a mean, median, gaussian, denoising or deblurring filter, which can be applied by using various kernel sizes and differing parameter values. Filtering attacks are performed by the convolution of an NxN sized kernel over each scene.</p>	<ul style="list-style-type: none"> <li>- [99] [101]</li> </ul>
Geometric Attack	<ul style="list-style-type: none"> <li>- Cropping</li> <li>- Rotation</li> <li>- Scaling</li> <li>- Translation</li> <li>- Shearing and bending</li> <li>- Row and column removal</li> </ul>	<p>The aim of a geometric attack is to lose spatial synchronisation of the watermark, within the data. Cropping attacks 'chop off' unwanted regions within the media and then convert the resolution of the media back to its original dimensions. For rotation attacks, the entire media is pivoted around the center point by a small offset amount of degrees. Scaling affects the media resolution and aspect ratio. (similar to cropping) A translation is achieved by moving every point a constant distance, in a specified direction and zero padding the open gaps. Shearing and bending occurs by interpolating the media to an entirely different shape. For row and column removal attack, random rows and columns are removed before interpolating the signal back to the original resolution.</p>	<ul style="list-style-type: none"> <li>- [103] [104] [99] [101]</li> <li>- [103] [105] [99] [101]</li> <li>- [103] [104] [105] [99] [100]</li> <li>- [103] [105]</li> <li>- [103]</li> <li>- [104]</li> </ul>
Compression	<ul style="list-style-type: none"> <li>- Lossy compression</li> <li>- Data Reformatting</li> </ul>	<ul style="list-style-type: none"> <li>- Compression attacks can non-intentionally remove embedded watermark data as lossy compression can modify coefficients to ensure highly efficient bit-rates.</li> <li>- Recompressing data into another format can also distort embedded information. (E.g JPEG2000, HEVC, H.264, MPEG-2)</li> </ul>	<ul style="list-style-type: none"> <li>- [105] [100]</li> <li>- [105] [100]</li> </ul>
Editing Attack	<ul style="list-style-type: none"> <li>- Border Modification</li> <li>- Feature Addition</li> <li>- Patching</li> </ul>	<p>Editing attacks are popular within media broadcasting sites such as Youtube. The pixels surrounding the frame edge are modified in border modification. Feature addition incorporates many techniques including adding subtitles and annotations. This usually non-intentional watermark attack is very common, especially with user-friendly commercially available software such as Photoshop. Patching comprises of pasting a cropped portion from one scene into another.</p>	<ul style="list-style-type: none"> <li>- [106]</li> </ul>

Table 3.4: Digital watermark attacks 2.

Category	Method	Brief Description	Ref
Temporal Attack	- Frame order changing	Temporal watermark attacks occur within video watermarking systems. The main aim is to destroy any temporal synchronisation between the watermark and media. Frame reordering (i.e swap every 8th and 9th frame) is the first conventional method to distort the watermark. Frame dropping removes selective individual frames from the media and can entirely eliminate any watermark information on those frames. Frame averaging attempts to remove any data protection information by calculating the mean between selected consecutive frames.	- [105] [100]
	- Frame dropping		- [98] [103] [104] [105] [100] [101]
	- Frame averaging		- [98] [105] [100] [101]
Other Attacks	- Gamma correction	- Gamma correction transforms the luminance by a non-linear equation to increase the overall contrast.	- [101]
	- Histogram equalisation	- Histogram equalisation enhances the frame contrast. The intensity values are spread over the maximum range by the collective cumulative frequency.	- [101]
	- Sharpening	- Sharpening media enhances the overall contrast between light and dark features.	- [101]
	- media recapturing	- Recapturing of the video using an external device is a major issue within multimedia security today. (i.e recording cinema film with a hand held camera) This has many problems associated with it, such as accidental geometric attacks.	- [107]
	- colour space conversion	- Values become clipped and rounded when converting colour-space. (e.g YUV to RGB)	- [103]
	- Mosaic Attack	- Mosaic attacks are performed by breaking up the entire watermarked media in many smaller components and reassembling.	- [102]
	- Collusion Attack	- Collect several frames and compare, using statistical averaging, to remove the watermark from the frames.	- [100]
	- Dithering	- Dithering is an intentionally applied form of noise used to randomly quantise the media. It has the illusion that more colours are present from a limited quantised frame.	- [98]

## 3.2 The Visual Attention Model

### 3.2.1 Approach to Visual Saliency

The Human Visual System (HVS) has a limited processing capacity which enforces competitive neural representation. When viewing a typical scene, the human or primate visual cortex must select specific regions to focus upon. Areas of visual interest stimulate nerve cells creating human gaze to fixate towards a

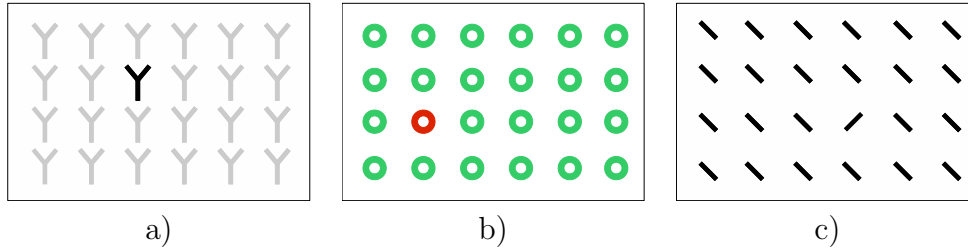


Figure 3.5: Visually stimulating features individually defined by a) Intensity contrast b) Colour contrast c) Orientation contrast.

particular region. This stimulation can correspond from unorthodox motion in a video sequence or prominent objects within static imagery. An object is classified as visually salient if it stands out from its surroundings and diverges human attention.

Computational modeling of the visual cortex commenced from early feature integration studies [51]. The theory states VA is derived from a parallel combination of components which excite neural stimuli. These visually stimulating features include intensity, colour and orientation contrast as shown in Figure 3.5. Various methodologies have been previously proposed to accurately model the human visual cortex for both image and video [74], many of which are described within Chapter 2.

### 3.2.2 Visual Saliency Model Evaluation

Saliency models are most suitably evaluated via subjective analysis, however this can be a sluggish procedure. The use of objective metrics are also important, providing a fast model evaluation.

Subsequent ground truth Region of Interest (ROI) frames, governed by the outcome of intensive visual testing [108], are manually created, from which Receiver Operating Characteristic (ROC) curves determine model accuracy [109]. The Area Under the ROC Curve (AUC) evaluates model performance, as shown in Figure 3.6. An AUC of 1 represents a perfect coherent saliency, whereas an area of 0.5 represents a worthless saliency map. The main drawback of evaluating

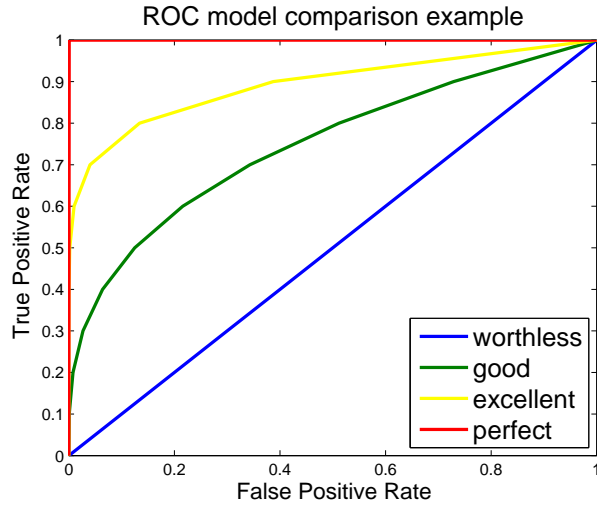


Figure 3.6: ROC curve example.

saliency models with an objective metric is generating accurate ground truth frames, which rely upon subjective assessment to classify salient regions into binary segments. Computational complexity of the saliency model is gauged by calculating the average scene processing time to generate a saliency map.

### 3.2.3 Applications of Visual Saliency

Visual saliency estimation is a highly powerful tool, providing great benefits towards applications such as watermarking, automatic frame resizing, auto-collage, compression and frame summarisation.

The VAM can be incorporated within the watermarking framework by embedding larger amounts of data within visually unattractive frame regions. Inattentive regions are less likely to catch human gaze, therefore the system provides a subjective increase in watermark imperceptibility. Figure 3.7 shows how incorporating VA within the watermarking framework can dramatically decrease overall embedding distortion. Further information is located within Chapter 5, which solely focuses upon VA-based watermarking. By employing saliency based frame compression, coefficients located far from conspicuous areas receive greater quantization, compared with the visually attentive counterparts. By preserving frame quality within salient locations, it is possible to dramatically increase the overall



Figure 3.7: Demonstrating the effect of VA within the watermarking framework. a) Watermark embedding without incorporating VA b) Watermark embedding incorporating VA.



Figure 3.8: Demonstrating the effect of VA-based compression. a) Compressed frame without VA consideration b) Compressed frame in coherence with VAM.

subjective frame quality, as shown in Figure 3.8. A visually appealing collage can be automatically formed as an output from various input images, summarising the overall theme depicted within the various frames. This is demonstrated in Figure 3.9 and further information can be found within the literature [110]. Figure 3.10 demonstrates automatic frame resizing in coherence with the VAM. This methodology is especially applicable to devices requiring a limited frame resolution, such as mobile phones, so any media content can be viewed across multiple platforms with optimal visual quality [111] [60]. Regions which require cropping will be selected from visually uninteresting frame areas. VA can be incorporated within individual images or an entire video sequence, to provide a video summarisation. By segmenting important aspects into a smaller contained abstract, key features within the media content are automatically highlighted. There are many other applications for visual saliency models across a wide variety fields including: advertising [112], medical research [113] and media



Figure 3.9: A collage, automatically formed from multiple salient frame regions.



Figure 3.10: Automatic frame cropping and resizing a) Original frame b) Automatically resized/cropped in coherence with VAM.

superresolution [114].

### 3.3 HEVC Codec

There is an increasing need to compress data efficiently as the demand for higher quality media distribution rises. H.264/AVC was released in 2003 [115], so advancements within video codec structures are required to cope with the expansion of high definition storage. The Joint Collaborative Team on Video Coding (JCT-VC) are currently working on a state of the art video codec system, HEVC, with imminent plans to finalise the standard. [3] The HEVC codec increases the video coding gain by the implementation of content adaptive prediction schemes. With the current HEVC Test Model under Consideration (TMuC) it is possible to acquire the same video quality as obtained with H.264/AVC and achieve bitrate



Figure 3.11: HEVC comparison with H.264/AVC predecessor.

savings of up to 50% [3], as shown in Figure 3.11.

The overall structure of HEVC bears resemblance to classical hybrid video codec architecture, however, there are numerous improvements compared with the H.264/AVC predecessor. New key features include an Advanced Motion Vector Prediction (AMVP), an enhanced Context-Adaptive Binary Arithmetic Coding (CABAC) and an improved intra coding scheme. HEVC is devised to cover a wide range of functionality for video content. Applications include but are not limited to: video camcorders, digital cinema, HD transmissions, internet streaming, medical imagery, mobile data streaming, real-time interactive conversational services (videoconferencing, facetime, videophone, etc.), video surveillance capture, multimedia storage (blu-ray, digital recorder for video, etc.) and wireless displays. Figure 3.12 shows a diagram of a typical HEVC encoder.

### 3.3.1 Coding Structure

HEVC employs a variable length Group of Pictures (GOP), dependant on the coding structure, containing either I, B, P or generalised B frames. Three coding structures are supported by the HEVC codec, namely: 1) All intra, 2) Low delay and 3) Random access. In the first case, each frame in the entire sequence is encoded as an I-frame or Instantaneous Decoding Refresh (IDR) picture. There



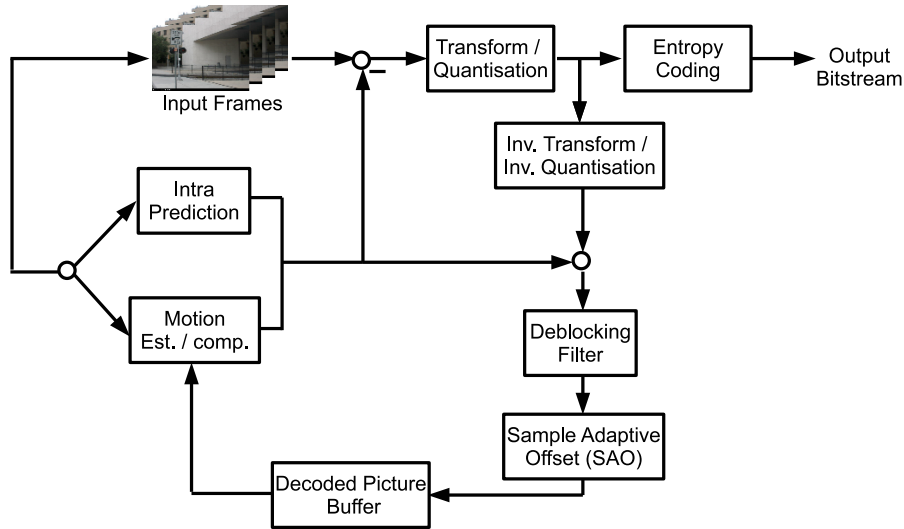


Figure 3.12: HEVC encoder diagram.

is no temporal prediction as each frame is individually encoded independent of one another. For low delay, the first picture is coded as an IDR frame with consequent frames coded as generalized P or B frames. To ensure computationally efficient coding, all temporal predictions are formed from past frames. The latter case, random access, implements a hierarchical B-frame structure, with an IDR frame inserted cyclically approximately once every second. Temporal predictions are formed from both past and future frames. Note that the Picture Order Count (POC) is not the same as the frame coding order, due to the future dependant B-Frame predictions. Figure 3.13 shows the 3 possible coding structure types.

### 3.3.2 Block Structure

HEVC employs an improved, more flexible quad-tree block partitioning structure within the prediction and transform stages of the codec. This ensures a more efficient block partitioning scheme as a greater number of block sizes are available. The HEVC block structure is determined by 3 main components; Coding Unit (CU), Prediction Unit (PU) and Transform Unit (TU).

The CU contains a hierarchical structure, as shown in Figure 3.14, defines a maximum block depth of 4 sizes. The quadtree syntax allows for splitting the CU

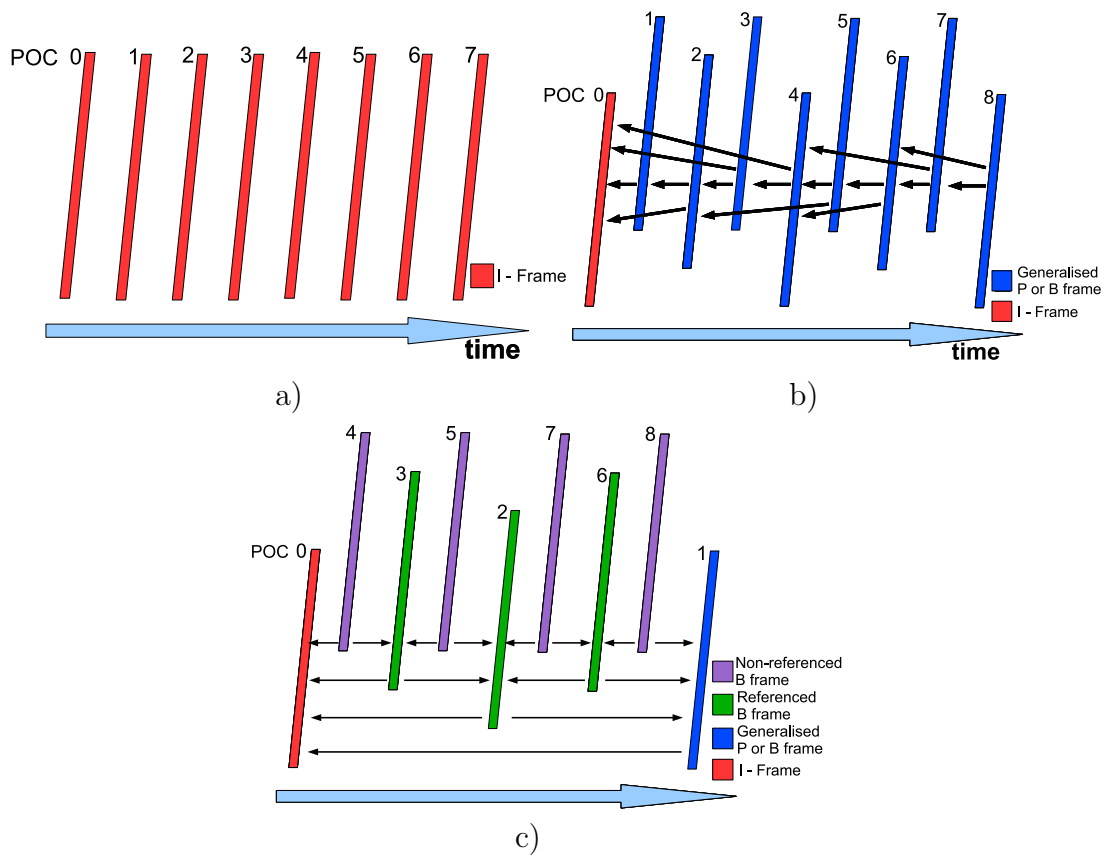


Figure 3.13: The three possible coding structures within HEVC a) All intra, b) Low delay and c) Random access.

into an appropriate size dependant upon the region characteristics by signalling a split flag. The CU is further decomposed into the PU, which comprises of 4, 8, 16, 32 and 64 block sizes only. Figure 3.15 displays the available PU block partitioning sizes available where  $n = N/2$  and  $U$ ,  $D$ ,  $L$  and  $R$  represent the upper, lower, left and right block dimensions, respectively. It is significant to note that intra mode prediction blocks can only take the form of  $PART\_2N \times 2N$  or  $PART\_N \times N$ , while inter prediction mode has full access to all 8 PU block dimensions. Further partitioning occurs within the TU, where obtainable block sizes range from 4x4 to 32x32.

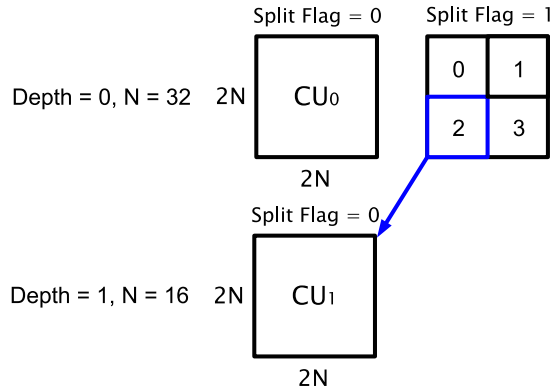


Figure 3.14: CU Block partitioning structure with the HEVC codec.

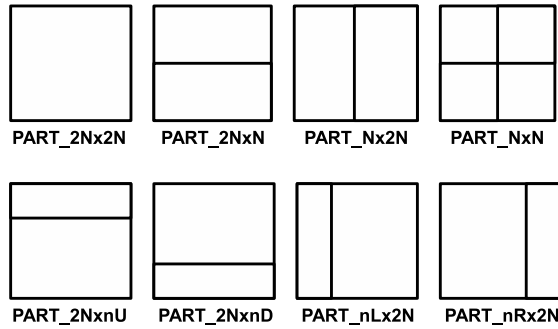


Figure 3.15: Possible PU Block partitioning options with the HEVC codec.

### 3.3.3 Intra Coding

Intra mode prediction is coded from coefficients within the same frame, without any temporal referencing, to exploit spatial correlation among pixels. HEVC incorporates a novel Arbitrary Directional Intra (ADI) prediction scheme to reduce residual errors, which arise from inaccurate block estimations in previous coding standards. Angular intra prediction is performed by interpolating the reference points, taken from the decoded boundary samples, surrounding each adjacent intra block. The best approximation from 33 differing angular predictions plus 2 additional modes from a DC average or planar prediction is chosen for the overall block prediction [116], as shown in Figure 3.16.

Each prediction mode utilizes the same set of reference points,  $R_{x,y}$ , which are depicted in Figure 3.17. DC mode provides a suitable block estimation when

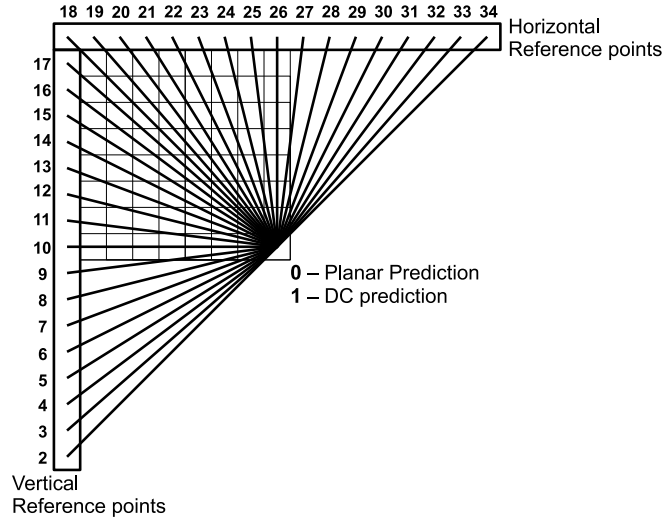


Figure 3.16: Possible ADI prediction modes.

predicting flat homogeneous regions. The prediction samples,  $P_{x,y}$ , for an  $N$  by  $N$  block are calculated from a zero order equation, described by:

$$P_{x,y} = \frac{\sum_{x=1}^N R_{x,0} + \sum_{y=1}^N R_{0,y}}{2N}. \quad (3.10)$$

where  $x$  and  $y$  represent the horizontal and vertical reference point indices, respectively. Essentially, DC prediction is derived from the mean of all reference points.

Planar prediction provides a suitable block approximation for smooth sample surfaces within homogeneous regions, without containing any prominent edge boundaries. [117] The prediction comprises from a combination of 2 first order linear equations,  $P1_{x,y}$  and  $P2_{x,y}$ , using 4 reference sample points:

$$\begin{aligned} P_{x,y} &= P1_{x,y} + P2_{x,y} \gg 1, & (3.11) \\ P1_{x,y} &= Ln_{x,y}(R_{N,y} + R_{0,y}), \\ P2_{x,y} &= Ln_{x,y}(R_{x,N} + R_{x,0}), \end{aligned}$$

where  $Ln_{x,y}$  is the linear interpolation between the respective reference samples. Planar prediction is a combination of both a horizontal and vertical estimation.

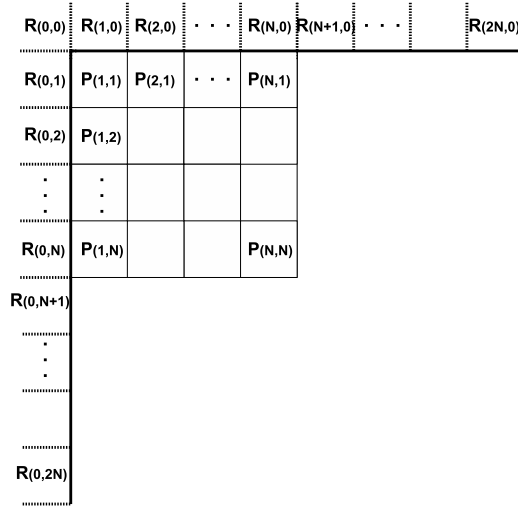


Figure 3.17: Reference samples  $R(x, y)$  used to predict prediction samples  $P(x, y)$  for an  $N * N$  block.

### 3.3.4 Inter Coding

Inter frame prediction requires a previous or future reference frame to estimate the current PU block, by exploiting temporal redundancies. A motion vector (MV) determines reference to this temporal coherence, which is calculated up to quarter-pel accuracy for the luma component and 1/8th-pel accuracy for chroma.

#### 3.3.4.1 Motion Compensation

HEVC implements an 8-tap filter,  $8tap[i]$ , for half-pel positions and a 7-tap filter,  $7tap[i]$ , for the quarter-pel locations. The long interpolation filter length improves precision compared to the H.264/AVC predecessor. Table 3.5 shows the 7 and 8 tap filter coefficients, which are derived from the Discrete Cosine Transform (DCT). [118] The integer values within the filters eliminate the need for intermediate rounding, reducing errors. Figure 3.18 shows how the original luma coefficients,  $(A(0,0), B(0,0), C(0,0)$  and  $D(0,0))$  displayed in a blue box, are interpolated. The 7-tap filter calculates the coefficients shown in red,  $(A(0,1), A(0,3), A(1,0), A(1,1), A(1,2), A(1,3), A(3,0), A(3,1), A(3,2)$  and  $A(3,3))$  whereas the yellow regions portrays coefficients within the half-pel position,  $(A(0,2), A(2,0), A(2,1), A(2,2)$  and  $A(2,3))$  calculated by an 8-tap filter.

Table 3.5: Filter coefficients for luma fractional sample interpolation.

<b>index[i]</b>	-3	-2	-1	0	1	2	3	4
$8tap[i]$	-1	4	-11	40	40	-11	4	1
$7tap[i]$	-1	4	-10	58	17	-5	1	



Figure 3.18: Integer and fractional sample positions for luma interpolation.

For YUV 4:2:0 video, the chroma components for the fractional sample interpolation process are highly comparable with the luma, except a 4-tap filter used to obtain 1/8-pel accuracy.

### 3.3.4.2 Advanced Motion Vector Prediction (AMVP)

HEVC enforces a unique merge mode to perform AMVP. Merge mode derives motion information (MV and reference picture indices) from spatially or temporally neighboring blocks. Index information from one of several selectable merge mode candidates is transmitted. Figure 3.19 shows the possible spatial (a1, b1, b0, a0, b2) and temporal (H, C3) candidates for an inter coded PU block. AMVP selects the most suitable candidate from the list so only merge mode index, predicted MV difference and residual data is required for transmission.

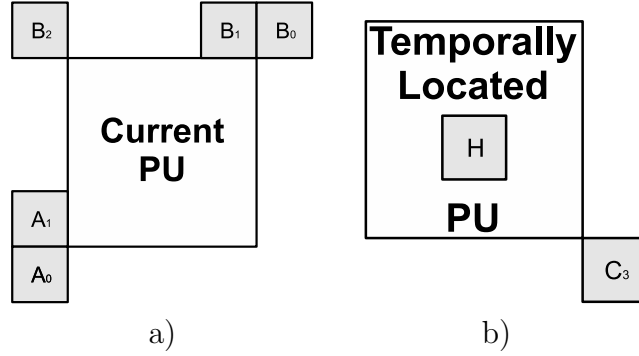


Figure 3.19: Possible inter merge mode candidates, located a) spatially b) temporally.

## 3.4 Evaluation Tools and Datasets

This section describes evaluation tools and common parameter values, within each experiment (unless otherwise stated), used throughout this thesis.

### 3.4.1 Subjective and Objective Evaluation Tools for Visual Quality

Visual quality can be evaluated by both subjective and objective evaluation. Objective metrics define a precise value, dependant upon mathematical modeling, to determine visual quality. Numerous existing techniques exist:

**PSNR** - Peak Signal to Noise Ratio (PSNR) [119] is one of the most commonly used visual quality metrics. It is based on the root mean square error (RMSE) of two images, or frames, with a dimension of  $XY$  as described in Equation (3.12) and Equation (3.13).

$$PSNR = 20 \log_{10} \left( \frac{255}{RMSE} \right) dB. \quad (3.12)$$

$$RMSE = \sqrt{\frac{1}{X * Y} \sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} (I(m, n) - I_0(m, n))^2}. \quad (3.13)$$

**SSIM** - Structural Similarity Index Measure (SSIM) [120] assumes that the HVS is highly adapted for extracting structural information from a scene. Unlike PSNR, where average error between two images is taken into consideration, SSIM focuses on a quality assessment based on the degradation of structural information. By using local luminance and contrast rather than average luminance and contrast, the structural information in the scene is calculated.

**JND** - Just Noticeable Difference (JND) [121] DCT transforms the host and original media before using thresholding. The thresholds are decided based on 1) luminance masking and 2) contrast masking of the transformed images. The threshold for luminance pattern relies on the mean luminance of the local image region, whereas the contrast masking is calculated within a block and particular DCT coefficient using a visual masking algorithm.

Objective measurements cannot always accurately assess media quality. Two distorted frames with comparable PSNR, SSIM or JND do not necessitate coherent media quality. Two independent viewers can undergo entirely different visual experiences, as two similarly distorted frames can provide a contrasting opinion for which contains higher visual quality. Subjective evaluation evaluates media quality by assessing human perception. In this thesis, two subjective evaluation measures are performed, namely:

**DSCQT**- Double Stimulus Continuous Quality Test (DSCQT) [122] subjectively evaluates any media distortion by using a continuous scale. The original and distorted media is shown to the viewer in a randomised order, who must measure the media quality of both data sets by Degradation Category Rating (DCR), as shown in 3.20a). A visual quality degradation rating value is calculated by the absolute difference between the two subjective results.

**DSIST**- Double Stimulus Impairment Scale Test (DSIST) [122] determines the perceived visual degradation between two media sources, implementing a discrete scale. Unlike the DSCQT case, the original and distorted media order are not randomised. A viewer must rank any quality deterioration, on a 5-point discrete Absolute Category Rating (ACR) scale, as shown in 3.20b).

Training images are first shown, in both subjective cases, to acclimatize viewers



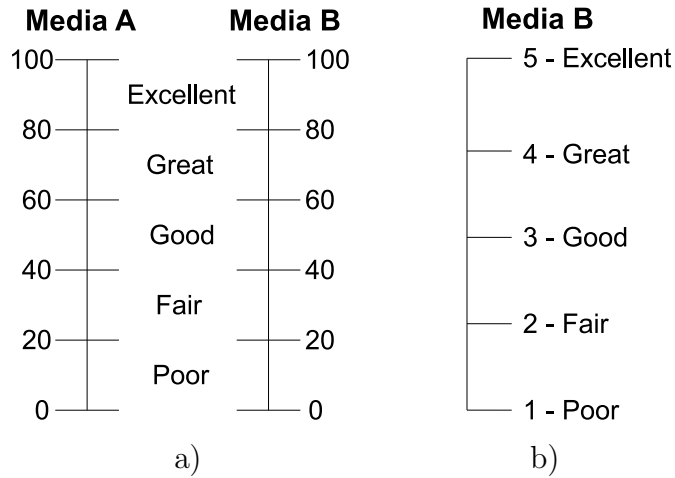


Figure 3.20: Subjective testing visual quality measurement scales a) DCR continuous measurement scale b)ACR ITU 5-point discrete quality scale.

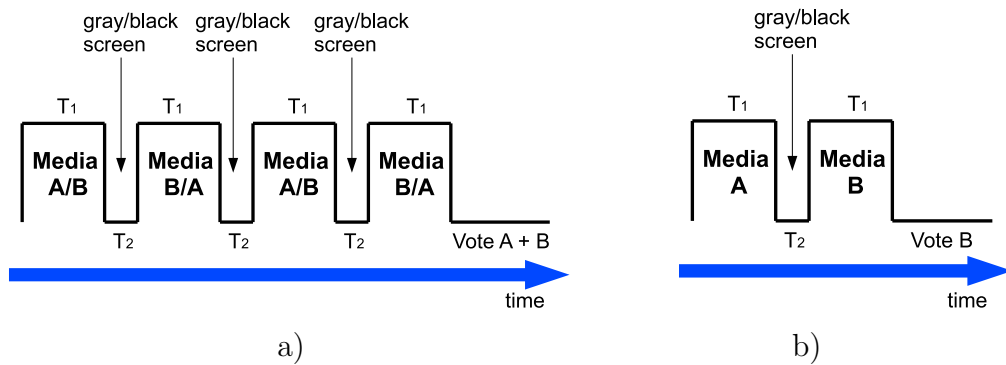


Figure 3.21: Stimulus timing diagram for a) DCR method b) ACR method.

to each ACR and DCR scoring system. In either of the two subjective tests, a higher DCR or ACR values represents a greater perceived viewer distortion. Both subjective evaluation measures follow appropriate testing standard criteria, defined within the International Telecommunication Union (ITU) [122]. An overall timing diagram for each subjective methodology is shown in Figure 3.21. Note that the media display time,  $T_1$ , and blank screen time,  $T_2$ , should satisfy the following condition:  $T_1 > T_2$ .

**VQM** - Video Quality Metric (VQM) [123] - evaluates video quality performance based upon subjective human perception. It incorporates numerous aspects of early visual processing, including both luma and chroma channels, a combination

of temporal and spatial filtering, light adaptation, spatial frequency, global contrast and probability summation. A numeric output is generated between 1 and 0 and higher video quality is distinguished by values closer to zero. The VQM is a commonly used video quality assessment metric as it eliminates the need for participants to provide a subjective evaluation.

As well as objective and subjective metrics, application-based evaluation measures also determine the quality of results. Dependant upon the target application, a model performance can vary massively.

This thesis incorporates both subjective and objective evaluation methodologies. Due to the simplicity, both SSIM and PSNR are used for objective watermarking evaluation. Notably each described objective metric performs in a similar manor, when dealing with a PSNR of 35 or above.

### **3.4.2 Experimental Datasets**

The experimental datasets can be devised into 2 sections for the image and video test sequences required throughout this thesis.

#### **3.4.2.1 Image Datasets**

The Microsoft Research Asia (MSRA) saliency datadase, by Liu et al [124], provides thousands of publicly available images, from which 1000 are selected to form the MSRA-1000. Subsequent ground truth ROI frames, governed by the outcome of subjective testing, have been manually created as part of the MSRA-1000 database. The data test set has been manually labeled by 3 users. Each of the users initially drew a bounding box around the most salient regions of 20,000 frames. The manual labelling took approximately 10-20 seconds per frame and around 3 weeks in total. The dataset was narrowed down to 5,000 frames by selecting the most consistent data. Salient portions within each of the 5,000 frames are labeled by 9 users into a binary ground truth map, segmenting the ROI, and the most consistant 1,000 frame make up the MSRA-1000 database. A test frame

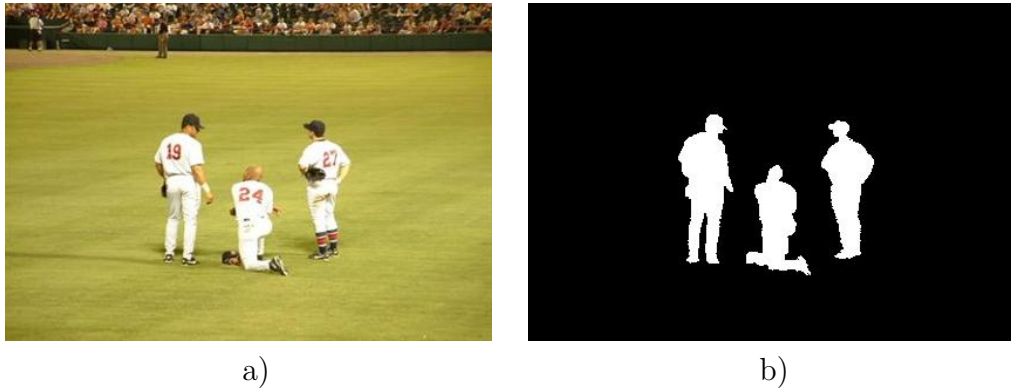


Figure 3.22: MSRA-1000 database a) Test frame from database b) Ground truth ROI frame.

example from the dataset, with corresponding manually segmented ground truth map, is shown in Figure 3.22.

The Kodak image test set, containing 24 colour scenes, is also utilised within this thesis and can be found from the link:  
<http://r0k.us/graphics/kodak/>

### 3.4.2.2 Video Datasets

The video dataset is taken from [125] and comprises of 15 video sequences, containing over 2000 frames in total. Ground truth video sequences have been generated from the database by subjective testing as in Section 3.4.2.1. A thumbnail from each of the 15 test sequences are shown in Figure 3.23.

Confidence intervals of 95% are implemented to eliminate any anomalous results, when dealing with multiple image or video frames. A Pseudorandom binary sequence,  $b \in 0, 1$ , determines the watermark data to be embedded within each sequence. All the saliency model and watermarking-based simulations were performed using Matlab.

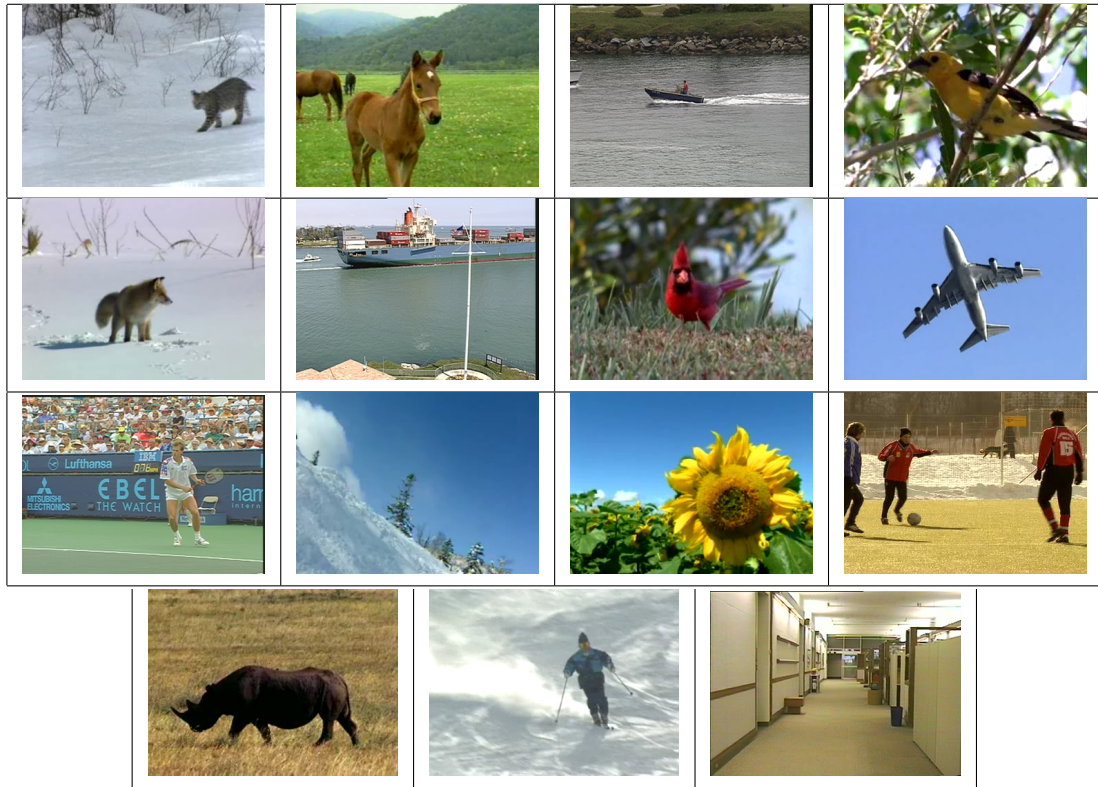


Figure 3.23: Video database - 15 thumbnails for each sequence.

### 3.5 Conclusions

In this chapter an overview and background detailing digital watermarking, VA and the HEVC codec are presented. Firstly, the basic properties, methodology and applications behind digital watermarking are briefed, followed by an insight into the VAM. Saliency model motivation, algorithm evaluation and numerous applications are consequently described. Finally, the HEVC codec is described, highlighting new key advancements in video coding technology, such as the new ADI prediction scheme, hierarchial block structure and new inter mode AMVP. In the next chapter the state-of-the-art study on wavelet based watermarking schemes, related to image and video watermarking are discussed. Existing fore-front approaches towards visual saliency models are also reviewed.

# Chapter 4

## Visual Saliency Estimation

The most attentive regions within media can be captured by exploiting and imposing characteristics from within the HVS. In this chapter, a novel method is proposed to detect any saliency information within an image or video sequence. The proposed methods incorporate wavelet decomposition combined with HVS modeling to capture any spatiotemporal saliency information which is directly integrable within the watermarking framework. A unique approach combining salient intensity, colour and orientation contrasts formulate the essential image domain saliency methodology where as temporal anomalies greatly contribute towards the generation of video saliency maps. Each of the image and video VA algorithms provided in this chapter can be simply implemented and are highly pertinent toward many wavelet-based applications such as watermarking, compression and fusion.

### 4.1 Introduction

Physiological and psychophysical evidence demonstrate visually stimulating regions occur at different scales within media [126]. Consequently, models proposed in this work exploit the identifiable multi-resolution property of the wavelet transform to generate both an image and video domain model. This thesis addresses wavelet-based watermarking methodologies based upon the VAM. It is highly

advantageous to provide a visual saliency model, generated directly from within the wavelet domain, which is directly integrable within the wavelet-based watermarking framework. Firstly, Section 4.2 presents a spatial saliency model applicable within the image domain. A consequent video domain saliency model is provided, in Section 4.3, fusing incoherent temporal characteristics within the spatial saliency framework.

## 4.2 Image Domain Saliency

By exploiting the multi-spatial representation of the wavelet transform, VA is estimated directly from within the wavelet domain. The image saliency model is divided into 3 sections. Firstly, Section 4.2.1 analyses the spatial scale implemented within the design and Section 4.2.2 describes the saliency algorithm. Finally, Section 4.2.3 shows the model performance.

### 4.2.1 Generating Scale Feature Maps

Due to the decreasing resolution after each wavelet decomposition iteration, the spatial synchronisation of objects within the frame distorts, limiting the useful contribution of coefficients towards the overall saliency map at very fine resolutions. Figure 4.1 shows the successive coefficient magnitude of each LH subband,  $LH_i$ , interpolated back to full frame resolution for  $i \Rightarrow \{1 - 7\}$  levels of wavelet transform using the luma channel. If  $M$  represents the maximum subband coefficient magnitude, the potential range of coefficients within each subband is  $LH_i \Rightarrow \{0 - M\}$ . This normalises the overall saliency contributions from each subband and prevents biasing towards the finer scaled subbands. Figure 4.1 shows after 5 levels of decomposition, the threshold to retain coefficient spatial synchronisation has been surpassed. Consequently, a highly distorted profile is displayed for the interpolation after 6 and 7 successive transform decompositions. The detailed investigation, shown in Figure 4.1, shows limited meaningful saliency data can be extracted after 5 levels, determining the upper bound of required transform iterations implemented in the overall algorithm design. This

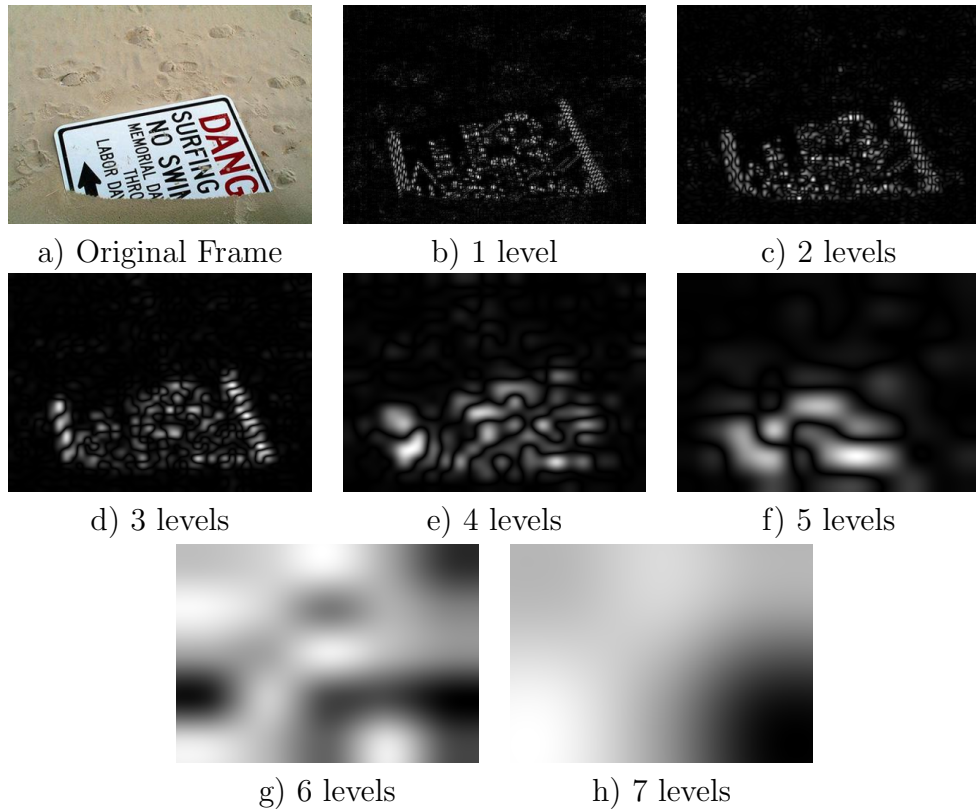


Figure 4.1: Interpolated LH subbands using a CIF resolution (352x288) image, for each successive wavelet decomposition level.

is true for all images within the database, described in Section 3.4.2.1.

## 4.2.2 Saliency Map Generation

The LH, HL and HH subbands emphasise horizontal, vertical and diagonal contrasts within a frame, respectively, portraying prominent edge boundaries. By combining these contrasts using an MRA approach, salient scene portions can be highlighted. This section describes the main body of the wavelet-based image saliency algorithm.

Each Y,U and V channel is firstly decomposed by 5 levels of wavelet decomposition and normalised within the range  $\{0 - M\}$ . The magnitude of each LH, HL and HH subband is computed to prevent negative salient regions as contrasting signs can potentially nullify salient regions when combined. To provide full

resolution output maps, each of the high frequency subbands are consequently interpolated up to full frame resolution as shown previously in Figure 4.1. Equation (4.1) depicts this process mathematically, showing how the absolute full resolution subband feature maps  $lh_i$ ,  $hl_i$  and  $hh_i$  are generated from the LH, HL and HH subbands in the luma channel, respectively:

$$\begin{aligned}lh_i &= (|LH_i|^{\uparrow 2^i}), \\hl_i &= (|HL_i|^{\uparrow 2^i}), \\hh_i &= (|HH_i|^{\uparrow 2^i}),\end{aligned}\tag{4.1}$$

where  $\uparrow 2^i$  is the bilinear upsample operation by a factor 2 for  $i$  levels of wavelet decomposition. Fusion of  $lh_i$ ,  $hl_i$  and  $hh_i$  provides a feature map for each subband in the luma channel. The normalised maps are combined by a weighted linear summation which is illustrated algebraically in Equation (4.2):

$$\begin{aligned}LH_Y &= \sum_{i=1}^5 lh_i * \tau_i, \\HL_Y &= \sum_{i=1}^5 hl_i * \tau_i, \\HH_Y &= \sum_{i=1}^5 hh_i * \tau_i,\end{aligned}\tag{4.2}$$

where  $\tau_i$  is the subband weightage parameter and  $LH_Y$ ,  $HL_Y$  and  $HH_Y$  are the subband feature maps for the luma channel. 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 finely tune the algorithm it's logical to apply a slight bias toward the middle scale subband maps, *i.e.*,  $\tau_3 < \tau_2, \tau_4 < \tau_1, \tau_5$ , although in practice this provides a minimal algorithm performance improvement over an equal subband weighting ratio. The reasoning is saliency is not specific towards a definite resolution [127].

Research suggests promoting feature maps which exhibit a low quantity of strong activity peaks [54], while suppressing maps flaunting an abundance of peaks pos-



sessing similar amplitude. Similar neighbouring features inhibit visual attentive selectivity, whereas a single peak surrounded by boundless low activity facilitates visual stimuli. If  $\bar{m}$  is the average of local maxima present within the feature map and  $M_g$  is the global maximum, the promotion and suppression normalisation is achieved by Equation (4.3):

$$\begin{aligned}\overline{LH}_Y &= LH_Y * (M_g - \bar{m})^2, \\ \overline{HL}_Y &= HL_Y * (M_g - \bar{m})^2, \\ \overline{HH}_Y &= HH_Y * (M_g - \bar{m})^2,\end{aligned}\tag{4.3}$$

where  $\overline{LH}_Y$ ,  $\overline{HL}_Y$  and  $\overline{HH}_Y$  are the normalised set of subband feature maps.

The entire process is repeated for each of the chroma channels and each of the LH, HL and HH subband feature maps are consequently forged into a respective Y, U and V saliency map ( $Ymap$ ,  $Umap$  and  $Vmap$ ) as shown in Equation (4.4), Equation (4.5) and Equation (4.6). Each Y, U and V feature map linearly combine in Equation (4.7) forming the final image saliency estimation,  $S_{Spat}$ :

$$Ymap = \overline{LH}_Y + \overline{HL}_Y + \overline{HH}_Y,\tag{4.4}$$

$$Umap = \overline{LH}_U + \overline{HL}_U + \overline{HH}_U,\tag{4.5}$$

$$Vmap = \overline{LH}_V + \overline{HL}_V + \overline{HH}_V,\tag{4.6}$$

$$S_{Spat} = Ymap + Umap + Vmap.\tag{4.7}$$

Normalisation only occurs only after the final combination in Equation (4.7). If the U or V channels portray sparse meaningful saliency information, only a minimal effect will occur from incorporating these features within the final map. An overall saliency model diagram is shown in Figure 4.2.

### 4.2.3 Results

Subjective and objective experimental results demonstrate the model performance against existing state-of-the-art methodologies. The 4 chosen state-of-the-art approaches are selected for a wide variety of differing approaches to estimate VA, all of which are described in more detail in Section 2.3. For all objective

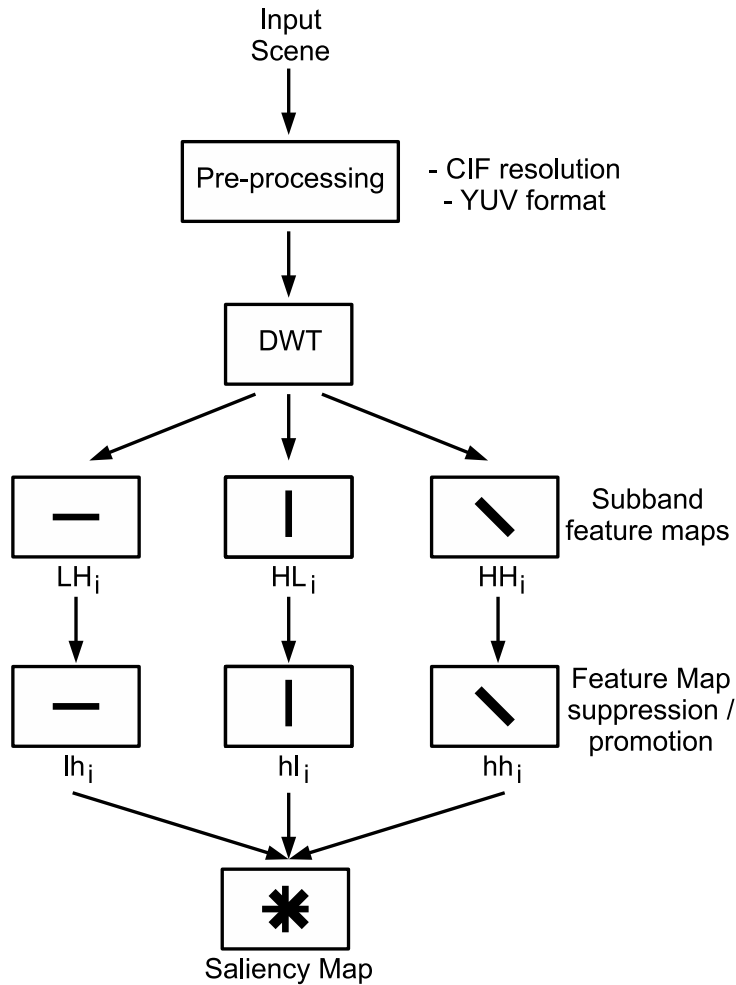


Figure 4.2: Overall Image Saliency Model Block Diagram.

measurements, the experimental database in Section 3.4.2.1 is utilised to evaluate the model performance. Figure 4.3, Figure 4.4 and Table 4.1 show the saliency model performance for image, static video and average scene computational time, respectively, comparing the proposed method against 4 differing state-of-the-art techniques. Saliency estimations for 4 images are shown in Figure 4.3 and 3 video frames in Figure 4.4, where as the mean computational time from 1000 independent MSRA database images was equated in the bottom row in Table 4.1.

For Figure 4.3 and Figure 4.4, row 2 demonstrates the performance of the Itti model [54], which portrays moderate saliency estimation, when subjectively compared to the ground truth frames in row 6. A drawback to this model is the added

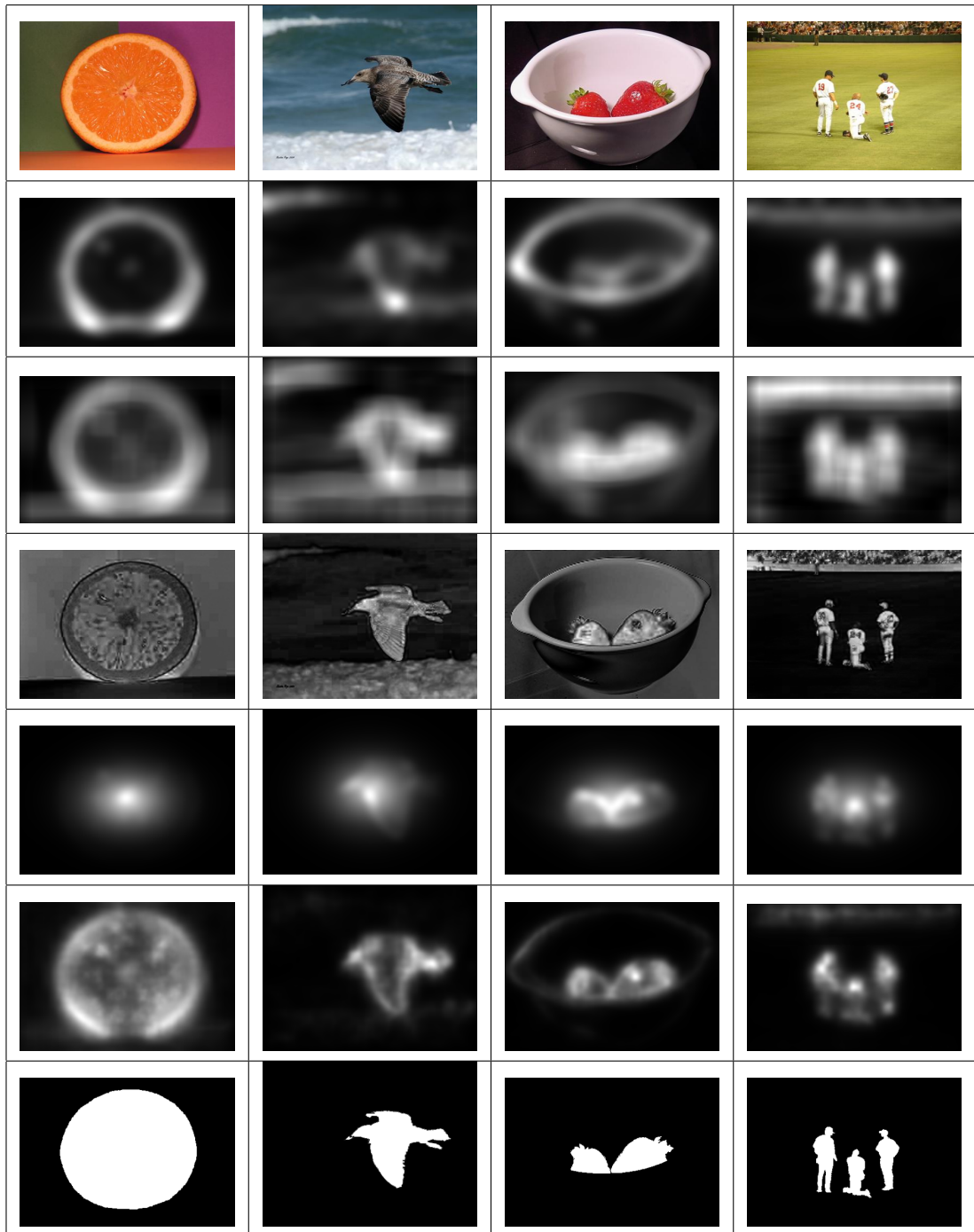


Figure 4.3: Image Saliency model state-of-the-art comparison: Row 1 - Original image from MSRA database. Row 2 - Itti model [54]. Row 3 - Rare model [67]. Row 4 - Ngau model [69]. Row 5 - Erdem model [68]. Row 6 - Proposed Method. Row 7 - Ground Truth.

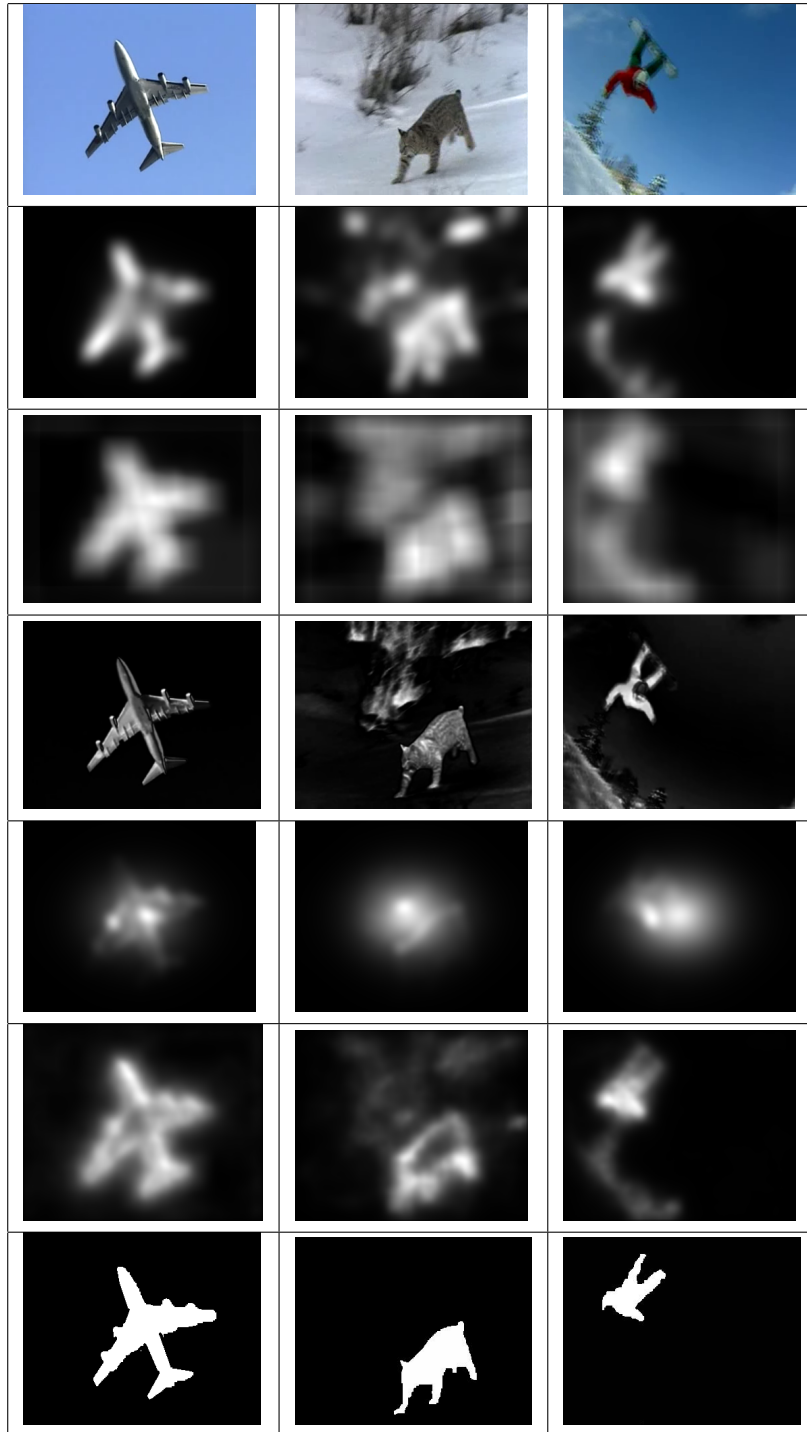


Figure 4.4: Static video frame comparison to state-of-the-art: Row 1 - Original sequence frame. Row 2 - Itti model [54]. Row 3 - Rare model [67]. Row 4 - Ngau model [69]. Row 5 - Erdem model [68]. Row 6 - Proposed Method. Row 7 - Ground Truth.

Table 4.1: Computational time comparing state-of-the-art image domain saliency models.

Saliency Model	Itti [54]	Rare [67]	Ngau [69]	Erdem [68]	Proposed
ROC AUC for 1000 frames	0.875	0.906	0.856	0.878	0.887
Computational time per frame (sec)	0.281	6.374	0.092	16.540	0.142

computational cost, persisting approximately twice the proposed algorithm. The Rare algorithm [67] 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 row 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 [69] is shown in row 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 frames containing a wide variety of intensities and colour, the model breaks down as shown in column 2, in Figure 4.3, where the white portion within the sea is visually misclassified as an interesting region. Row 5 shows the generated saliency maps from the Erdem model [68]. The proposed model is shown in row 6 and clearly highlights any salient activity within in each of the 7 frames, by locating the presence of intensity and colour contrasts. By accurately highlighting any salient regions, within the wavelet domain, an authentic visual attention based watermarking scheme can be provided. For example, the proposed method clearly highlights the orange, bird, strawberries and cricketers, in Figure 4.3, and plane, wolf and snowboarder, in Figure 4.4.

Subjective assessment of the saliency model alone does not provide an adequate algorithm evaluation. ROC curves in Figure 4.5 and ROC AUC values in the top row of Table 4.1 contribute an objective measure, as described in Section 3.2.2, to determine the saliency model performance against numerous state-of-the-art methodologies. The proposed technique shows superior performance compared with the Itti and wavelet-based methods having an ROC AUC 1.4% and 3.6% 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, shown in Figure 4.4. Further results,

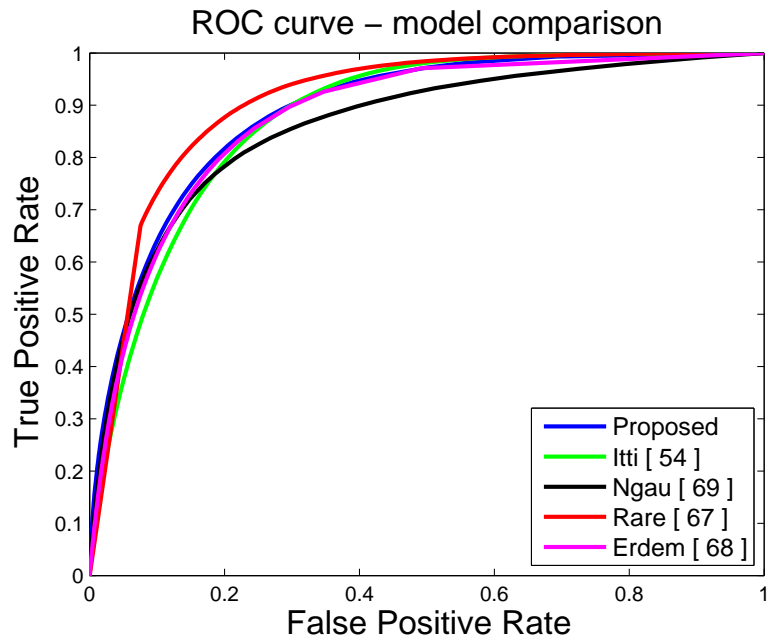


Figure 4.5: ROC curve comparing state-of-the-art image domain saliency algorithms.

portraying scenes with varying frame resolutions, taken from the MSRA-1000 database are shown for inspection in Figure 4.6.



Figure 4.6: Image saliency model from MSRA-1000 database - column 1: host image - column 2: proposed saliency map - column 3: ground truth map.

## 4.3 Video Domain Saliency

This subchapter concentrates upon wavelet-based video saliency estimation and it comprises of 5 parts. Firstly, Section 4.3.1 provides an incite into the 2D+t domain and why this is a logical choice for saliency estimation. Section 4.3.2 describes the temporal saliency map generation, Section 4.3.3 describes compensating for global motion and the final video saliency algorithm is explained in Section 4.3.4. Finally, Section 4.3.5 evaluates the proposed model performance.

### 4.3.1 2D+t Wavelet Domain

To provide a complete wavelet-based solution towards video domain saliency, logically, the 3D wavelet transform is utilised. Video coding research provides evidence that differing texture and motion characteristics occur after wavelet decomposition from the t+2D domain [128] and incorporating its alternative technique, the 2D+t transform [129] [130]. The t+2D domain decomposition compacts most of the transform coefficient energy within the low frequency temporal subband and provides efficient compression within the temporal high frequency subbands. Vast quantities of the high frequency coefficients have zero magnitude, or very close, which severely limits the transform usefulness within the watermarking framework. Alternatively, 2D+t decomposition produces greater transform energy within the higher frequency components, i.e a greater amounts of larger and non-zero coefficients. The extra energy stored within the high frequency subbands facilitates any potential robust watermarking schemes which are dependant upon embedding information within the high-pass temporal coefficients. Consequently, the overall video embedding distortion is greatly reduce.

This thesis focuses upon digital watermarking based upon VA characteristics so consequently the 2D+t transform domain is adopted, as VA-based watermarking within the low frequency subbands can leave highly perceptible artifacts. Previous visually uninteresting regions can attract human gaze resultant from low frequency domain embedding. A diagram of 2D+t decomposition is shown in Figure 4.7 for 3 levels of spatial and 1 level of temporal haar wavelet decomposition.



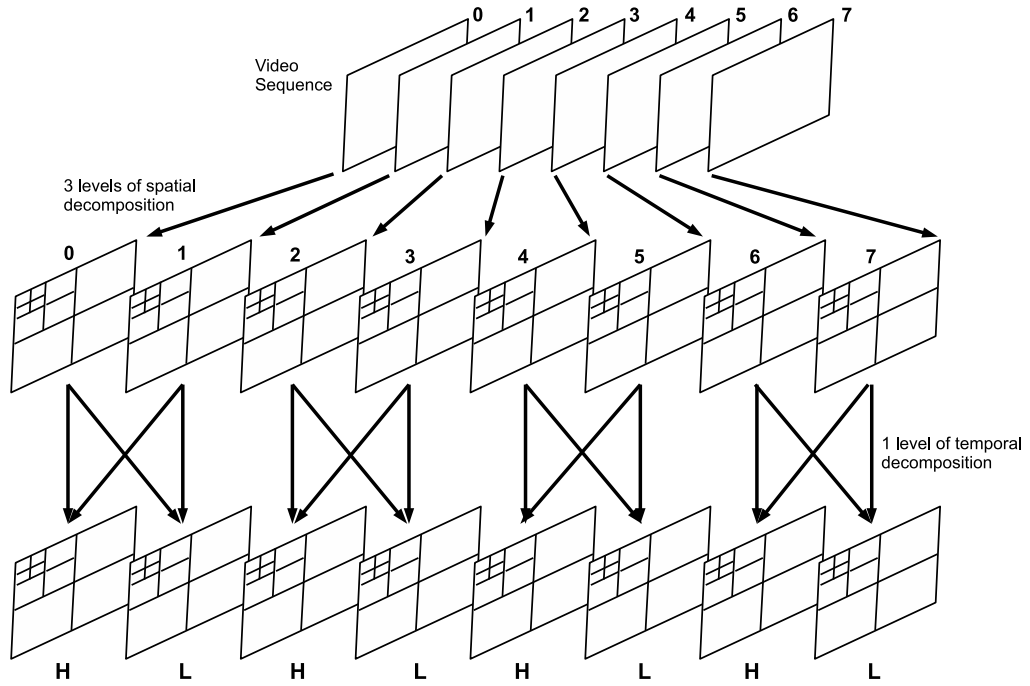


Figure 4.7: 2D+t wavelet decomposition.

### 4.3.2 Temporal Saliency Feature Map

To acquire accurate video saliency estimation, both spatial and temporal features within the wavelet transform are considered. The wavelet-based image domain solution, described in Section 4.2 constitutes the spatial element for the video saliency model where as this section concentrates upon establishing temporal saliency maps,  $S_{Temp}$ .

Correlatable methodology to expose temporal conspicuousness is implemented in comparison to the spatial model in Section 4.2. Firstly, the existence of any palpable local object motion is determined within the sequence. Figure 4.8 shows the histograms of 2 globally motion compensated frames. Global Motion is any frame motion due to camera movement, whether that be panning, zooming or rotation. This is explained further in Section 4.3.3. Change within lighting, noise and global motion compensation error account for the peaks present within Figure 4.8a), whereas the contribution from object movement is also present within Figure 4.8b). A local threshold,  $T$ , segments frames containing sufficiently noticeable local motion, from an entire sequence. If  $F_1(x, y)$  and  $F_2(x, y)$  are

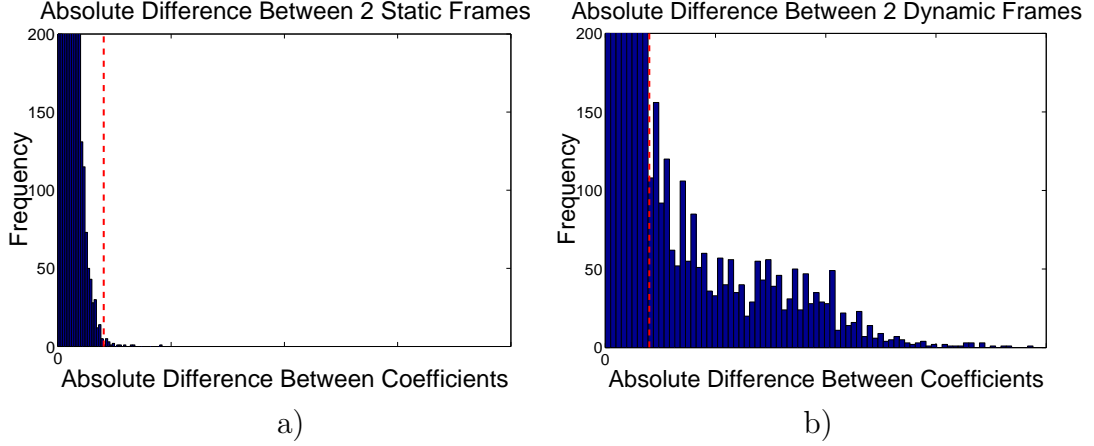


Figure 4.8: Difference frames after global motion compensation: a) Without local motion b) Containing local motion.

consecutive 8-bit luma frames within the same sequence, Equation (4.8) classifies temporal frame dynamics:

$$|F_1(x, y) - F_2(x, y)| > T, \quad (4.8)$$

From the histograms shown within Figure 4.8a) and Figure 4.8b), a local threshold value of  $T = D_{max}/10$  determines motion classification, where  $D_{max}$  is the maximum possible frame pixel difference, and  $T$  is highlighted by a red dashed line within both figures. A 0.5 percent error ratio of coefficients must be greater than  $T$ , to reduce frame misclassification.

For each temporally active frame, the Y channel renders sufficient information to estimate salient object movement without considering the U and V components. 5 levels of spatial wavelet decomposition are adopted to provide optimum performance. After 5 decomposition levels, temporal disfiguration arises in a comparable manner to the spatial model as shown in Figure 4.1. 1 level of temporal decomposition provides adequate information to extract any valid salient motion features.

The  $S_{Temp}$  methodology bears a distinct similarity to the spatial domain approach as the high pass temporal subbands:  $LHt_i$ ,  $HLt_i$  and  $HHt_i$ , for  $i$  levels of spatial decomposition, combine after full 2D+t wavelet decomposition, which is shown in Figure 4.7. The decomposed data is forged using comparable logic as Equa-

tion (4.1) and Equation (4.2), as all transformed coefficients are segregated into 1 of 3 temporal subband feature maps. This process is described in Equation (4.9):

$$\begin{aligned}
LHt_t &= \sum_{i=1}^n (|LHt_i|^{\uparrow 2^i} * \tau_i), \\
HLt_t &= \sum_{i=1}^n (|HLt_i|^{\uparrow 2^i} * \tau_i), \\
HHt_t &= \sum_{i=1}^n (|HHt_i|^{\uparrow 2^i} * \tau_i),
\end{aligned} \tag{4.9}$$

where  $LHt_t$ ,  $HLt_t$  and  $HHt_t$  are the temporal LH, HL and HH combined feature maps, respectively. The method captures any subtle conspicuous object motion, in both horizontal, vertical and diagonal directions. This subsequently fuses the coefficients into a meaningful visual saliency approximation by merging the data across multiple scales.  $S_{Temp}$  is finally generated from Equation (4.10):

$$S_{Temp} = LHt_t + HLt_t + HHt_t, \tag{4.10}$$

### 4.3.3 Global Motion Compensated Frame Difference

Compensation for global motion is dependant upon homogeneous MV detection, consistent throughout the frame. Figure 4.9 considers the motion estimation between 2 consecutive frames, taken from the coastguard sequence. A fixed block size based upon the frame resolution determines number of MV blocks. The magnitude and phase of the MVs are represented by the size and direction of the arrows, respectively, whereas the absence of an arrow portrays a MV of zero. Firstly, it is assumed there is a greater percentage of pixels within moving objects than the background, so large densities of comparative MVs are the result from dynamic camera action. To compensate for camera panning, the entire reference frame is spatially translated by the most frequent MV, the global camera MV,  $M_{global}^{\rightarrow}$ . This process is applied, prior to the 2D+t wavelet decomposition to deduce global motion compensated saliency estimation. The global motion compensation is described in Equation (4.11):

$$M_{object}^{\rightarrow} = M_{total}^{\rightarrow} - M_{global}^{\rightarrow}, \tag{4.11}$$



Figure 4.9: Motion block Estimation

where  $M_{object}^{\rightarrow}$  is the local object MV and  $M_{total}^{\rightarrow}$  is the complete combined MV.

Compensating for other camera movement can be achieved by searching for a particular pattern of MVs. For example a circular MV pattern will determine camera rotation and all MVs converging or diverging from a particular point will govern camera zooming. An iterative search over all possible MV patterns can cover each type of global camera action. [131] Speeded Up Robust Features (SURF) detection [132] could be used to directly align key feature points between consecutive frames but this would be very computationally exhaustive. This model only requires a fast rough global motion estimate to neglect the effect of global camera motion on the overall saliency map.

#### 4.3.4 Spatial-Temporal Saliency Map Combination

The spatial and temporal maps are combined to form an overall saliency map. The primary visual cortex is extremely sensitive to object movement so if enough local motion is detected, within a frame, the overall saliency estimation is dominated by any temporal contribution. Hence, the temporal weightage parameter,

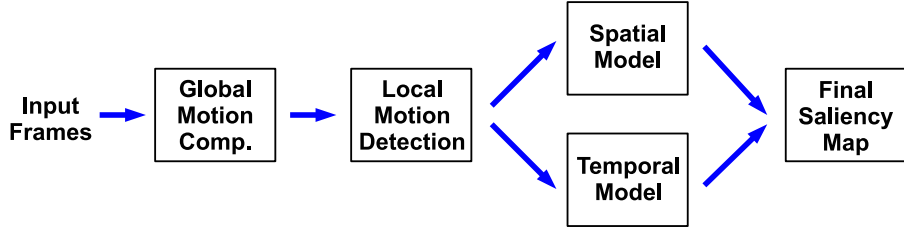


Figure 4.10: Proposed Saliency Model Block Diagram.

$\alpha$ , determined from Equation (4.8) is calculated in Equation (4.12):

$$\alpha = \begin{cases} 1 & \text{if } |F_1(x, y) - F_2(x, y)| > T, \\ 0 & \text{otherwise.} \end{cases} \quad (4.12)$$

If significant motion is detected within a frame, the complete final saliency map comprises solely from the temporal feature. Previous studies support this theory, providing evidence that local motion is the most dominant feature within low level VA [133]. Consequently, if no local motion is detected with a frame, the spatial model contributes towards the final saliency map in its entirety, hence  $\alpha$  is a binary variable. The equation forging the overall saliency map is,

$$S_{Final} = S_{Temp} * \alpha + S_{Spat} * (1 - \alpha), \quad (4.13)$$

where  $S_{Spat}$ ,  $S_{Temp}$  and  $S_{Final}$  are the spatial, temporal and combined overall saliency maps, respectively. An overall diagram for the entire proposed system is shown in Figure 4.10.

### 4.3.5 Results

The proposed algorithm is compared with the Itti [74] and Dynamic [134] video saliency models, in terms of accurate salient region detection and computational efficiency. The Itti framework is seen as the foundation and benchmark used for saliency model comparison, whereas the Dynamic algorithm is more recent, dependant upon locating energy peaks within incremental length coding. The video dataset, described in Section 3.4.2.2, is used for model comparison. The top row in Table 4.2 shows the complexity of each algorithm in terms of average

Table 4.2: Computational time comparison of video saliency models.

<b>Saliency Method</b>	Itti [74]	Dynamic [134]	Proposed
<b>Average frame computational time (sec)</b>	0.244	0.194	0.172
<b>ROC AUC</b>	0.804	0.769	0.832

frame computational time. The values in the table are calculated from the mean computational time over every frame within the video database and provide the time required to form a saliency map from the original raw frame. All calculations include any transformations required. From the table, the proposed low complex methodology can produce a video saliency map around 70% and 88% of the time for an Itti and Dynamic model frame, respectively.

Figure 4.11 shows the performance of the proposed model opposed to the Itti and Dynamic algorithms subjected to the presence of significant object motion, which dominates the saliency maps. The Itti motion model saliency maps are depicted in row 2 and the Dynamic model saliency maps in row 3. Any distinguishes between local and global movement are not fully accounted for and the maps are dominated by spatially attentive features, leading to salient object misclassification. For example, the trees within the background of the snowboard sequence are estimated as an attentive region, when a man is performing acrobatics within the frame foreground. The proposed model is shown in row 4. From left to right, the locally moving snowboarder, flower and bird are clearly identified as salient objects. Corresponding ground truth frames are shown in row 5, which depict all salient local object movement.

The ROC curves and corresponding ROC AUC values, shown in Figure 4.12 and the bottom row in Table 4.2, respectively, display an objective model evaluation. The results show the proposed method exceeds the performance of the Itti motion and Dynamic models having a 3.5% and 8.2% higher ROC AUC, respectively.

Further results demonstrating the video saliency estimation model across 4 video sequences are shown in Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16. Video saliency becomes more evident when viewed as a sequence rather than from still frames. The video sequences with corresponding saliency maps are

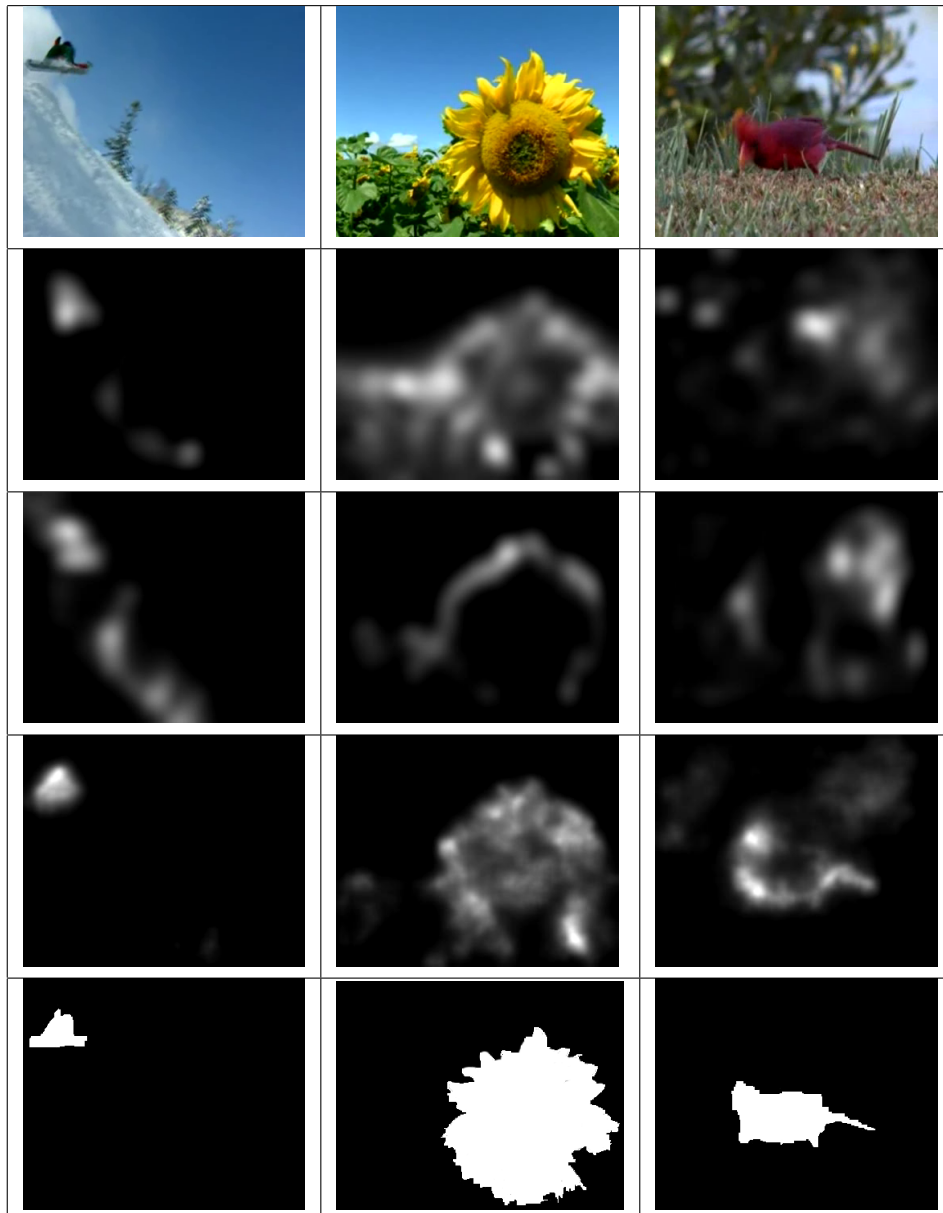


Figure 4.11: Temporal Saliency model comparison table: Row 1 - Original frame from sequence. Row 2 - Itti video model. [74] Row 3 - Dynamic model. [134] Row 4 - Proposed Method. Row 5 - Ground Truth.

available for viewing at the following website address:

<http://svc.group.shef.ac.uk/va-video.html>.

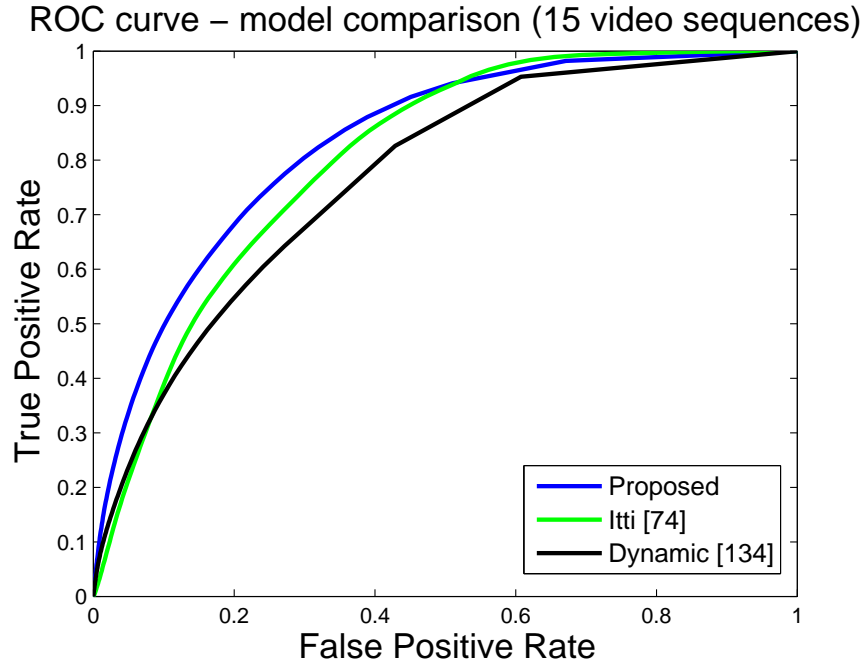


Figure 4.12: ROC curve comparing performance of proposed model.

## 4.4 Conclusions

In this chapter, a wavelet-based image and video saliency model is provided by merging spatial and temporal features to locate visually attentive regions. The image saliency algorithm detects conspicuous scene regions by combining intensity, colour and orientation contrasts within the wavelet domain. For the video saliency model, both camera and object movement jointly contribute to the perceived motion between consecutive frames. By estimating and compensating for the dynamic camera trajectory, local motion from salient objects can be extracted. The degree of accuracy, when performing global motion compensation, limits the overall speed of the proposed method due to the iterative nature of the operation. However, real-time video saliency estimation is attainable for scenes containing low global camera movement due to the fast and simple algorithm. Experimental simulations show the wavelet-based saliency models have significant improvements, in terms of accurately estimating salient regions and frame computational time, against existing state of the art methodologies. This is verified by a 3.5% or greater increase in ROC AUC and a computational time 88% or lower than state-of-the-art video saliency algorithms. The image saliency proposal, compared with existing approaches, provides a 1.2% or greater increase in



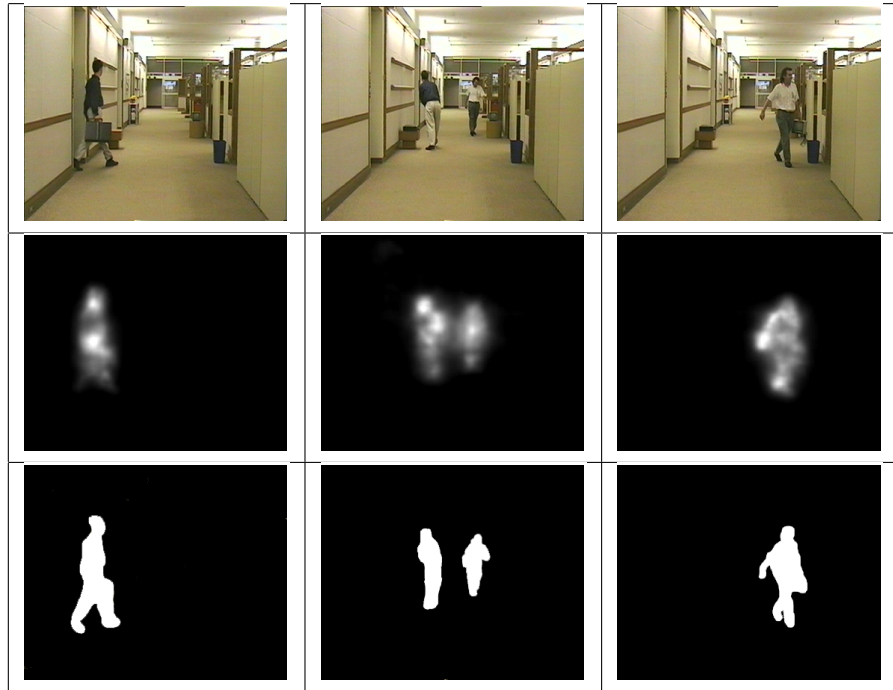


Figure 4.13: Video saliency estimation results sequence 1: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth.

ROC AUC for algorithms exhibiting a similar computational complexity time. The next chapter focuses upon the application of the wavelet-based saliency algorithms within the watermarking domain.

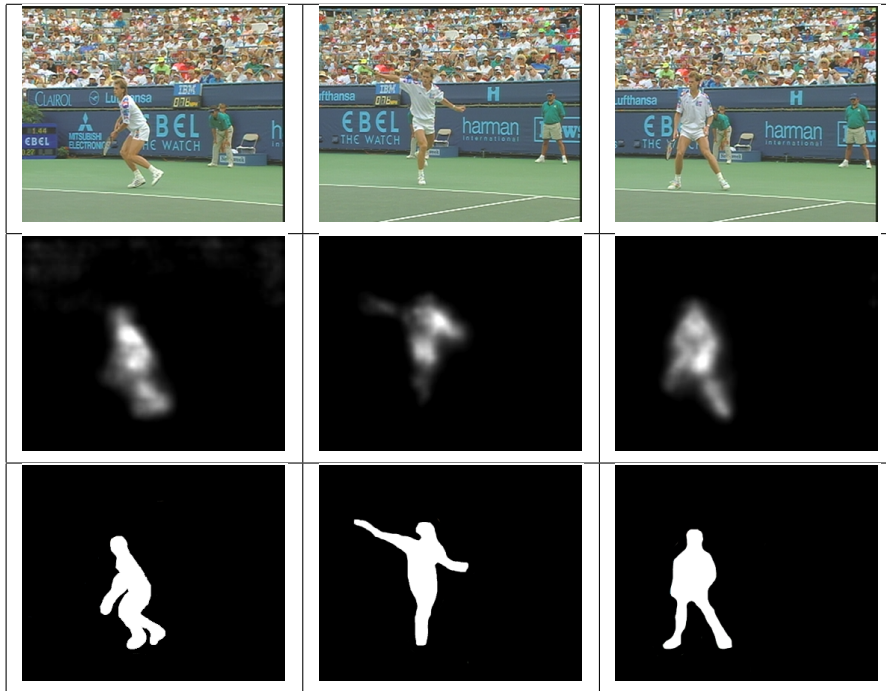


Figure 4.14: Video saliency estimation results sequence 2: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth.

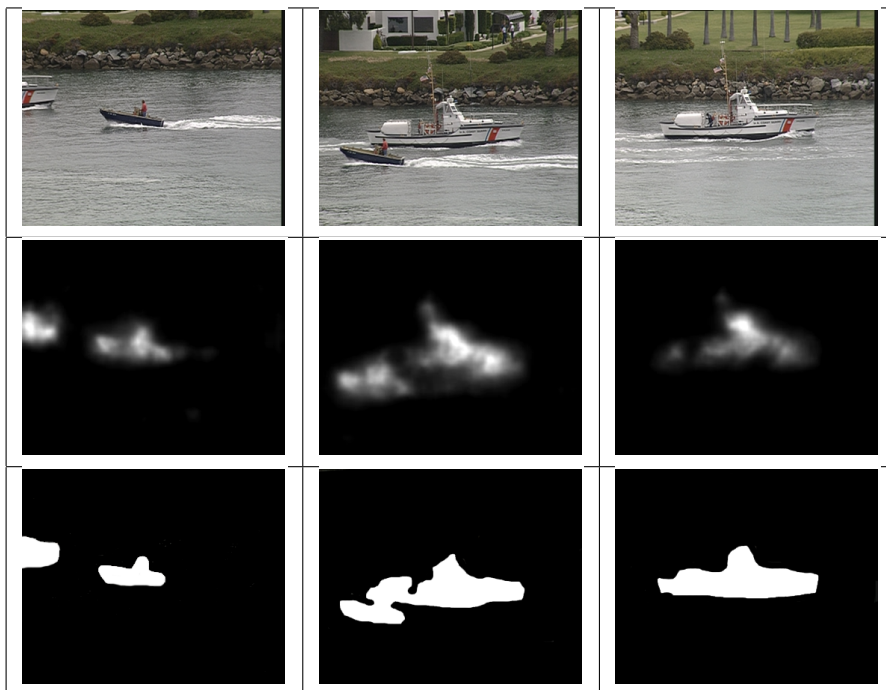


Figure 4.15: Video saliency estimation results sequence 3: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth.

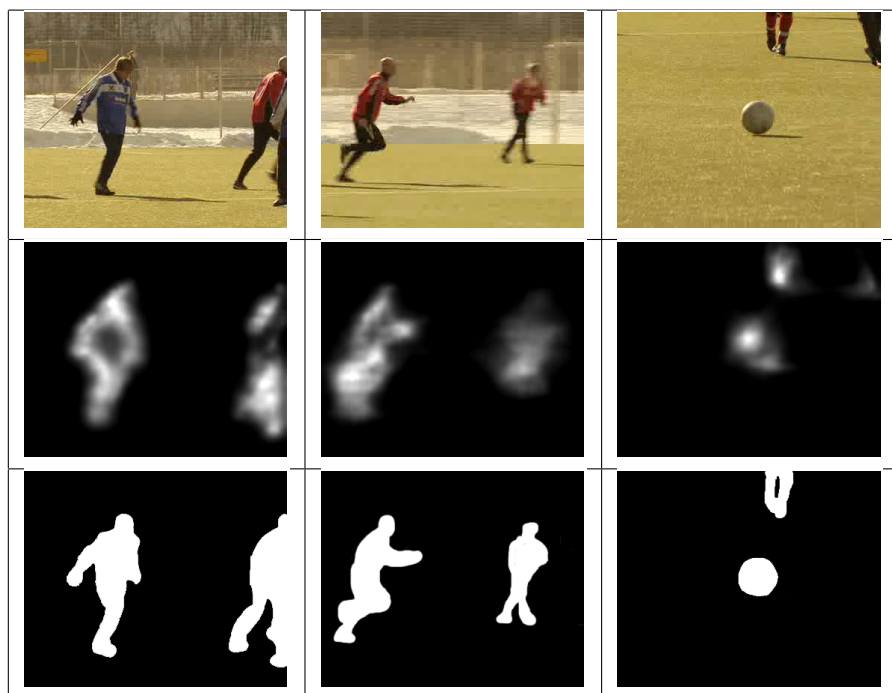


Figure 4.16: Video saliency estimation results sequence 4: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth.



# Chapter 5

## Visual Attention-based Watermarking

A ROI dictates the most important visible aspects within media, so distortion within these areas will be highly noticeable to any viewer. The VAM computes such regions, which is a highly useful concept within the watermarking field. Substantial previous work has been published based on image domain saliency, as discussed in Chapter 2, contrasting to the seldom literature available on video saliency algorithms. In this chapter, a novel image and video watermarking scheme is presented using the VAM, where the visual saliency map is computed within the wavelet domain as described in Chapter 4. By embedding greater watermark strength within the less visually appealing regions, within the media, a highly robust scheme is attained without compromising the visual quality of the data. Consequent joint watermark robustness and imperceptibility results are also provided in this chapter.

### 5.1 VA-based Image Watermarking

The VAM identifies the ROI, most perceptible to human vision, which is a highly exploitable property when designing watermarking systems. The subjective effect of watermark embedding distortion can be greatly reduced if any artifacts oc-

cur within inattentive regions. By incorporating VA-based characteristics within the watermarking framework, algorithms can provide a retained subjective media quality and increased overall watermark robustness, compared with non-VA methodologies. Recalling previous VA-based research in Chapter 2, the saliency map is usually generated within the pixel domain. Section 4.2.1 provides efficient wavelet domain saliency map generation for the image domain, in order to easily adapt any wavelet-based watermarking schemes within the VA framework. Section 5.1.1, Section 5.1.2 and Section 5.1.4 provide the novel saliency-based watermarking scheme.

### 5.1.1 Saliency Map Thresholding

Recalling blind and non-blind watermarking schemes, in Chapter 2, the host media source is only available within non-blind algorithms. Identical saliency reconstruction might not be possible within the watermark extraction process so VA-based quantisation is implemented to threshold saliency maps into similar regions of visual attentiveness. The employment of a threshold reduces saliency map reconstruction errors, which may occur resultant of any watermark embedding distortion. This statement is justified further in Section 5.1.3.

Figure 5.1a) and Figure 5.1b) show an original host image and corresponding saliency map, respectively, generated from the methodology in Section 4.2. In Figure 5.1b), the light and dark regions, within the saliency map, represent the attentive and visually uninteresting areas, respectively. A cumulative histogram of the coefficients within the saliency map is shown in Figure 5.1c). Histogram analysis depicts automatic segmentation of the map into 2 independent levels by employing threshold  $T_{sal}$ .  $T_{sal}$  is dependant upon the cumulative frequency of coefficients to segment highly conspicuous locations within a scene. From the graph, the maximum frequency of coefficients and max saliency values are represented by  $f_{max}$  and  $S_{max}$ , respectively and  $T_{sal} = f_{max} * 0.75$ . An adaptive watermark strength map is determined from the threshold selections in Figure 5.1c), by utilizing a high watermark embedding strength in regions of low visual saliency and a low embedding strength in highly conspicuousness areas. To ensure VA-based embedding, the watermark weighting parameter strength,  $\alpha$ , is made variable,

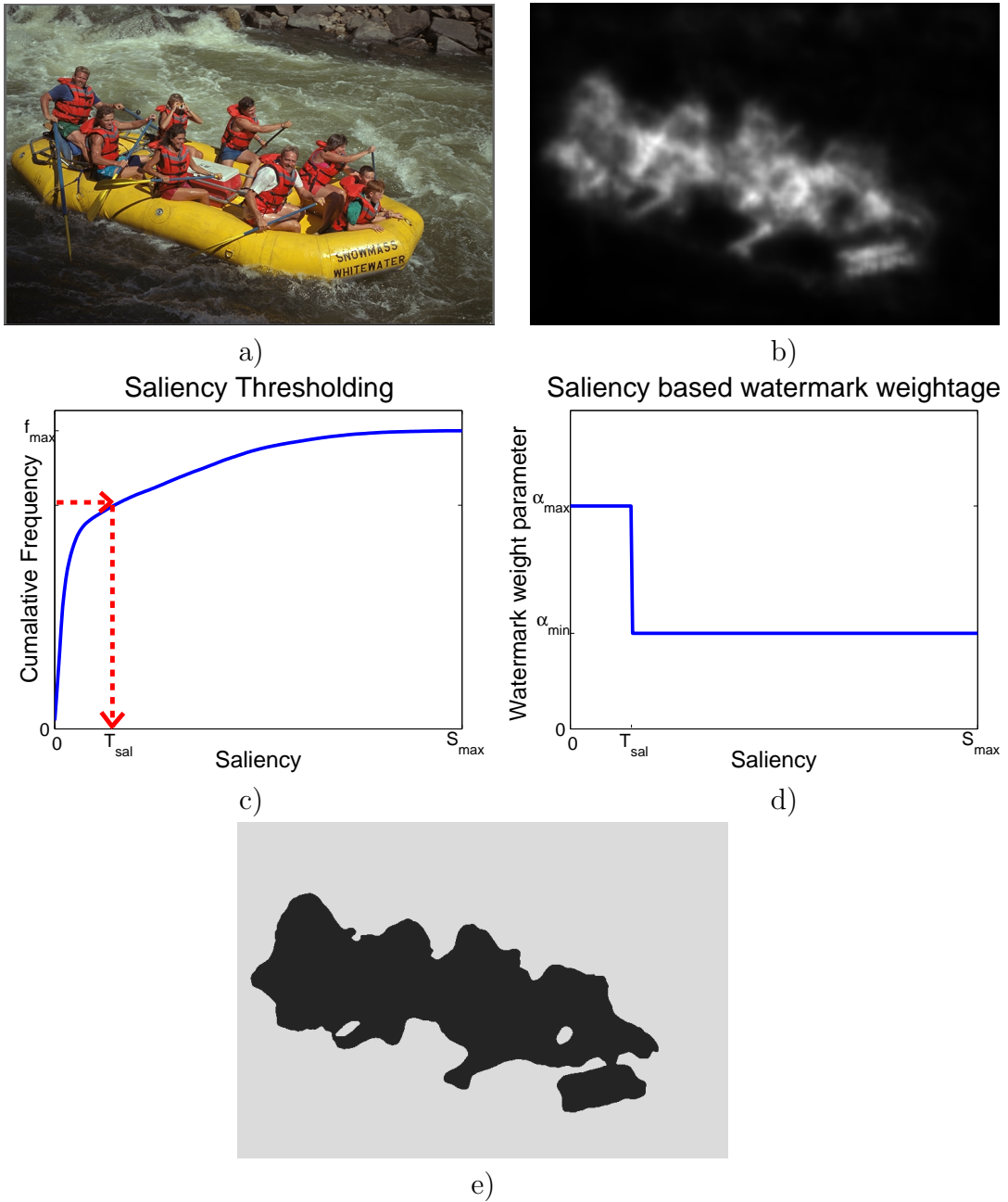


Figure 5.1: a) Host image b) VAM saliency map c) Cumulative saliency histogram d)  $\alpha$  step graph e)  $\alpha$  strength map.

dependant upon the thresholded saliency level. As shown in Figure 5.1d), the most and least salient regions are given watermark weighting parameters of  $\alpha_{min}$  and  $\alpha_{max}$ , respectively. The actual  $\alpha_{min}$  and  $\alpha_{max}$  values are determined from the analysis within Section 5.1.2. The final VA-based alpha watermarking strength map is shown in Figure 5.1e), where a brighter intensity represents an increase in  $\alpha$ . Further test images, with corresponding alpha maps are shown in Figure 5.2.



Figure 5.2:  $\alpha$  strength map examples - A - Original Image B -  $\alpha$  strength map.



### 5.1.2 Watermark Embedding Strength Calculation

The watermark weighting parameter strengths,  $\alpha_{max}$  and  $\alpha_{min}$  can be calculated from the visible artifact PSNR limitations within the image. Visual distortion becomes noticeable as the overall PSNR drops below 40db [135], 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 data can be embedded into any host image to enhance watermark robustness without substantially distorting the media quality. Recalling the combination of Equation (3.12) and Equation (3.13) from Section 3.4.1, PSNR is calculated by:

$$PSNR = 10\log_{10} \left( \frac{M^2}{\frac{1}{X * Y} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (I'_{(n,m)} - I_{(n,m)})^2} \right), \quad (5.1)$$

where  $M$  is the maximum coefficient value of the data. The equation for non-blind magnitude based additive watermarking, shown in Equation (5.2):

$$C'_{(m,n)} = C_{(m,n)} + \alpha\omega_{(m,n)}C_{(m,n)}, \quad (5.2)$$

can be rearranged and substituted into Equation (5.1) forming Equation (5.3), as follows:

$$PSNR = 10\log_{10} \left( \frac{M^2}{\frac{1}{X * Y} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (\alpha\omega_{(m,n)}I_{(m,n)})^2} \right). \quad (5.3)$$

By rearranging for  $\alpha$ , an equation determining the watermark weighting parameter, depending on the required PSNR value is derived for non-blind watermarking in Equation (5.4) as:

$$\alpha = \frac{M}{\sqrt{\frac{1}{X * Y} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (\omega_{(m,n)}I_{(m,n)})^2 \cdot 10^{\left(\frac{PSNR}{10}\right)}}} \quad (5.4)$$

Determining  $\alpha_{max}$  and  $\alpha_{min}$ , for the blind watermarking scheme described in Section 3.1.3.2, is achieved in a similar manor to Equation (5.4), by rearranging Equation (5.1). Consequently, substituting the median and modified median coefficients,  $C_{(med)}$  and  $C'_{(med)}$ , respectively, gives:

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |C'_{(med)} - C_{(med)}| = \frac{M}{\sqrt{\frac{1}{X * Y} \cdot 10^{\left(\frac{PSNR}{10}\right)}}, \quad (5.5)$$

where  $C'_{(med)}$  is a function of  $\alpha$ ,  $C_{min}$ ,  $C_{max}$  and  $b$  as shown in Equation (5.6):

$$C'_{(med)} = f(\alpha, C_{min}, C_{max}, b). \quad (5.6)$$

Equation (5.5) determines the total coefficient modification for a given PSNR requirement, hence determining  $\alpha$ .

### 5.1.3 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 become distorted. Thresholding the saliency map into 2 levels, as described in Section 5.1.1, ensures high accuracy within the saliency model reconstruction for blind watermarking. Figure 5.3 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 34.97, using the blind watermarking scheme described in Section 3.1.3.2. The figure 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 Figure 5.3c) and Figure 5.3d) appear indistinguishable, although objective analysis determines only 55.6% of coefficients are identical. In Figure 5.3e) and Figure 5.3f) 99.4% of saliency coefficients match, hence reconstruction errors are greatly reduced due to the thresholding.

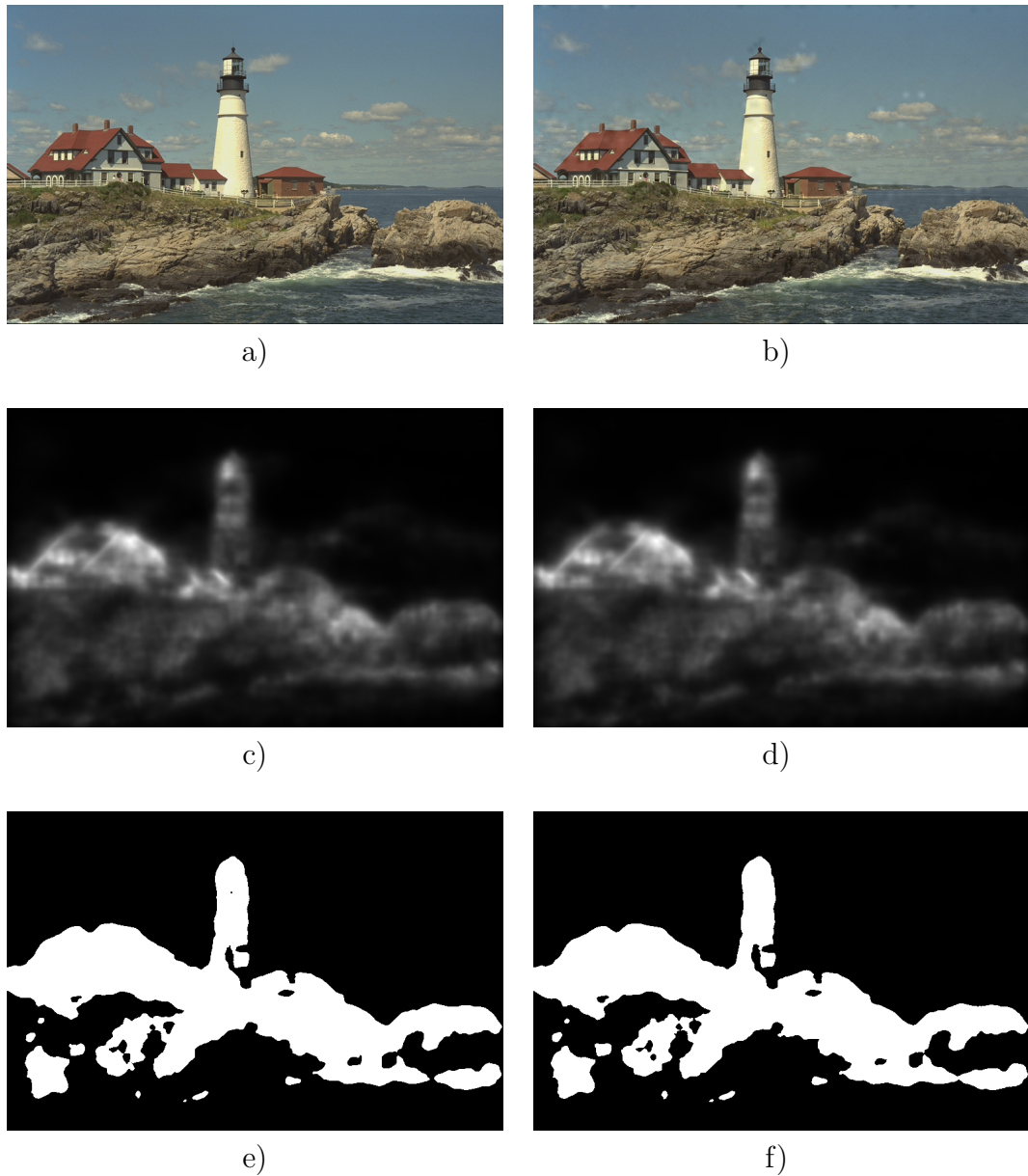


Figure 5.3: Saliency map reconstruction - a) Original host frame b) Watermarked frame embedded using a constant  $\alpha_{max}$  c) Host frame saliency map d) Saliency map of watermarked frame e) Original thresholded saliency map. f) Reconstructed saliency map thresholded after blind watermark embedding.

#### 5.1.4 System Architecture

From the adaptive VA-based  $\alpha$  maps, the proposed novel architecture is incorporated within the classical watermarking framework, illustrated in Figure 3.3.

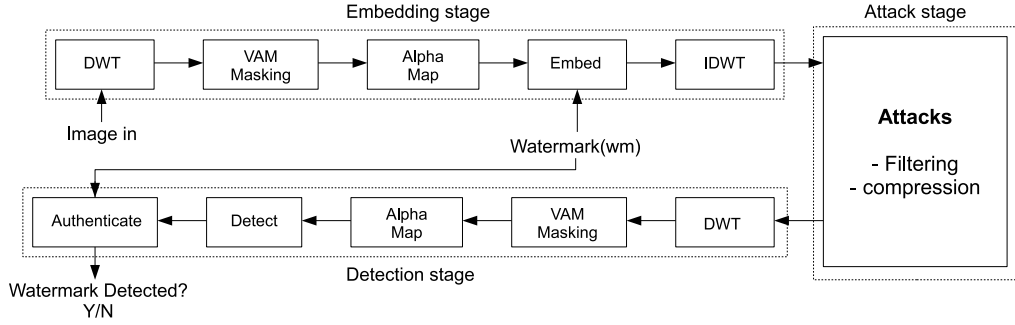


Figure 5.4: Visual Attention-based Watermarking Scheme.

The resultant overall system diagram, adaptive toward both blind and non-blind watermarking scenarios, is shown in Figure 5.4.

### 5.1.5 Experiments, Results and Analysis

Experimental results are gathered by embedding watermark data within selected wavelet subbands, for both the blind and non-blind watermarking scenarios. For each embedding case,  $\alpha_{min}$  and  $\alpha_{max}$  were calculated from the minimum and maximum PSNR requirements of 35 and 40dB, using Equation (5.4) and Equation (5.5), as described in Section 5.1.2. For all experimental simulations, common test set parameters include: the orthogonal Daubechies-4 (D4) wavelet and a database containing the 24 test images as described in Section 3.4.2.1. Each of the subjective evaluations, described further in Section 3.4.1, are determined from a panel 30 individuals.

Throughout the entire experimental results section, 4 differing scenarios are analysed, with  $\alpha$  varying in each instance. The 4 watermarking scenarios consist of:

- 1) a uniform  $\alpha_{min}$  for the entire scene;
- 2) the proposed watermarking scheme which implements an adaptive VA-based  $\alpha$ ;
- 3) a constant average watermark strength,  $\alpha_{ave}$ , where  $\alpha_{ave} = (\alpha_{min} + \alpha_{max})/2$ ; and

- 4) a constant  $\alpha_{max}$  throughout the embedding procedure.

The experimental results are consequently deciphered into 2 sections: imperceptibility and robustness. The imperceptibility of the watermarking schemes are determined by measuring any embedding distortion through objective and subjective evaluation tools. Robustness is considered against natural image processing and filtering attacks as implemented by Checkmark [136], and content adaption by Watermarking Evaluation Bench for Content Adaptation Modes (WEBCAM) [137].

#### 5.1.5.1 VA-based Experimental Image Data Sets

Image test sets from the subjective and objective evaluation are provided, displaying: the host image, a low strength watermarked scene, VA-based watermarked image and a low strength watermarked image. Figure 5.5 demonstrates the dangers of VA-based LL subband embedding within media containing large homogeneous regions of low texture. Previously uninteresting regions increase in visual saliency, due to the visual artifacts and fixate human gaze towards the distortion. Past studies indicate the human eye is less sensitive towards pixel modification in regions containing high texture but more sensitive near any edges, within these high texture areas [138]. Therefore in Figure 5.6, a scene containing large portions of high texture is displayed and the VAM-based embedding artifacts appear less noticeable. Figure 5.7 and Figure 5.8 show VAM-based watermarking results for the HL and LH subbands, respectively. The figures display the luma channel only to depict the watermark magnitude, although the embedding can be simply applied to each of the chroma components for colour images. In all cases the VAM-based watermarking scheme contains a significant visual quality improvement over using the high strength watermarking scheme.

#### 5.1.5.2 Embedding Distortion

Embedding distortion results are shown in Table 5.1 and Table 5.2, which display PSNR and SSIM measures for both non-blind and blind watermarking cases, re-



a)



b)

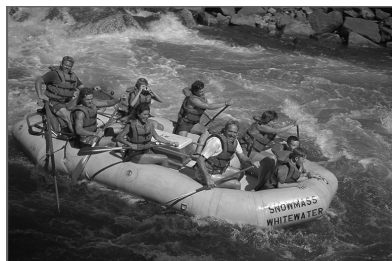


c)

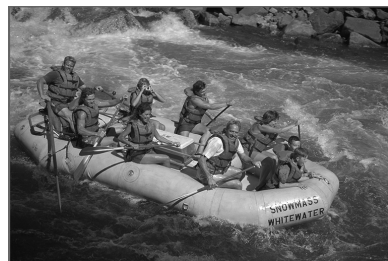


d)

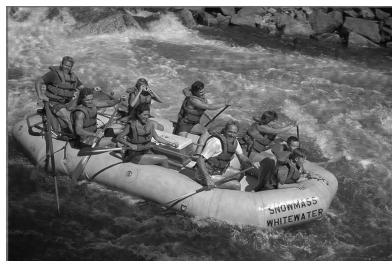
Figure 5.5: LL low texture subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.



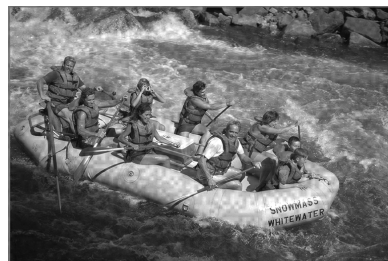
a)



b)



c)



d)

Figure 5.6: LL high texture subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.



a)



b)



c)



d)

Figure 5.7: HL subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.



a)



b)



c)



d)

Figure 5.8: LH subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.

Table 5.1: PSNR and SSIM values - non-blind watermarking.

<b>LL Subband</b>				
	<b>Low Strength</b>	<b>Proposed</b>	<b>Average Strength</b>	<b>High Strength</b>
PSNR	39.91 $\pm$ 0.06	36.07 $\pm$ 0.24	37.37 $\pm$ 0.07	34.92 $\pm$ 0.04
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>HL Subband</b>				
PSNR	39.92 $\pm$ 0.07	36.42 $\pm$ 0.26	37.28 $\pm$ 0.08	34.95 $\pm$ 0.06
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>LH Subband</b>				
PSNR	39.90 $\pm$ 0.05	36.18 $\pm$ 0.28	37.39 $\pm$ 0.09	34.94 $\pm$ 0.05
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>HH Subband</b>				
PSNR	39.94 $\pm$ 0.06	36.42 $\pm$ 0.29	37.45 $\pm$ 0.08	34.97 $\pm$ 0.06
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00

Table 5.2: PSNR and SSIM values - blind watermarking.

<b>LL Subband</b>				
	<b>Low Strength</b>	<b>Proposed</b>	<b>Average Strength</b>	<b>High Strength</b>
PSNR	39.93 $\pm$ 0.08	37.17 $\pm$ 0.26	37.44 $\pm$ 0.08	34.94 $\pm$ 0.06
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>HL Subband</b>				
PSNR	39.92 $\pm$ 0.08	37.21 $\pm$ 0.29	37.38 $\pm$ 0.09	34.96 $\pm$ 0.08
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>LH Subband</b>				
PSNR	39.95 $\pm$ 0.07	36.98 $\pm$ 0.29	37.35 $\pm$ 0.08	34.96 $\pm$ 0.08
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00
<b>HH Subband</b>				
PSNR	39.96 $\pm$ 0.08	37.08 $\pm$ 0.31	37.46 $\pm$ 0.09	34.96 $\pm$ 0.08
SSIM	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00	0.99 $\pm$ 0.00

spectively. All the results are portrayed for watermark embedding after 3 levels of wavelet decomposition. From the tables, PSNR improvements of approximately 2dB are achieved when comparing the proposed and constant high strength models. Unlike SSIM, which considers HVS characteristics, PSNR acknowledges every pixel within the scene with an equal visual importance. The SSIM measures remain consistent for each scenario, with only a slight decrease of 1% for the high strength watermarking model in most cases, due to the increase in watermark capacity.

Two frames accommodating indistinguishable objective readings for PSNR and SSIM do not necessarily radiate identical perceived visual media quality. To pro-



vide a complete media quality study, subjective testing analyzes the impact of each applied watermarking scheme on the overall perceived human viewing experience. Subjective analysis of the proposed model comprises of DSCQT and DSIST and is shown in Figure 5.9, for both blind and non-blind watermarking schemes. The top and bottom rows show subjective results for the blind and non-blind watermarking cases, respectively, whereas the left and right columns subsequently correlate to the DSCQT and DSIST evaluation tools. Consistent results are portrayed for both the blind and non-blind scenarios. For the DSCQT, the DCR measurements only deviate by approximately 1 unit when comparing the proposed and low strength embedding methodologies, suggesting a subjectively similar media quality. The high strength watermarking scheme shows a significantly higher subjective media quality degradation compared with the VA-based methodology. Similar outcomes are determined from the DSIST graphs, where the low and VA-based scheme both generate a similar mean opinion score 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.

From both objective and subject analysis, only minimal added visual distortion is perceived when comparing the low strength and VA-based methodologies. The proposed method successfully exploits visually uninteresting areas to mask extra embedded watermark information, in comparison to the low strength scheme.

### 5.1.5.3 Robustness

Results were generated to test the watermark ability to withstand intentional and non-intentional adversary attacks. Robustness against Joint Photographic Experts Group (JPEG) 2000 compression is shown in Figure 5.10 and Figure 5.11 for the non-blind and blind watermarking schemes, respectively, by plotting Hamming distance of the recovered watermark against the JPEG2000 compression ratio. 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

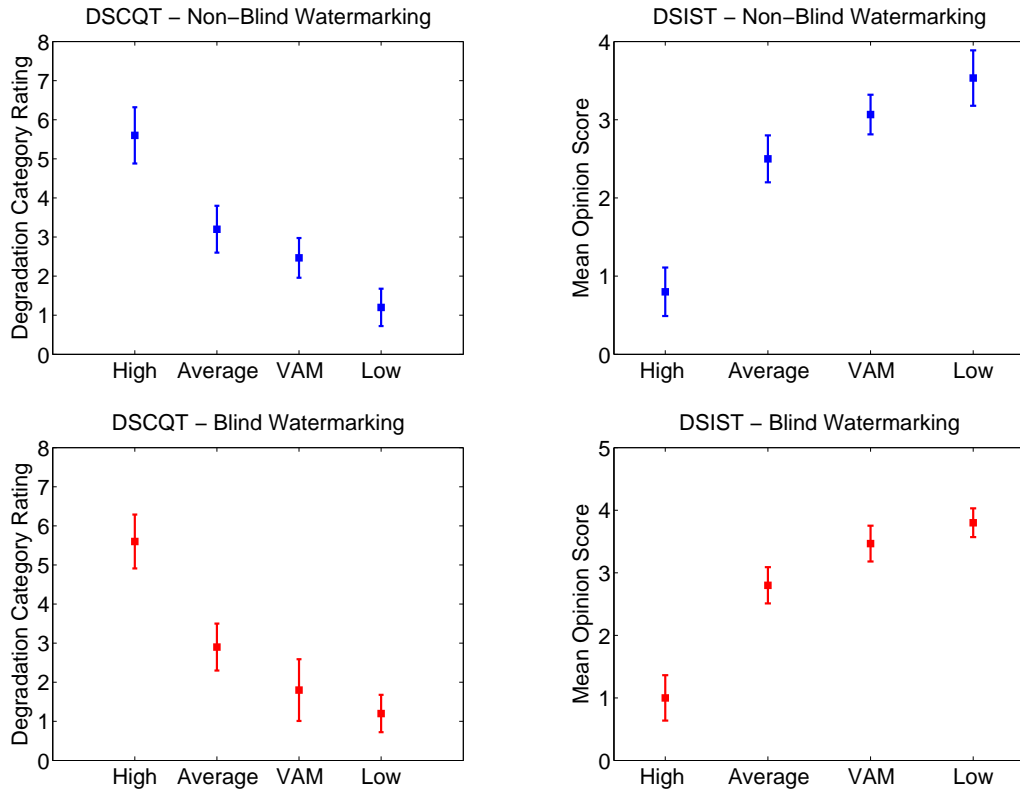


Figure 5.9: Subjective Image Watermarking Imperceptibility Testing - (top row) non-blind watermarking, (bottom row) blind watermarking.

proposed VA-based watermarking scheme, compared with the low strength watermark. Adversary filtering attacks, for each of the 3 scenarios, are simulated by convoluting the watermarked data with a filtering kernel, to distort any embedded information. Table 5.3 shows the watermark robustness against various low pass kernel types, namely: a 5x5 and a 3x3 mean filter, a 5x5 and a 3x3 median filter and a 5x5 Gaussian kernel. An increase in watermark robustness, ranging between 10% and 40%, is shown between the low strength and proposed method, for the various types of kernel. For both filtering attacks and JPEG2000 compression, a maintained or slight improvement within watermark robustness is seen in Figure 5.10, Figure 5.11 and Table 5.3 between the proposed VA-based technique compared using an average watermark strength.

The proposed VA-based method 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

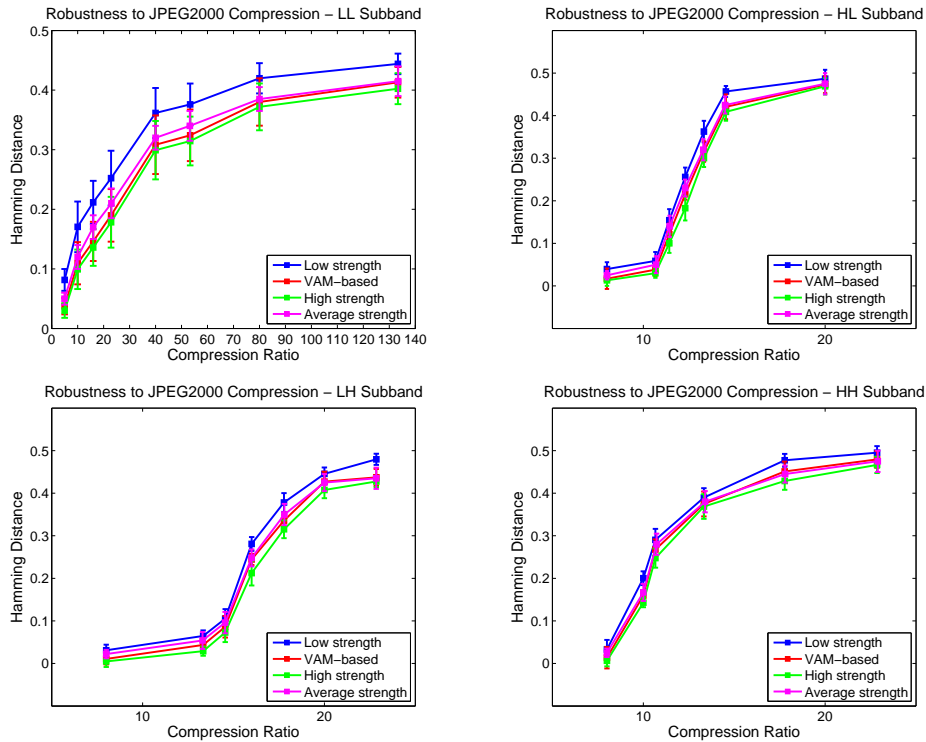


Figure 5.10: Robustness to JPEG2000 Compression - non-blind watermarking.

imperceptibility, verified by subjective and objective metrics in Section 5.1.5.2 and Section 5.1.5.3, deem the VA-based methodology highly suitable towards providing an efficient watermarking scheme.

## 5.2 VA-based Video Watermarking

By utilizing VA-based mechanisms to control the watermarking embedding strength, a comparable algorithm to VA-based image watermarking, as described in Section 5.1, can be implemented within the video domain. It is unfeasible to simply extend the previous VA-based image domain algorithm into a frame-by-frame video watermarking scheme, as temporal factors must first be considered within the video watermarking framework.

A viewer has unlimited time to absorb all information within an image, so potentially could view all conspicuous and visually uninteresting aspects in a scene.

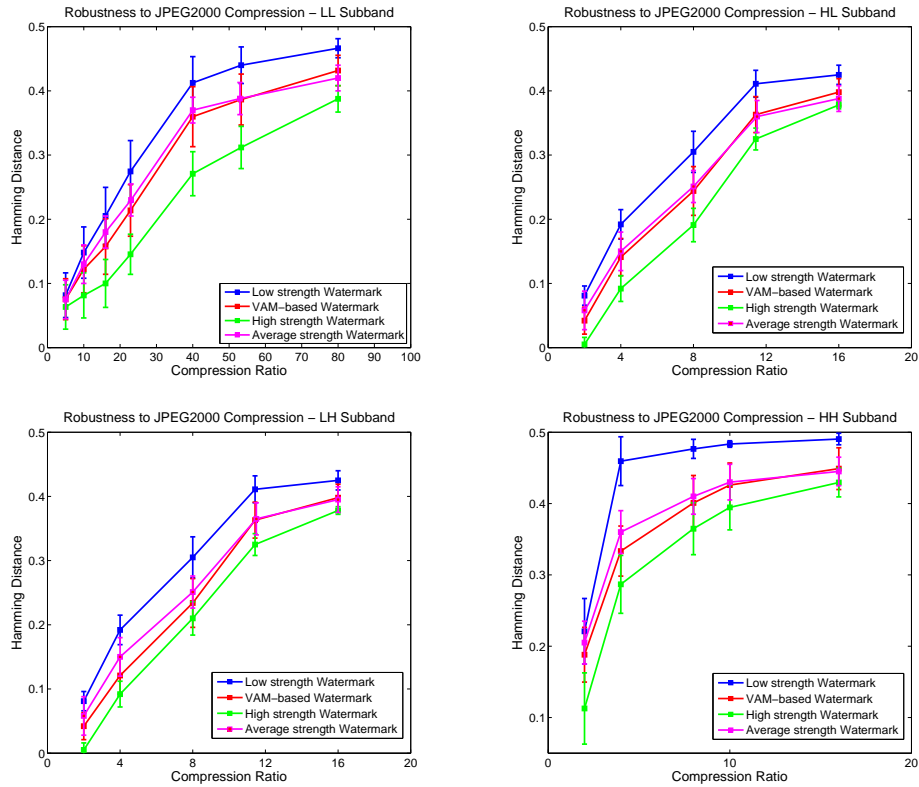


Figure 5.11: Robustness to JPEG2000 Compression - blind watermarking.

However, in a video sequence, the visual cortex has very limited processing time to analyse each individual frame. Human attention will naturally be diverged only towards visually striking regions. Frame cropping is a highly common procedure to eliminate visually redundant scene regions, while still retaining all worthwhile scene information [60]. Image cropping is a simple procedure and many readily, commercially available tools exist, such as Photoshop. Video frame cropping is a less practical option to potential adversaries. Attentive regions can switch spatial location during a sequence which highly limits any conceivable frame regions to be discarded. Consequently, VA-based watermarking algorithms are realistically more applicable within the video domain.

This section provides a solution towards blind and non-blind VA-based video watermarking. Blind watermarking schemes are a more viable solution, within the video domain, due to possible limited access obtaining the raw host media. The video saliency model described in Section 4.3 is utilized within the video watermarking framework to determine the watermarking embedding strength. Coin-

Table 5.3: Image Filtering Robustness.

LL Subband - non-blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.17 ± 0.03	0.13 ± 0.02	0.13 ± 0.02	0.06 ± 0.01
5x5 median	0.22 ± 0.02	0.16 ± 0.02	0.17 ± 0.02	0.07 ± 0.02
3x3 median	0.12 ± 0.03	0.09 ± 0.03	0.09 ± 0.02	0.06 ± 0.02
5x5 mean	0.18 ± 0.02	0.13 ± 0.02	0.14 ± 0.02	0.06 ± 0.01
3x3 mean	0.06 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.03 ± 0.00
LL Subband - blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.21 ± 0.02	0.18 ± 0.02	0.18 ± 0.02	0.15 ± 0.01
5x5 median	0.25 ± 0.04	0.18 ± 0.04	0.19 ± 0.03	0.11 ± 0.03
3x3 median	0.16 ± 0.02	0.11 ± 0.02	0.11 ± 0.02	0.09 ± 0.01
5x5 mean	0.23 ± 0.02	0.19 ± 0.02	0.19 ± 0.02	0.17 ± 0.01
3x3 mean	0.10 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.04 ± 0.00
HL Subband - non-blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.28 ± 0.03	0.24 ± 0.02	0.24 ± 0.02	0.19 ± 0.01
5x5 median	0.29 ± 0.02	0.21 ± 0.02	0.23 ± 0.03	0.17 ± 0.02
3x3 median	0.24 ± 0.02	0.19 ± 0.02	0.20 ± 0.02	0.15 ± 0.01
5x5 mean	0.27 ± 0.02	0.22 ± 0.02	0.21 ± 0.02	0.18 ± 0.02
3x3 mean	0.21 ± 0.01	0.17 ± 0.01	0.17 ± 0.01	0.14 ± 0.01
HL Subband - blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.28 ± 0.03	0.24 ± 0.02	0.24 ± 0.03	0.19 ± 0.01
5x5 median	0.29 ± 0.02	0.21 ± 0.02	0.21 ± 0.02	0.17 ± 0.02
3x3 median	0.24 ± 0.02	0.19 ± 0.02	0.20 ± 0.02	0.15 ± 0.01
5x5 mean	0.27 ± 0.02	0.22 ± 0.02	0.23 ± 0.03	0.18 ± 0.02
3x3 mean	0.21 ± 0.01	0.17 ± 0.01	0.18 ± 0.01	0.14 ± 0.01
LH Subband - non-blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.29 ± 0.02	0.23 ± 0.02	0.23 ± 0.03	0.19 ± 0.01
5x5 median	0.28 ± 0.02	0.21 ± 0.02	0.22 ± 0.02	0.17 ± 0.01
3x3 median	0.24 ± 0.02	0.20 ± 0.02	0.20 ± 0.02	0.14 ± 0.01
5x5 mean	0.28 ± 0.02	0.22 ± 0.02	0.23 ± 0.03	0.18 ± 0.02
3x3 mean	0.22 ± 0.01	0.18 ± 0.01	0.18 ± 0.01	0.14 ± 0.01
LH Subband - blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.40 ± 0.02	0.35 ± 0.03	0.36 ± 0.02	0.31 ± 0.02
5x5 median	0.38 ± 0.02	0.33 ± 0.03	0.33 ± 0.02	0.29 ± 0.02
3x3 median	0.34 ± 0.01	0.29 ± 0.02	0.30 ± 0.03	0.24 ± 0.01
5x5 mean	0.39 ± 0.02	0.36 ± 0.02	0.36 ± 0.02	0.31 ± 0.02
3x3 mean	0.34 ± 0.01	0.29 ± 0.01	0.29 ± 0.01	0.23 ± 0.01
HH Subband - non-blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.38 ± 0.03	0.35 ± 0.03	0.35 ± 0.02	0.32 ± 0.02
5x5 median	0.36 ± 0.03	0.34 ± 0.03	0.34 ± 0.02	0.33 ± 0.02
3x3 median	0.23 ± 0.02	0.22 ± 0.03	0.22 ± 0.01	0.20 ± 0.02
5x5 mean	0.38 ± 0.02	0.35 ± 0.04	0.36 ± 0.02	0.34 ± 0.02
3x3 mean	0.23 ± 0.02	0.21 ± 0.03	0.22 ± 0.01	0.20 ± 0.01
HH Subband - blind				
Filtering	Low	Proposed	Average	High
Gaussian	0.43 ± 0.02	0.40 ± 0.02	0.40 ± 0.02	0.38 ± 0.01
5x5 median	0.41 ± 0.03	0.39 ± 0.02	0.40 ± 0.02	0.38 ± 0.02
3x3 median	0.28 ± 0.02	0.27 ± 0.03	0.27 ± 0.02	0.25 ± 0.02
5x5 mean	0.42 ± 0.03	0.40 ± 0.03	0.40 ± 0.03	0.39 ± 0.02
3x3 mean	0.30 ± 0.02	0.28 ± 0.03	0.28 ± 0.02	0.26 ± 0.01

ciding with the previous video VA model, watermark data is embedded within the 2D+t wavelet domain as outlined in Section 4.3.1. Figure 5.12 shows the complete overall VA-based watermarking system.

## 5.2.1 Experimental Results

A series of experimental results are generated, analysing both watermark robustness and imperceptibility. Complimenting the image domain watermarking,

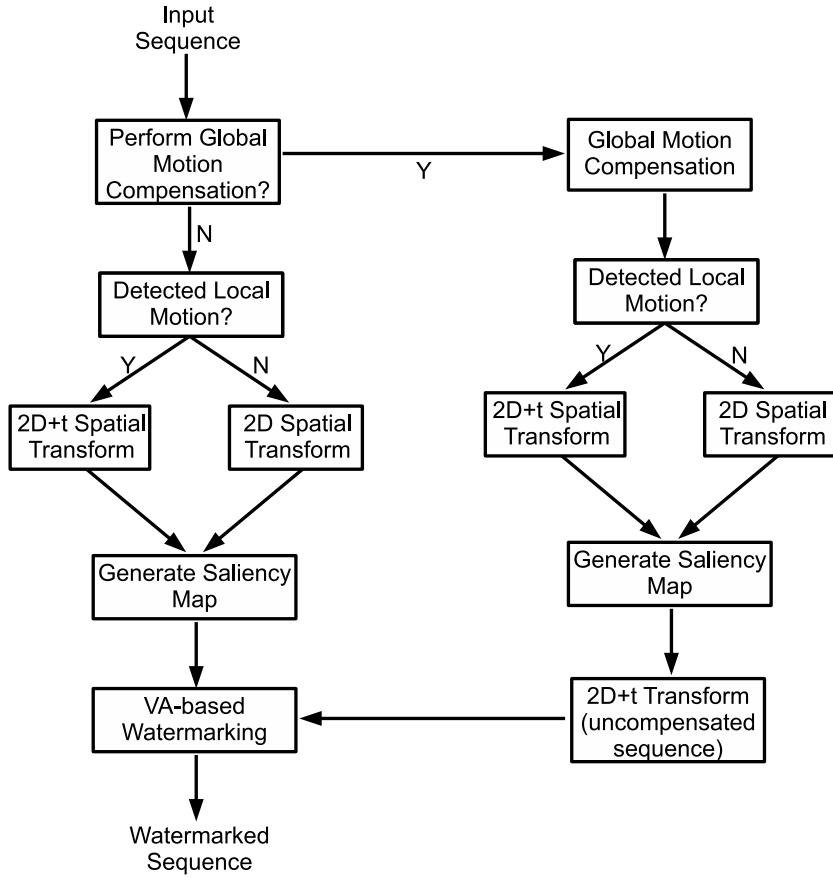


Figure 5.12: Overall VA-based Video Watermarking Scheme - Block Diagram.

objective and subjective evaluation tools are enforced to provide a comprehensive embedding distortion measure. Robustness against H.264/AVC compression is provided, as common video attacks comprise of platform reformatting and video compression. Common test set parameters, used throughout all performed experiments, include: the orthogonal D4 wavelet for 3 levels of 2D spatial decomposition and 1 level of temporal Haar decomposition, implemented using the video test sequences described in Section 3.4.2.2. An  $\alpha_{max}$  and  $\alpha_{min}$  approximating a PSNR of 35dB and 40dB, respectively, is utilised by applying Equation (5.4) and Equation (5.5). Consistent with the image watermarking analysis in Section 5.1, 4 scenarios of varied watermark embedding strength are also implemented for video watermarking comparison:

- 1) a constant  $\alpha_{min}$  throughout the entire sequence;
- 2) the proposed VA-based  $\alpha$  strength;
- 3) a constant average watermark strength,  $\alpha_{ave}$ ; and
- 4)  $\alpha_{max}$  used entirely throughout the video.



a)



b)



c)



d)

Figure 5.13: 'Soccer' sequence video watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.

### 5.2.1.1 Imperceptibility

The PSNR and SSIM values corresponding to each video watermarking scenario are shown in the top 2 rows in Table 5.4. The performance of both SSIM and PSNR are rank ordered in terms of highest imperceptibility. Naturally, the methodology ranking list is: low strength embedding > VA-based algorithm / average strength embedding > high strength embedding. To provide a complete media quality investigation, the VQM is also provided for the video watermarking



a)



b)



c)



d)

Figure 5.14: 'Stefan' sequence video watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image.

algorithms as described in Section 3.4.1. The bottom rows in Table 5.4 show the VQM for both blind and non-blind video watermarking schemes, for each of the 4 scenarios. Figure 5.15 shows the subjective test results for DCQST and DSIST averaged over 4 video test sequences. For each of the blind and non-blind watermarking cases in both the objective and subjective visual quality evaluation, the low strength watermark and VA-based watermarking sequences yield similar visual quality, whereas the high strength embedded sequence appears severely more distorted. Low strength watermarking provides a high imperceptibility but is fragile as discussed in Section 5.2.1.2. Coherent with the subjective test-



Table 5.4: PSNR, SSIM and VQM average of 4 video sequences for blind and non-blind watermarking.

	<b>Low Strength</b>	<b>Proposed</b>	<b>Average Strength</b>	<b>High Strength</b>
<b>non-blind method</b>				
PSNR	40.15 $\pm$ 0.80	37.39 $\pm$ 0.87	37.47 $\pm$ 0.76	34.93 $\pm$ 0.73
SSIM	0.99 $\pm$ 0.00	0.97 $\pm$ 0.00	0.98 $\pm$ 0.00	0.95 $\pm$ 0.01
VQM	0.10	0.15	0.18	0.25
<b>blind method</b>				
PSNR	40.23 $\pm$ 1.03	36.80 $\pm$ 1.02	37.20 $\pm$ 0.92	34.85 $\pm$ 0.90
SSIM	0.99 $\pm$ 0.00	0.98 $\pm$ 0.00	0.98 $\pm$ 0.00	0.96 $\pm$ 0.01
VQM	0.08	0.13	0.15	0.22

ing of VA-based image watermarking, shown in Figure 5.9, the visual quality of the proposed watermarking scheme is higher than a constant average watermark strength.

Frames taken from the test sequences are shown in Figure 5.13 and Figure 5.14. For a comprehensive incite into the subjective media quality, the results are best displayed as a full sequence rather than a still image. Full supporting video watermarking sequences can be viewed from the website:

<http://svc.group.shef.ac.uk/va-video-wm.html>.

### 5.2.1.2 Robustness

Video reformatting and compression is a frequent and typically unintentional adversary attack, hence watermark tolerance for H.264/AVC compression is calculated. Robustness against H.264/AVC compression for both non-blind and blind video watermarking schemes are shown in Figure 5.16a) and Figure 5.16b) respectively. For simulating the watermark robustness, 5 constant Quantisation Parameter (QP) values are implemented to compress the high strength, average strength, VAM-based and low strength test sequences. In both scenarios from the graphs, the VAM-based, proposed methodology shows an increase in robustness compared with the low strength watermark counterpart. From the graphs in Figure 5.16, Hamming distance reductions up to 39% for the non-blind case and 22% for the blind case are possible, when comparing the low and VAM-based

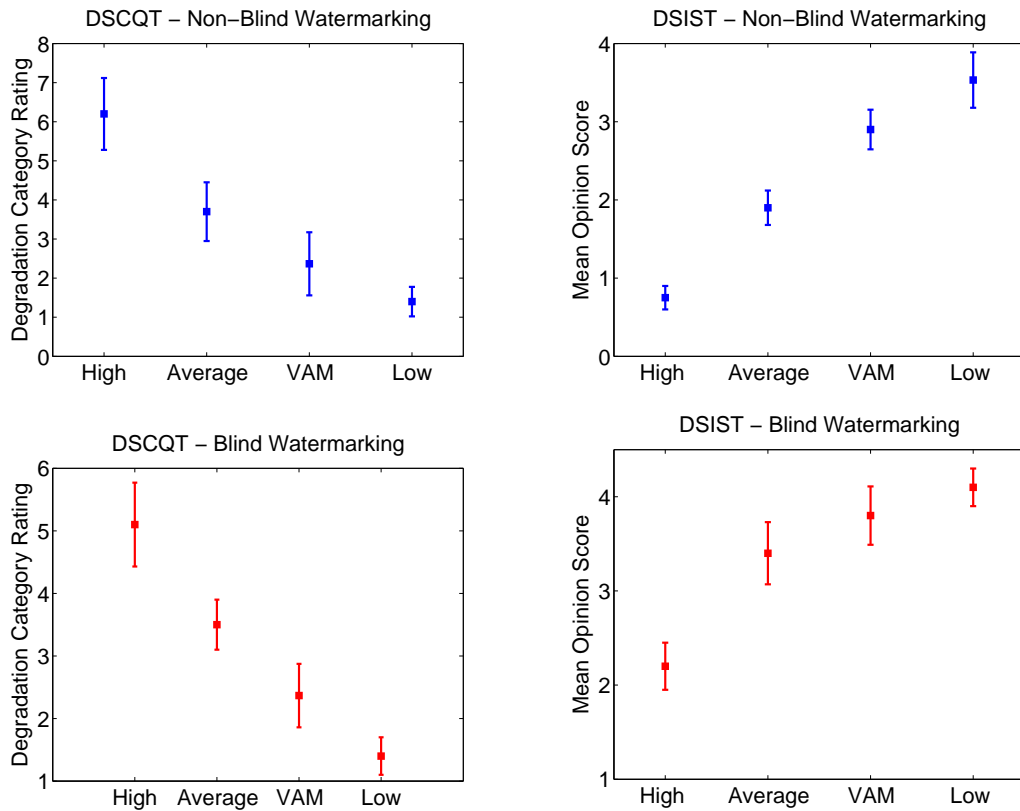


Figure 5.15: Subjective Video Watermarking Imperceptibility Testing - (top row) non-blind watermarking, (bottom row) blind watermarking.

models. Naturally, the high strength watermarking scheme portrays a strong Hamming distance but is highly perceptible, as described previously. The proposed watermarking scheme has a slight increased robustness towards H.264/AVC compression, as shown in Figure 5.16, when compared against a constant average strength watermark.

It is of suitable note that for a constant QP value, the compression ratio is inversely proportional to the increase in watermark strength, i.e as the watermark strength increases, the overall compression ratio decreases due to the extra watermark capacity.

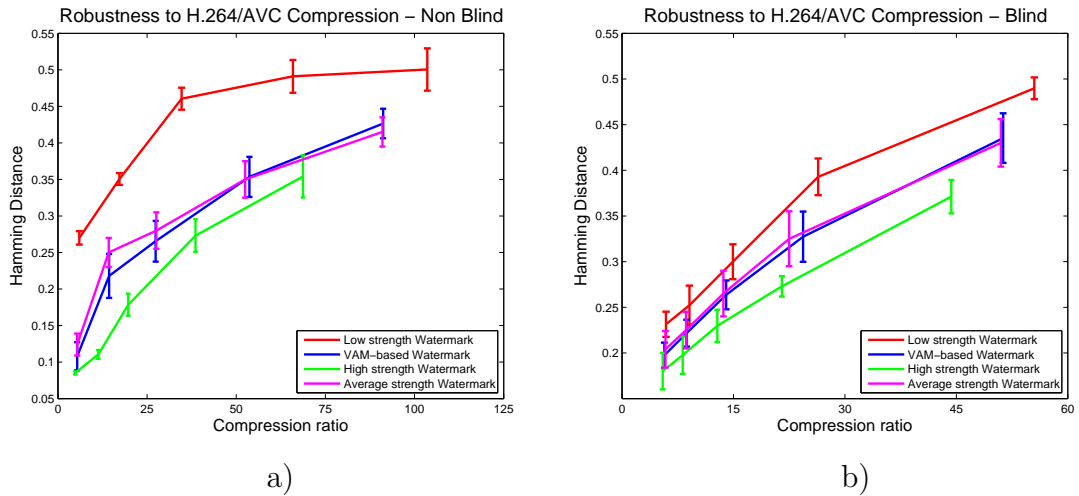


Figure 5.16: Robustness to H.264/AVC compression - average of 4 video sequences: a) non-blind watermarking b) blind watermarking.

### 5.3 Conclusions

This chapter implements the VA model within the watermarking framework to determine a value for watermark embedding strength. The proposed algorithms can embed more information into visually uninteresting areas within the host media, determined by the VAM, while maintaining the subjective scene quality. The novel methodologies provide an increased robustness to various natural processing and filtering attacks, while minimally affecting the media visual quality as verified by both subjective and objective evaluations. The proposed algorithm shows up to 39% and 22% improvements in Hamming distance against H.264/AVC compression for the non-blind and blind scenarios, respectively, and robustness against JPEG2000 is increased by up to 25%, compared with a constant low strength watermark. The proposed model addresses both non-blind and blind watermark extraction scenarios. VA plays a crucial role when the primary cortex must select attentive scene regions within a limited time constraint. For this reason, visually inattentive watermarking algorithms are ideally suitable for video sequences or for static frames grouped in a slide show.

The algorithms presented within this chapter are provided directly from within the uncompressed domain. However, the uncompressed video is not always available, therefore, a compressed sequence must firstly undergo full bitstream decod-

ing prior to applying the VA-based watermarking algorithm. By incorporating a VA-based watermarking scheme directly from within the compressed domain, large computational savings will be possible. The next chapter focuses on compressed domain saliency algorithms, in particular the HEVC codec.

# Chapter 6

## HEVC Domain Saliency Estimation

The computation of VA is an exhaustive procedure to locate conspicuous regions within a frame, which contrast with the surrounding background. In this chapter a unique algorithm is proposed to estimate visual saliency in the compressed domain using coding decisions, from HEVC encoded video sequences. By exclusively combining data obtained from the coding unit structure, intra mode block predictions, MV estimation and the residual data, a visual saliency approximation is obtained. The proposed model can accurately detect salient regions without the need to fully decode the HEVC bitstream. Experimental results show the proposed algorithm compares positively against multiple methods in the literature, highlighting accurate saliency detection with minimal time additions to the video coding computation. The new methodology can provide aid to a wide variety of fields such as advertising, watermarking, video editing and spatial-temporal adaptation.

### 6.1 Introduction

As described in Section 2.3, most state-of-the-art saliency models [53–55] rely upon the early feature integration theory [51], which is an exhaustive procedure.

By incorporating a saliency model from directly within the video codec, significant computational savings are attainable as the majority of stored video sequences exist in compressed format. Visually stimulating regions arise from the presence of conspicuous scene abnormalities, which can originate from multiple feature mechanisms. Neural stimuli are extremely sensitive to contrasts within scene brightness, orientation and motion, which largely contribute to low level VA. [51] This is highly relatable within the modern video codecs, such as, Advanced Video Coding standard (H.264/AVC) [115] and HEVC [3]. In the HEVC codec, much of the coding gains have been obtained by using extensive analysis of frames for optimising the predictions. HEVC dissects each video frame to determine various coding decisions based upon the analysis of the frame characteristics. Numerous coding decisions for intra and inter-frame coding include: a flexible quad-tree block size partitioning scheme based on the regions homogeneousness, a scene orientation approximation using ADI prediction and scene motion estimation through MV's.

The aim of this chapter is to explore such coding decisions and other data available within the HEVC bitstreams to estimate visual saliency in frames. Figure 3.13 shows any video sequence can be encoded incorporating either intra-only or a combination of intra and inter-frame prediction. The intra and inter-frame as well as coding decisions and prediction errors (usually known as residuals) are explored to propose an HEVC compressed domain saliency model. The saliency maps can be generated without fully decoding the HEVC bitstream, which is highly advantageous. For existing methodologies [54, 55, 67, 69, 139, 140], the HEVC bitstream data would first have to be fully decoded before a saliency model is applied. Therefore, the proposed model is highly suitable towards any application or device limited by low computational complexity, such as mobile phones.

The rest of the chapter is arranged as follows: Section 6.2 provides a saliency model for an intra-only encoded sequence, where as Section 6.3 implements inter-frame saliency prediction. The combined saliency model is derived in Section 6.4, with Section 6.5 describing how encoder parameters influence the model. Finally, Section 6.6 provides the model evaluation.

## 6.2 Intra-Frame Saliency Estimation

To establish the prospect of a visually salient block, numerous features must be exploited from within the HEVC codec. The main methodology behind intra-frame saliency estimation comprises of 3 elements, namely: block structure, intra mode difference and residual energy. These features are highly relatable to visual stimuli mechanisms such as intensity and orientation contrast. The block structure is complementary to salient features as the split flag in Figure 3.14a) is signalled based upon the presence of a high colour or intensity contrast. Intra modes determine a suitable orientation approximation for the scene. By locating the presence of inconsistent variations within the determined intra modes, conspicuous regions can emerge. Large residuals arise when an accurate block prediction cannot be formulated, usually from the presence of peculiar block patterns. By pinpointing these block abnormalities, potential salient regions can be detected.

Firstly, the relationship between various coding modes is established with visual saliency mechanisms using the MSRA database described in Section 3.4.2. Pixel values given a particular block partition size, intra mode difference or residual magnitude are compared against the MSRA ground truth saliency maps. Section 6.2.1, Section 6.2.2 and Section 6.2.3 analyse the saliency relationship with each of the coding modes, whereas the combined intra-frame model is proposed in Section 6.2.4.

### 6.2.1 Block Size

Based on the PU block structure, a saliency estimation can be made. Figure 6.1a) and Figure 6.1b) show the original frame and corresponding PU block structure, respectively. A relationship between PU block size,  $s$ , and VA is derived from the results shown in Figure 6.2a). Pixels located within the manually segmented ROI are compared with the partitioned block dimensions. A clear correlation can be established between decreasing block size and salient regions, from which the block size saliency probability, can be determined. The block size partitioning is derived by the level of homogeneous activity within each block, as greater

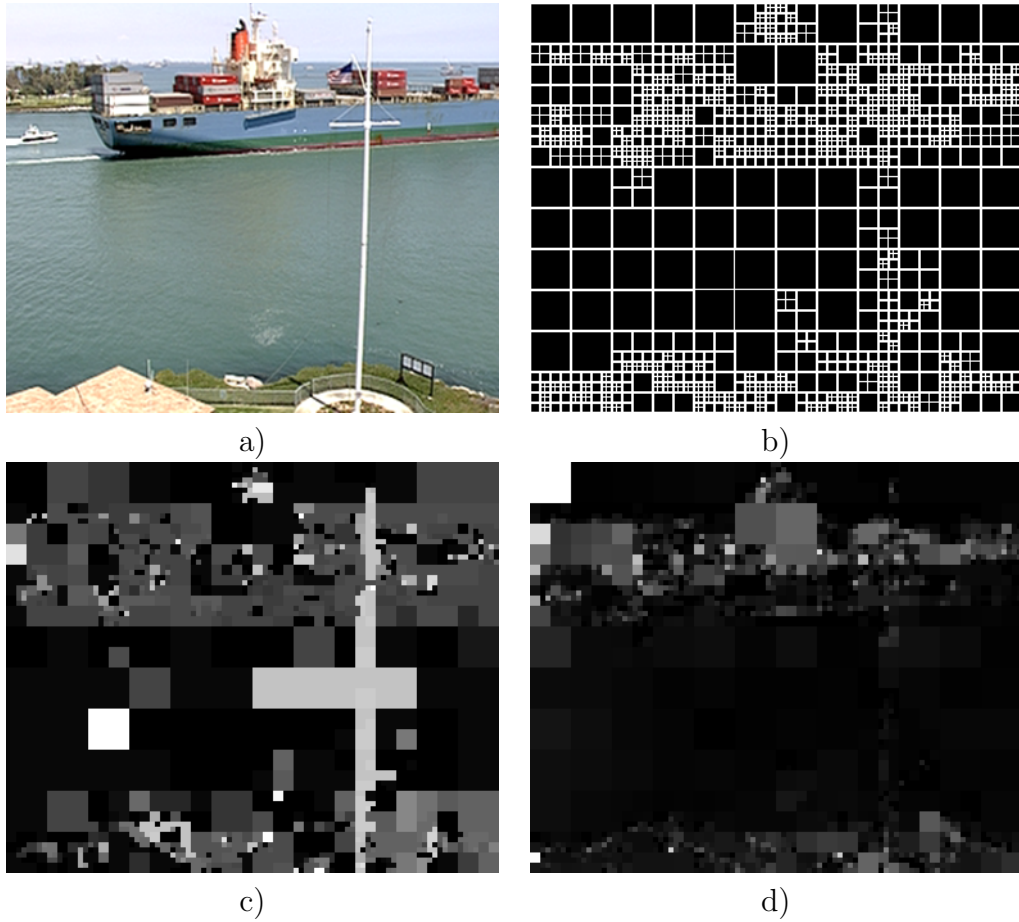


Figure 6.1: a) Original frame b) PU block structure c) ADI prediction mode number (0-34) d) Average block residual.

partitioning occurs under the influence of an overall high block variance. Homogeneous regions within a frame portray low saliency characteristics, therefore, larger block sizes are less likely to contain visual conspicuousness in comparison to finely partitioned areas.

### 6.2.2 Intra mode differences

By modelling the inconsistencies within angular intra prediction modes, an estimation to locate salient features can be performed. For example, a textured background, solely predicted from ADI modes 16 and 17, could contain prominent foreground objects depicted from a various combination of directional modes. By



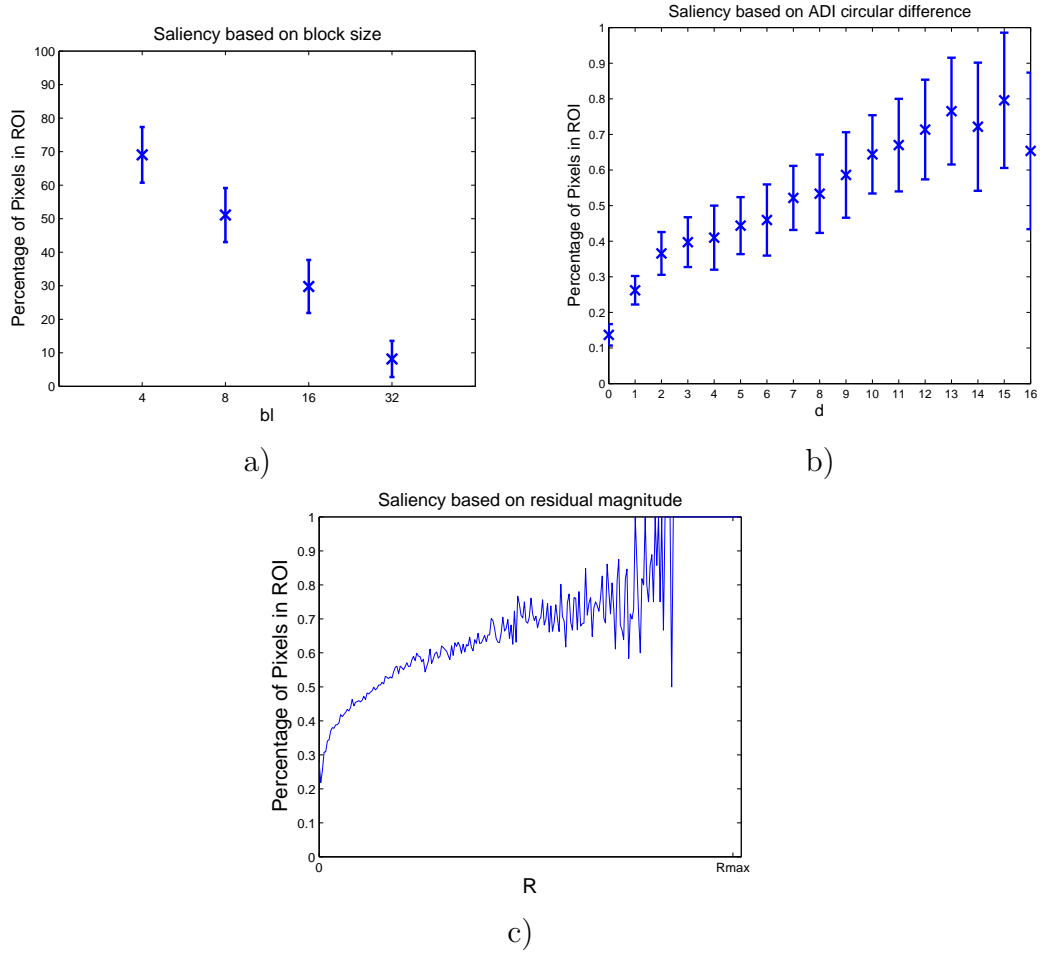


Figure 6.2: Saliency correlation with a) block size b) intra mode difference c) average block residual.

locating mode irregularities, salient conspicuous regions can be identified. Modes 2 and 34 are highly correlated so simple difference calculation between adjoining partitions is insufficient. In Figure 6.3 each prediction mode is re-mapped to a circular plain, where modes 34 and 2 are adjacent.

$m_d$  is defined as the absolute difference between consecutive modes  $m1$  and  $m2$ , i.e  $m_d = |m1 - m2|$ . Let  $\ominus$  represent the circular difference between 2 adjacent modes, which is calculated by:

$$m1 \ominus m2 = \begin{cases} m_d & \text{if } m_d \leq M/2 \\ M - m_d & \text{if } m_d > M/2, \end{cases} \quad (6.1)$$

where  $M$  is the maximum range of possible modes, i.e., 34-2. Modes 0 and 1

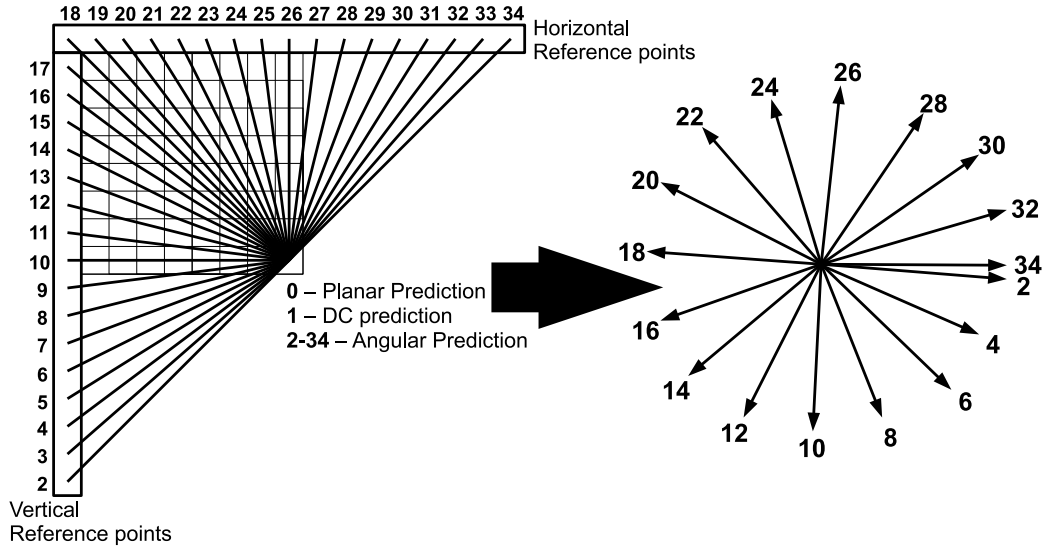


Figure 6.3: Mapped ADI circular difference modes.

are omitted from the algorithm as they bear negligible resemblance towards angular prediction. Figure 6.1c) shows an intra mode map, derived from the ADI prediction mode choices, described in Figure 6.3. The methodology determines the circular difference between the horizontal and vertical consecutive modes, at block boundaries. If the intra mode map is given by  $In_{x,y}$ , with  $x$  and  $y$  describing the spatial coefficient location within the frame, the horizontal and vertical circular differences,  $D_{x,y}^H$  and  $D_{x,y}^V$ , at the block boundaries can be determined by:

$$\begin{aligned} D_{x,y}^H &= |In_{x,y} - In_{x+1,y}| + |In_{x,y} - In_{x-1,y}|, \\ D_{x,y}^V &= |In_{x,y} - In_{x,y+1}| + |In_{x,y} - In_{x,y-1}|. \end{aligned} \quad (6.2)$$

The block boundary differences are integrated over  $bl$ , in both horizontal and vertical directions, before linear combination into the absolute circular difference map,  $d$ , by Equation (6.3):

$$d = (\int_0^s D_{x,y}^H dx + \int_0^s D_{x,y}^V dy)/4. \quad (6.3)$$

A correlation linking ADI prediction mode circular differences against ROI is shown in Figure 6.2b). The graph shows a positive relationship between the location of high intra mode circular differences and salient data. Previous studies

show neural stimuli are extremely sensitive towards orientation contrast creating visual saliency [51, 54]. The ADI prediction determines an overall suitable orientation which best defines each block, from which highly contrasting adjacent modes can decipher salient portions within the frame. Likewise, an abundance of neighbouring modes of similar orientation are highly probable to part of a common visually uninteresting region or background. Figure 6.4 demonstrates the ADI prediction of an input frame and corresponding  $d$ .

### 6.2.3 Residual Data

Residual data is the prediction error between the original frame and the block estimation, so low residual values arise when accurate block estimations are made. The relationship between the quantised residual energy and visual saliency is explored, to determine a block saliency prediction based upon the residual magnitude,  $R$ . Figure 6.1d) shows the average quantised residual energy present within each block from an encoded frame. The residual energy for the entire block is generated from the DC coefficient within the transform, which determines the average residual energy in the untransformed domain. Figure 6.2c) shows the positive correlation depicting the normalised residual magnitude of each block with the ROI, within the range  $0 - R_{max}$ , where  $R_{max}$  is the maximum possible residual value. From the graph, partitioned blocks containing a higher residual magnitude are more likely to be part of a visually salient region. High residual energy arises from large inaccuracies between the intra mode prediction and the original frame. These imprecisions can originate from obscure, salient regions, which are difficult to predict. Therefore it is highly probable that if a block contains large residual data it is also very likely to be visually salient.

### 6.2.4 Intra Saliency Map Generation

Criteria for predicting visually salient blocks within the HEVC codec, can be determined. There is a high probability a block will be visually salient if it: is partitioned into a small size, accommodates a high quantised residual magnitude and contains large orientation inconsistencies between the surrounding prediction

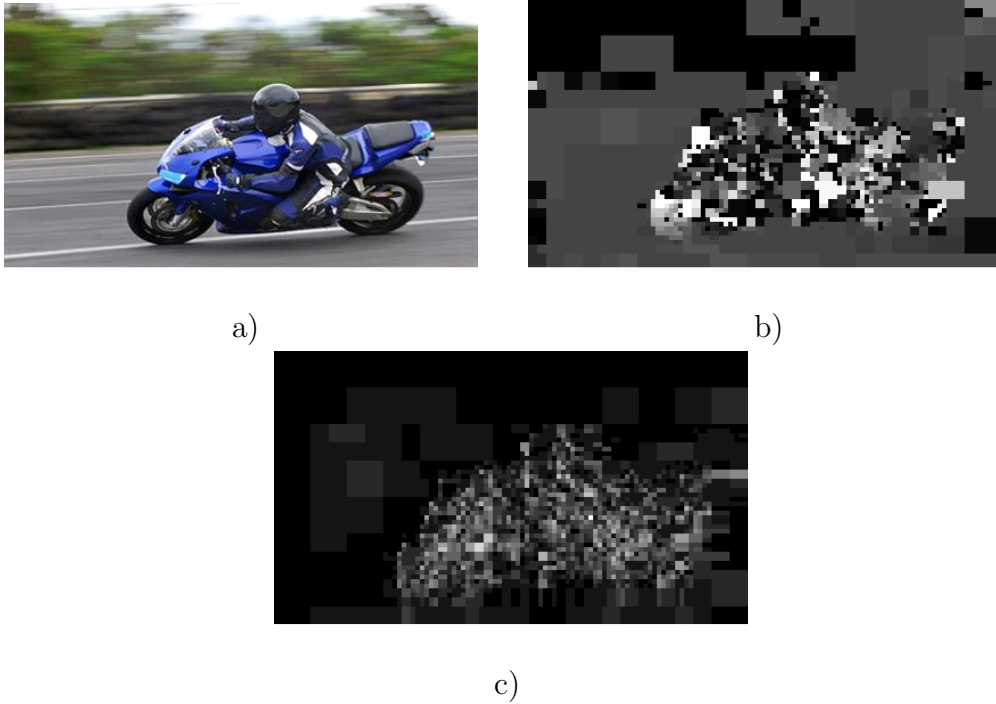


Figure 6.4: Intra mode circular difference a) Original frame b) ADI intra mode prediction c) Intra mode circular difference.

modes.

Adjacent intra mode circular differences and residual data portraying similar saliency characteristics are generically classified into a smaller number of bins. A local threshold,  $T_{res}$ , classifies the residual magnitude into 2 partitions based upon the saliency value.  $T_{res}$  is based upon histogram analysis of the residual data, shown in Figure 6.5a), which is taken from a typical frame within the MSRA database.  $T_{res}$  divides the majority of coefficients having near zero residual magnitude, which also portray a very low saliency probability as shown in Figure 6.5b). Generic partitioning, thresholded by  $T_{res}$ , is shown in Figure 6.5c), giving low and high probabilities of 0.26 and 0.68, respectively, from  $T_{res} = R_{max} * 0.1$ . The generic partitioning for residual magnitude is defined mathematically in Equation (6.4) by:

$$R' = \begin{cases} Low & \text{if } R < T_{res}, \\ High & \text{if } R \geq T_{res}, \end{cases} \quad (6.4)$$

where  $R'$  is the thresholded residual magnitude.

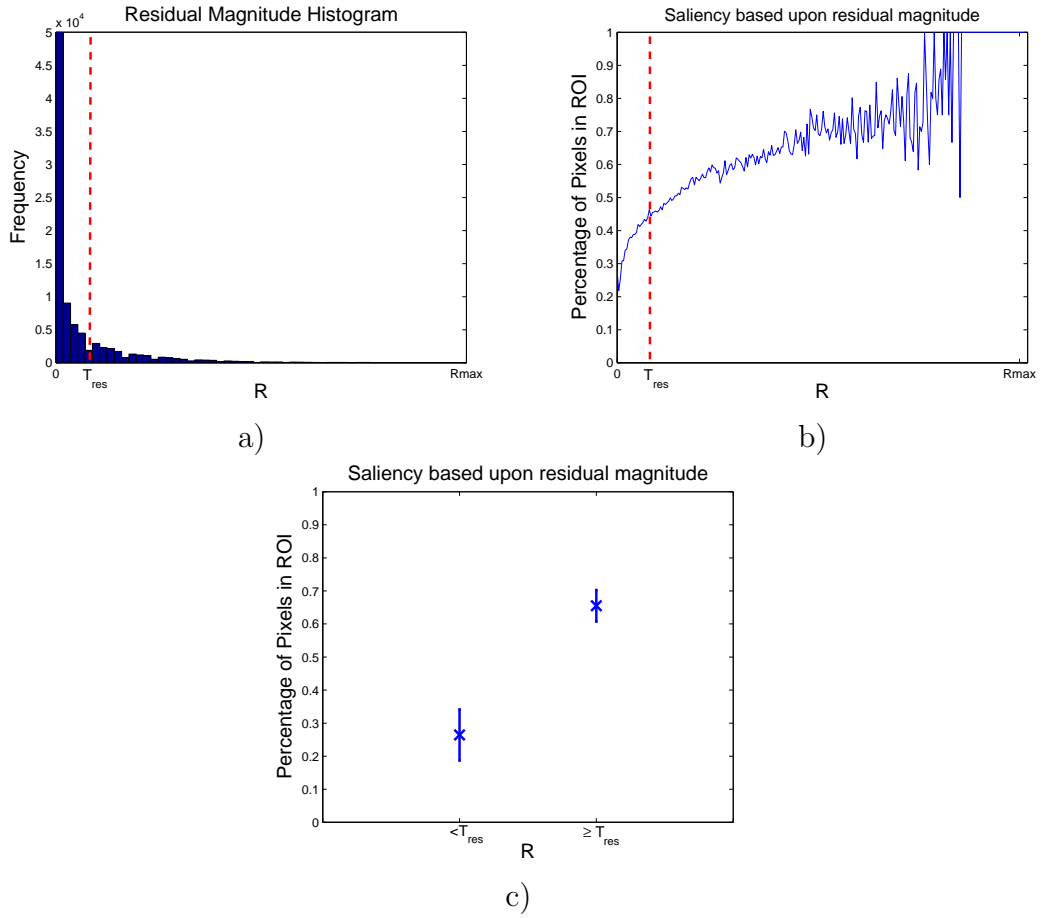


Figure 6.5: a) Histogram of residual magnitudes b) Residual magnitude threshold location c) Thresholded residual magnitude.

Histogram analysis defines logical partitions for  $d$  into 2 saliency levels by a threshold,  $T_{int}$ , as shown in Figure 6.6a) and Figure 6.6b). Setting a threshold of  $T_{int} = 2$  is logical as it segments a large quantity of coefficients portraying a low saliency probability. Figure 6.6c), shows the low and high thresholded intra mode difference probabilities of 0.19 and 0.62, respectively. The intra mode difference partitioning is calculated in Equation (6.5) by:

$$d' = \begin{cases} Low & \text{if } d < T_{int}, \\ High & \text{if } d \geq T_{int}, \end{cases} \quad (6.5)$$

where  $d'$  is the thresholded absolute circular difference data.

The conditional probability a partitioned block is salient,  $P(Sal)$ , is calculated

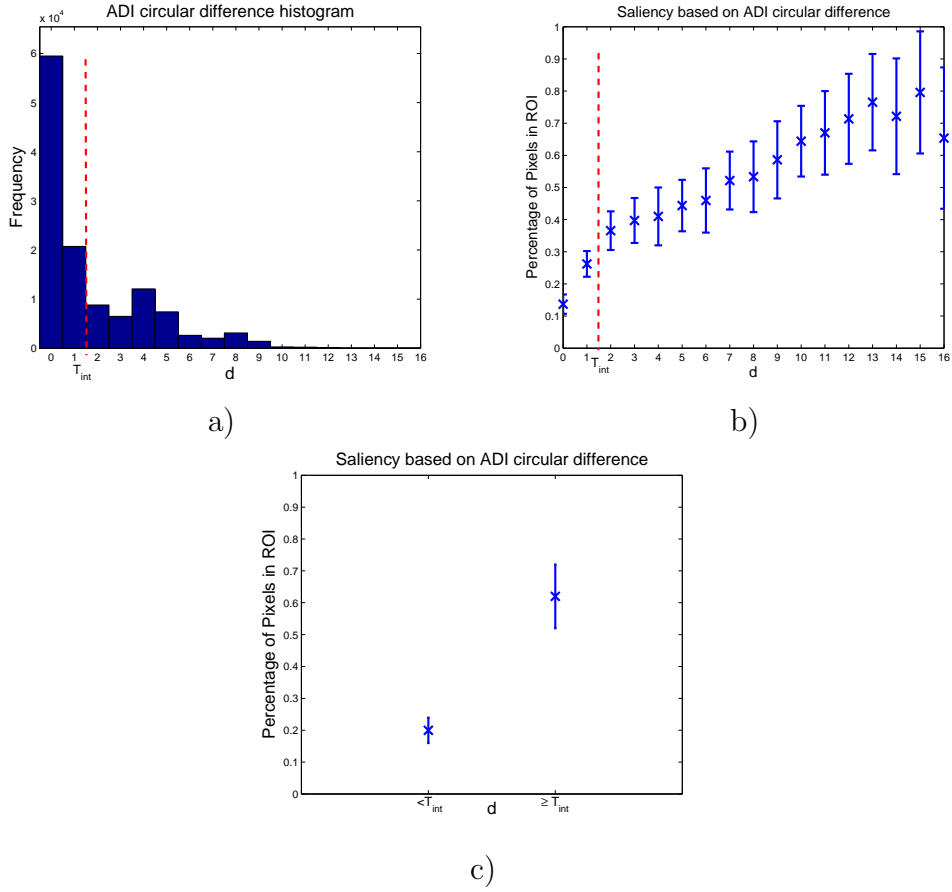


Figure 6.6: a) Histogram of intra mode differences b) Intra mode difference threshold location c) Thresholded intra mode difference.

given the PU block size, residual and intra mode difference,  $P(Sal|(s \cap R' \cap d'))$ , yielding 16 possible outcomes. The 16 combinations occur resultant from combining the 4  $s$  levels and 2 different values for  $R'$  and  $d'$ . Figure 6.7 shows the saliency estimation can range between  $0.03 \geq P(Sal|(s \cap R' \cap d')) \leq 0.85$ .

A consequent morphological non-linear opening and closing filter [141] is applied to the final saliency map to distort gaps defined by adjacent block partitions, as any region enclosed by high salient activity is also probable to draw human gaze. The kernel size,  $\psi_{sz}$ , is based upon the frame resolution as substantial filter dimensions are required to connect gapping regions without completely distorting the overall saliency shape. 3 values for  $\psi_{sz}$ , based upon the frame dimensions,  $D_X$  and  $D_Y$ , are shown in Figure 6.8. Figure 6.8c) highlights a small kernel size of  $\psi_{sz} = \frac{D_X + D_Y}{80}$ , where consequent gaping holes between salient regions exist

due to the inadequate kernel size. A large kernel size,  $\psi_{sz} = \frac{D_x+D_y}{10}$ , is shown in Figure 6.8d) and completely distort the spatial structure of any salient elements. A suggested  $\psi_{sz} = \frac{D_x+D_y}{30}$  is shown in Figure 6.8e). Figure 6.9 shows the ROC curves over the entire 1000 frame database and gives AUC values of 0.861, 0.884 and 0.867 for the small, suggested and large  $\psi_{sz}$ , respectively.

Equation (6.6) mathematically describes the final intra mode saliency prediction,  $S_{T1}$ , by:

$$S_{T1} = \psi.(P^{(x,y)}(Sal|(s \cap R' \cap d'))), \quad (6.6)$$

where  $\psi$  is the morphological filtering operation. The overall system diagram is shown in Figure 6.10.

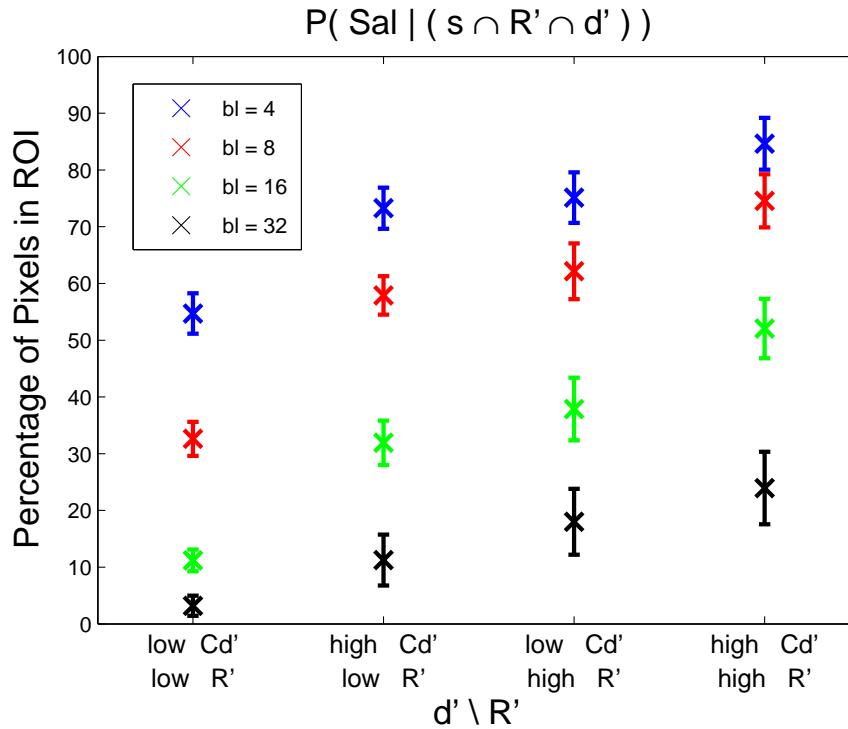


Figure 6.7: Graphs showing  $P(Sal|(s \cap R' \cap d'))$  for all 16 possible outcomes.

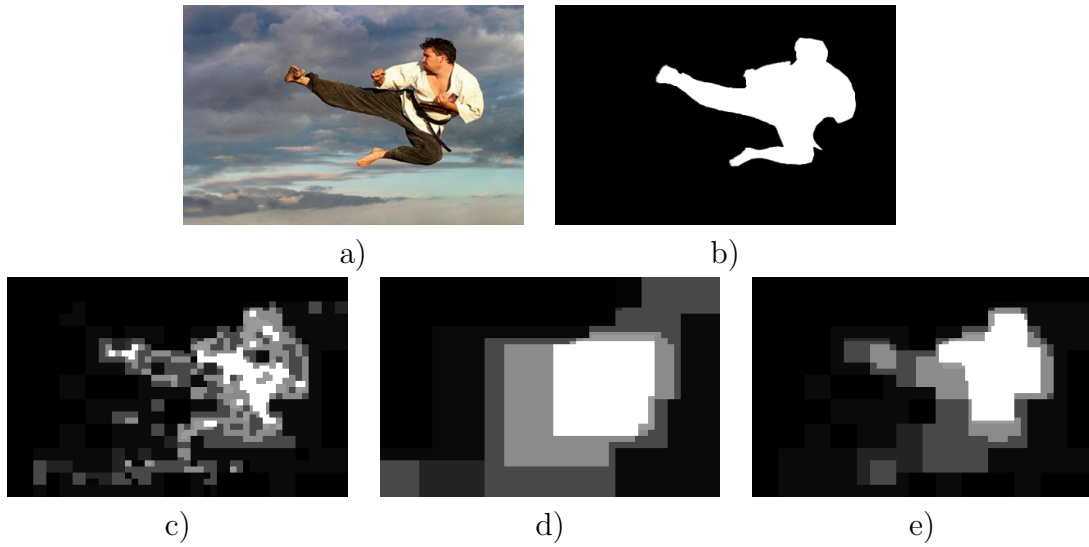


Figure 6.8: Morphological filtering for a 288x416 resolution frame - a) Original frame b) Ground truth frame c) Small kernel size d) Large kernel size e) Suggested kernel size.

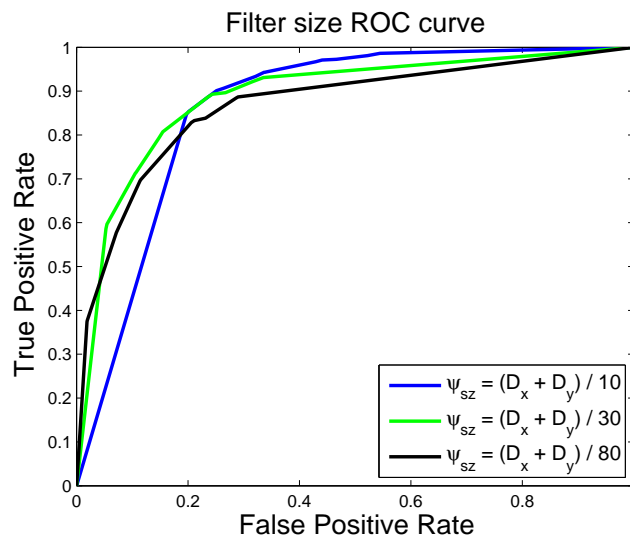


Figure 6.9: ROC curve comparing morphological kernel size.

### 6.3 Inter-Frame Saliency Estimation

The inter-frame saliency algorithm comprises of 3 elements, namely: quantised motion block residual data, motion vector difference magnitude and inter mode PU partition size, which are shown in Figure 6.11. These features are directly



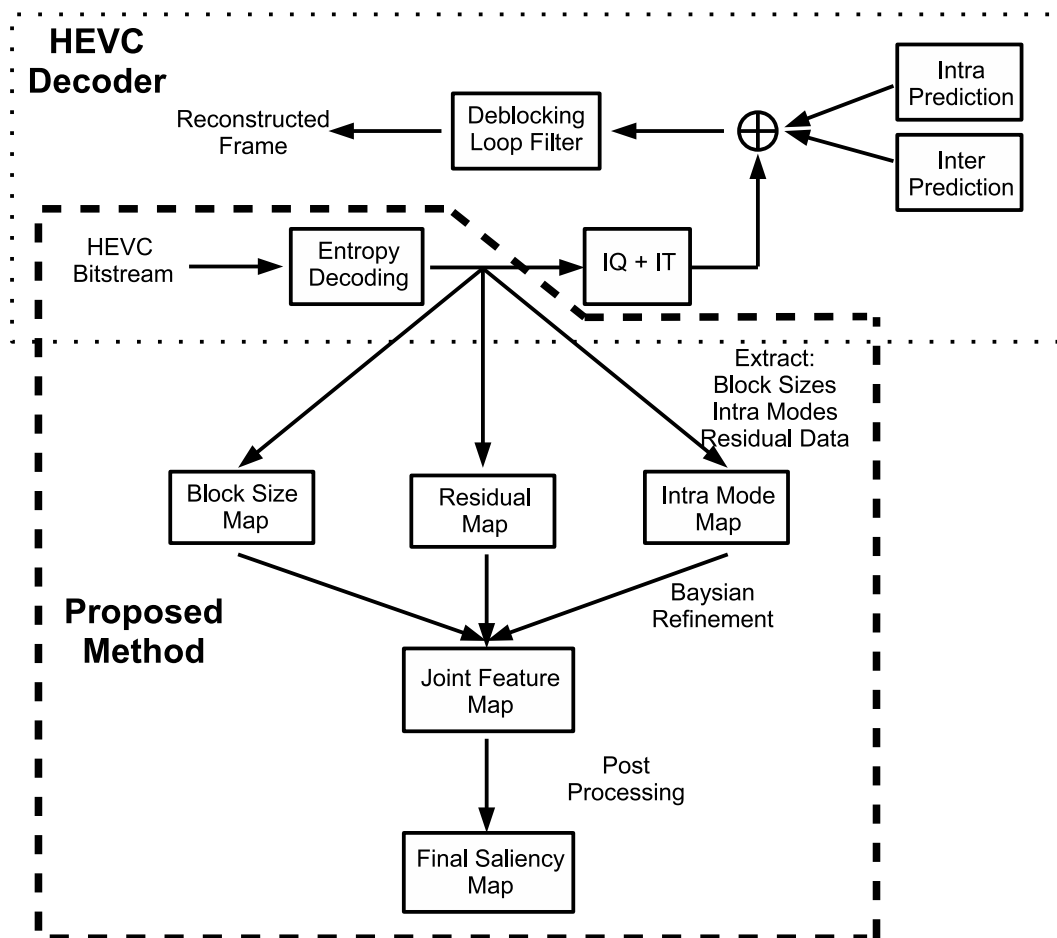


Figure 6.10: Overall saliency model diagram for an intra encoded frame.

extractable from within the HEVC decoder and exhibit highly correlatable characteristics with neural selectivity. The remainder of this subchapter is arranged as follows: Section 6.3.1, Section 6.3.2 and Section 6.3.3 describe saliency correlations with inter mode PU size, quantised residual magnitude and MV data, respectively. The experimental video dataset to derive the algorithms in the consequent subchapter is described in Section 3.4.2.

### 6.3.1 Block Size

As described in Section 6.2.1, block partition size is a highly correlatable feature compared with neural selectivity. Inter mode PU partitioning comprises of 8 possible selections, each of which are shown in Figure 3.15. Consequently, for the entire hierarchical block structure, a total of 24 possible block sizes are available,

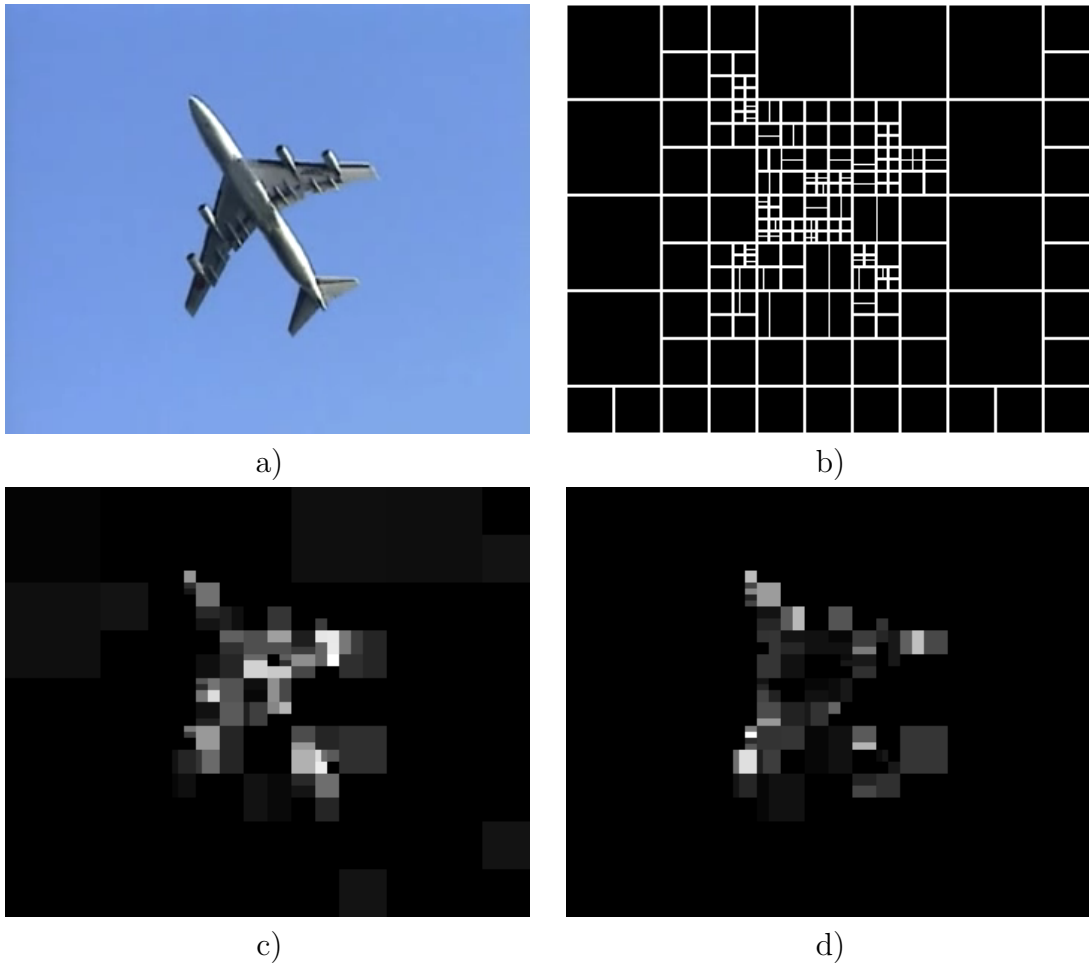


Figure 6.11: Inter-Frame prediction features - a) Original frame b) Block structure c) Residual magnitude d) Motion vector difference magnitude.

excluding the non-viable 4x4 selection [3]. Consistent with the intra mode feature maps in Section 6.2.4, generic classification of each block size allows for a broader saliency approximation. The generic block size classification for inter mode PU block sizes,  $s'$ , are determined by block pixel area,  $a$ , and is shown mathematically in Equation (6.7) and Table 6.1.  $a$  is calculated from the multiplication of the PU block dimensions  $s_1$  and  $s_2$ .

$$s' = \begin{cases} 1 & \text{if } a = 4096, \\ 2 & \text{if } 1024 \geq a < 4096, \\ 3 & \text{if } 256 \geq a < 1024, \\ 4 & \text{if } a < 256. \end{cases} \quad (6.7)$$

Table 6.1: Generic classification of PU block sizes.

$s1 \times s2$	$a$	$s'$
64x64	4096	1
64x48, 48x64	3072	2
32x64, 64x32	2048	
32x32, 64x16, 16x64	1024	
32x24, 24x32	768	3
32x16, 16x32	512	
16x16, 32x8, 8x32	256	
12x16, 16x12	192	4
16x8, 8x16	128	
8x8, 4x16, 16x4	64	
4x8, 8x4	32	

The correlation between  $s'$  and saliency is shown in Figure 6.12. The graph characterises a strong correlation between decreasing block size and increasing visual saliency, as smaller block sizes are more likely to be salient. A similar relationship was previously derived for intra predicted frames in Section 6.2.1.

### 6.3.2 Motion Residual Data

Inter-frame residuals arise from conspicuous frame motion which cannot be accurately predicted by the HEVC encoder. Visual neural stimuli are particularly sensitive to object motion within video sequences [133], which suggests motion residual data will be highly correlatable with visual saliency. Figure 6.13a) shows the relationship between the quantised DC coefficient residual magnitude within each transform block,  $R_m$ , and salient frame regions. The graph shows  $R_m$  dramatically increases in saliency after threshold  $T_{resm}$  is surpassed.  $T_{resm}$ , is based upon histogram analysis and the maximum residual magnitude,  $R_{mmax}$ , comparable to the intra mode residual magnitude partitioning in Figure 6.5. A threshold parameter of  $T_{resm} = R_{mmax} * 0.02$  generates the generic motion residual magnitude,  $R'_m$ , as shown in Figure 6.13b). This classifies the saliency likelihood of each block, dependant upon the residual magnitude. This is mathematically

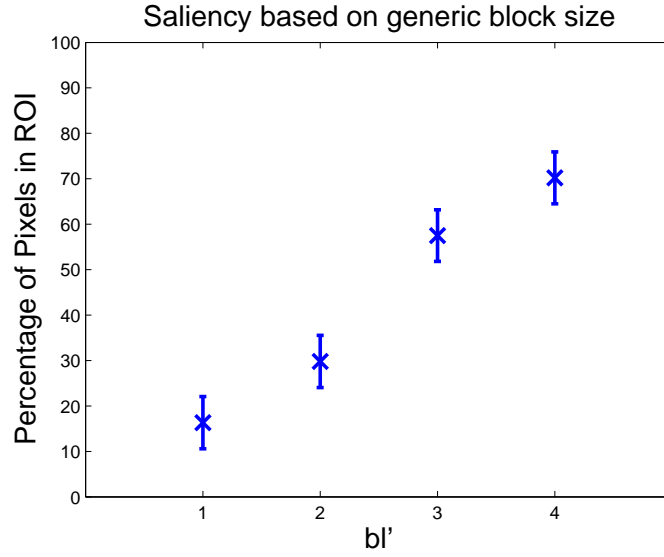


Figure 6.12: Correlation between inter mode PU block size and saliency.

defined in Equation (6.8) by:

$$R'_m = \begin{cases} Low & \text{if } R_m < T_{resm}, \\ High & \text{if } R_m \geq T_{resm}. \end{cases} \quad (6.8)$$

The graph demonstrates partitioned blocks are approximately 3.5 times more likely to be salient by containing high residual magnitude data, compared with the low residual counterpart.

### 6.3.3 Motion Vector Difference Magnitude

The codec MV's describe the optimum 2D translation, from a reference frame, to estimate the current PU block. MV's can help detect and track moving objects [142], which draw human gaze and bear large a resemblance toward the scene optical flow.

The smooth change of motion between neighbouring blocks is exploited by the spatial predictors within the HEVC codec, compensating for any global motion due to translational camera trajectory. Figure 6.14 shows a predicted motion vector from surrounding candidates. A MV deviating from the predicted value

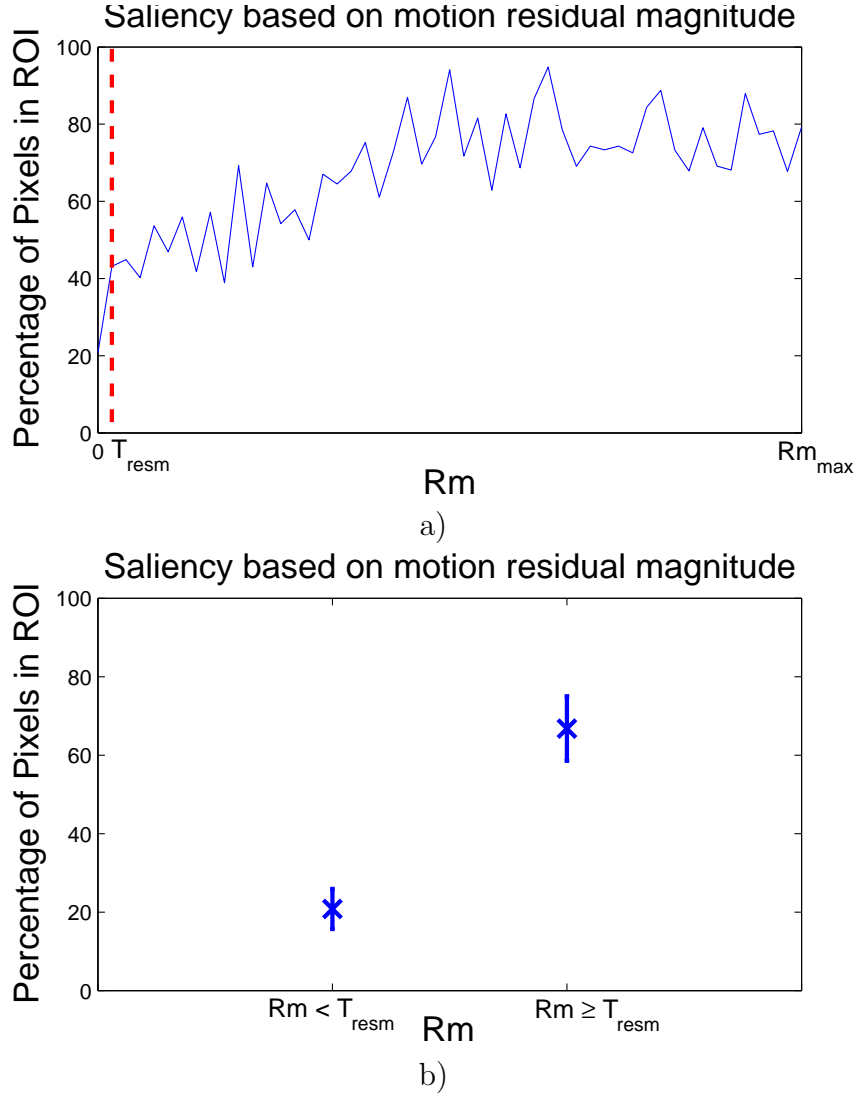


Figure 6.13: Correlation between inter mode block residual magnitude and saliency.

will consequently be encoded by a large MV difference, which is highly probable to originate from object motion within the sequence. The vertical,  $M\vec{V}D_V$ , and horizontal,  $M\vec{V}D_H$ , MV difference components are directly extractable, for every PU partition within the HEVC decoder, so the overall MV difference magnitude,  $MVD$ , can easily be calculated from Equation (6.9) by:

$$MVD = |\sqrt{M\vec{V}D_V^2 + M\vec{V}D_H^2}|. \quad (6.9)$$

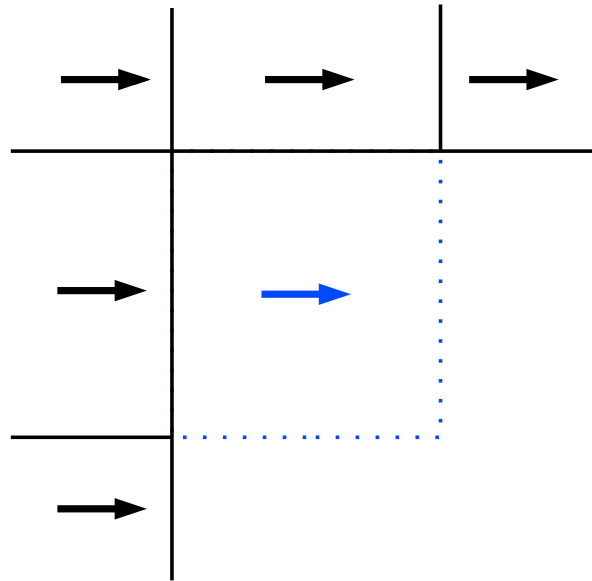
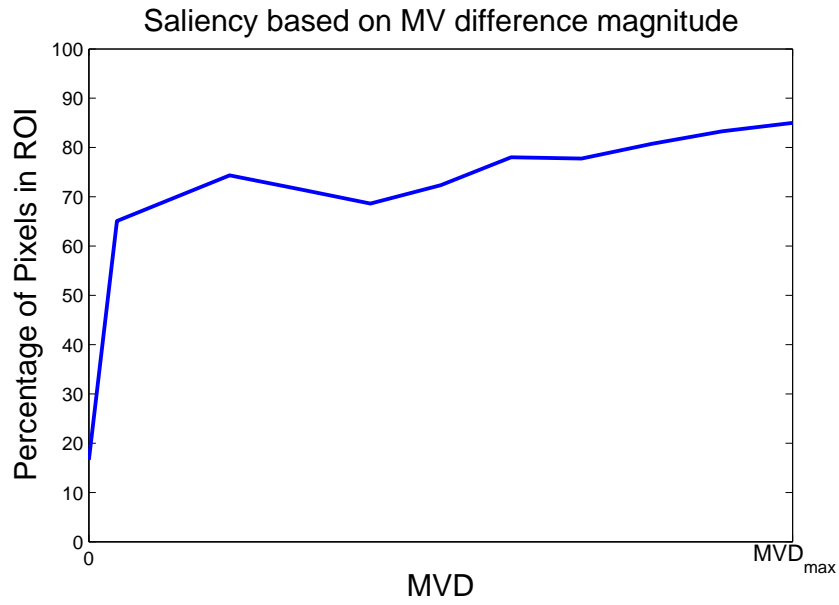
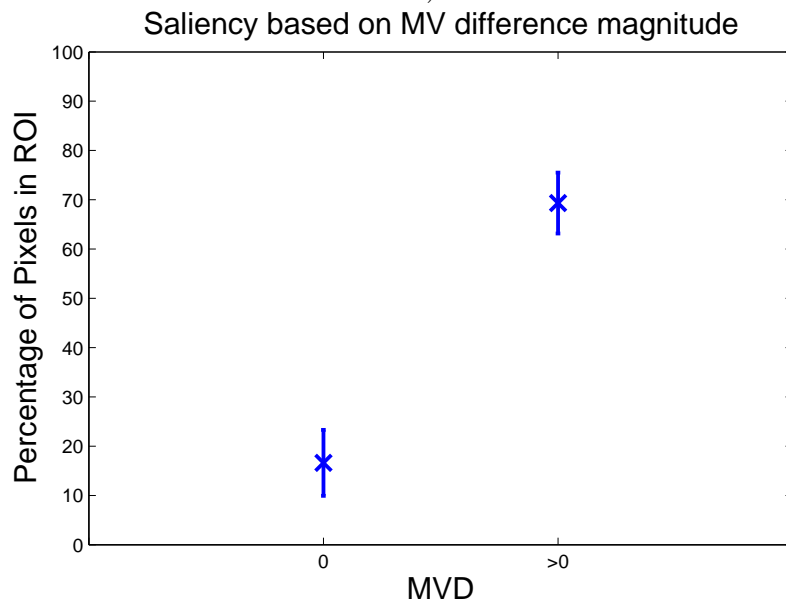


Figure 6.14: Motion vector prediction from spatial candidates.

The MV difference orientation bears negligible resemblance towards salient object prediction, however, the MV difference magnitude is correlatable. Figure 6.15a) shows the relationship between  $MVD$  and saliency, where  $MVD_{max}$  is the maximum possible MV difference limit which coincides with the inter PU candidate search range. From the graph there is a clear saliency deviation between positive MV differences and MV differences of zero magnitude. Figure 6.15b) shows the saliency of partitioned blocks containing a zero and non-zero MV difference magnitude,  $MVD'$ . Naturally, the non-zero magnitude elements have higher saliency correspondence. The MV difference generic partitioning is mathematically de-



a)



b)

Figure 6.15: Correlation between MV difference magnitude and saliency.

defined in Equation (6.10) by:

$$MVD' = \begin{cases} Low & \text{if } MVD = 0, \\ High & \text{if } MVD > 0. \end{cases} \quad (6.10)$$

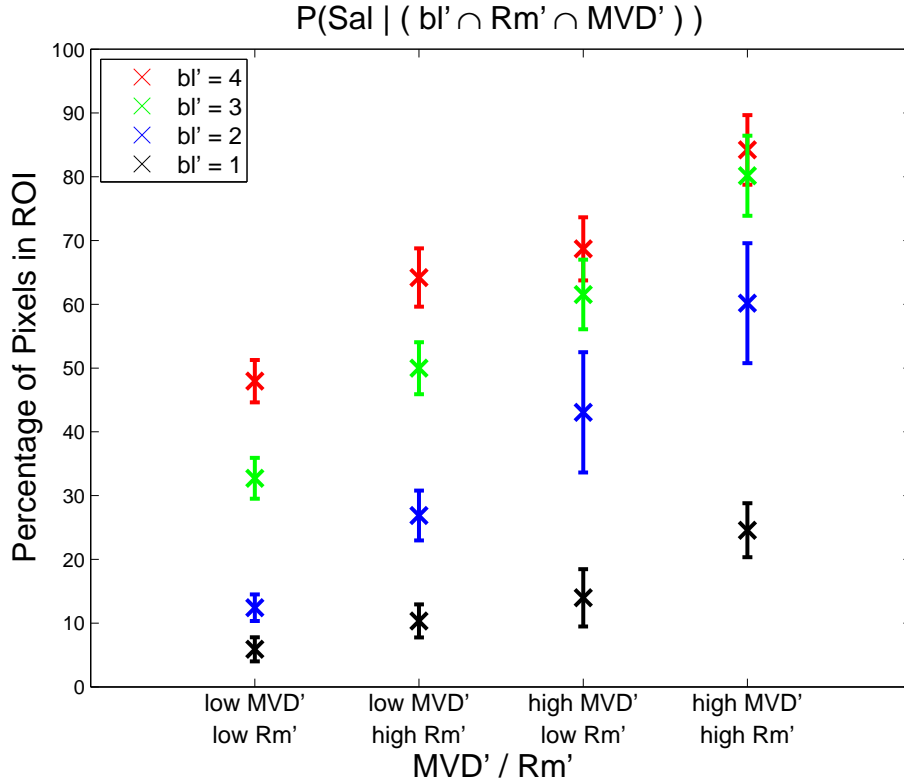


Figure 6.16: Graphs showing  $P(\text{Sal} | (s' \cap R'_m \cap MVD'))$  for all 16 possible outcomes.

### 6.3.4 Inter Saliency Map Generation

An inter predicted PU block will more likely be salient if it contains a high comparative quantised DC residual magnitude, a non-zero MV difference and is finely partitioned into a small PU size. Inter-frame saliency maps are generated by employing conditional probability, complementary to the intra-frame prediction in Section 6.2.4. The combined inter block saliency probability is subsequently expressed by  $P(\text{Sal} | (s' \cap R'_m \cap MVD'))$ . There are 16 possible saliency probabilities for  $P(\text{Sal} | (s' \cap R'_m \cap MVD'))$ , each of which are shown in Figure 6.16. This coincides with the 4 possible values for  $s'$  and 2 thresholded levels for  $R'_m$  and  $MVD'$ .

Consistent with the intra-frame saliency prediction, in Section 6.2.4, a morphological opening and closing operation is applied to the refined output map. An



equation describing the final inter predicted saliency map,  $S_{T_2}$ , is described in Equation (6.11) by:

$$S_{T_2} = \psi.(P(Sal|(s' \cap R'_m \cap MVD'))). \quad (6.11)$$

## 6.4 Combined HEVC Saliency Model

The final saliency map for an HEVC encoded frame,  $S_T$ , is derived from a combination of  $S_{T_1}$  and  $S_{T_2}$  as expressed mathematically in Equation (6.12) by:

$$S_T = \begin{cases} \psi.(P(Sal|(s \cap R' \cap d'))) & \text{if Intra-Frame Prediction} \\ \psi.(P(Sal|(s' \cap R'_m \cap MVD'))) & \text{if Inter-Frame Prediction.} \end{cases} \quad (6.12)$$

Dependant upon whether a frame is encoded by inter or intra-frame prediction, the relevant features can be extracted directly from within the bitstream after entropy decoding. A saliency estimation for each partitioned block can be consequently formulated via a look-up table approach, by extracting data from Figure 6.7 and Figure 6.16. The complete saliency model is shown in Figure 6.17.

## 6.5 Encoder Setup

The encoder setup, in particular the QP, determines the desired output video bitrate and can affect various encoding decisions. If a user requires a low bitrate, the codec selects a high QP value which will highly compress the video sequence. For larger QP, less block partitioning occurs, amounting to generally larger CU block sizes. Also, fewer block prediction errors are transmitted due to an increase in transform residual compression. This in turn can affect the MV configuration and intra mode coefficients within the inter and intra-frame prediction scheme.

The encoder setup affects the block size saliency probabilities slightly as shown in Figure 6.18a), where 3 differing QP values have been used to encode the database

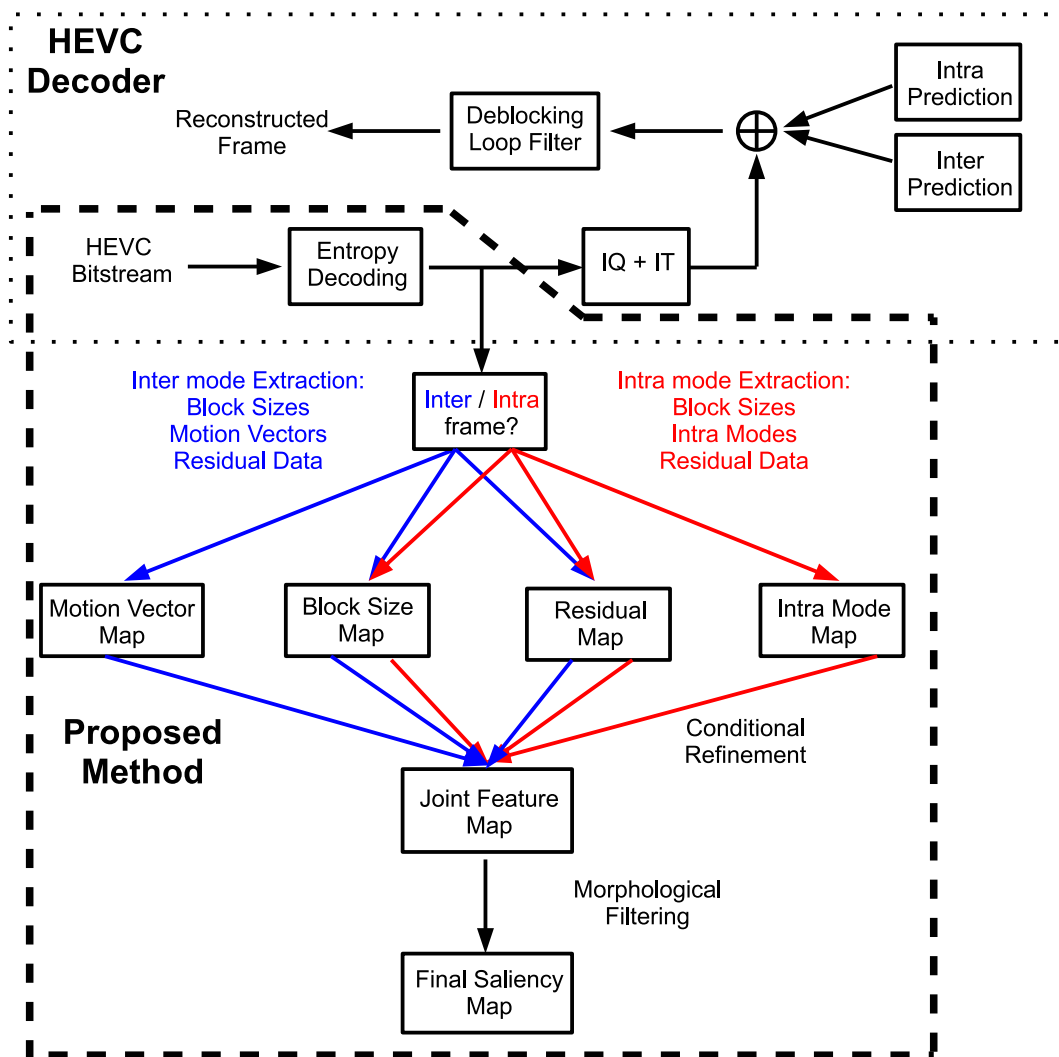


Figure 6.17: Overall HEVC saliency model diagram.

frames. A consistent saliency correlation remains independent of QP. Visual saliency increases as block size decreases, although the previous hard coded probabilities may slightly alter. The residual magnitude, intra mode difference, MV difference and motion residual magnitude saliency probabilities all retain their saliency probability, irrespective of the QP. Figure 6.18b) shows the QP value has negligible impact upon quantised DC residual magnitude saliency probability. The partitioned block size is the main attribute altered by QP. Figure 6.18c), Figure 6.18d) and Figure 6.18e) show the Intra-coded PU frame partitions for the frame in Figure 6.8a) using  $QP = 15$ ,  $QP = 30$  and  $QP = 45$ , respectively.

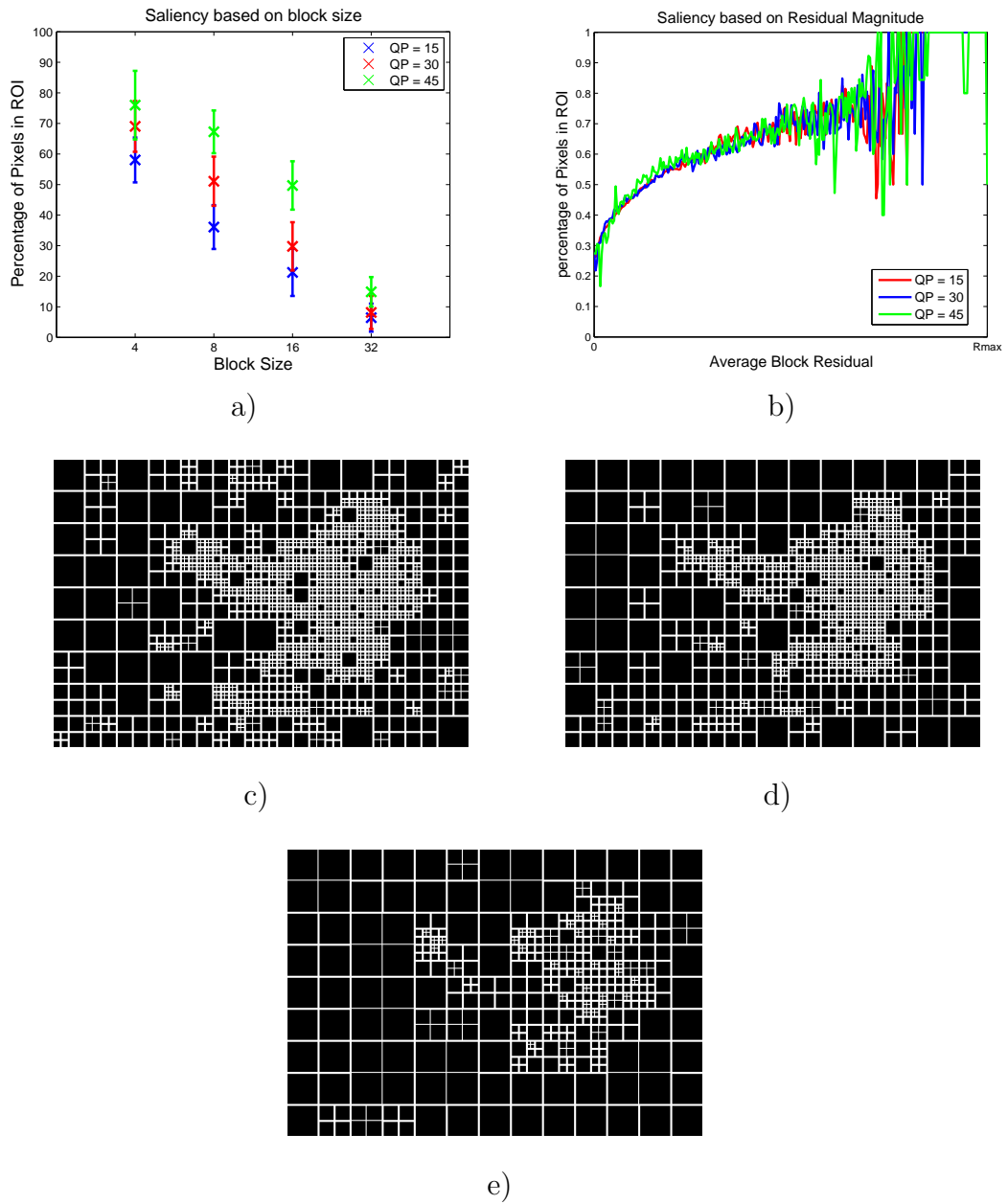


Figure 6.18: Effect of QP on saliency prediction - a) Block size b) Residual magnitude c) Intra-frame predicted block structure for QP = 15 d) Intra-frame predicted block structure for QP = 30 e) Intra-frame predicted block structure for QP = 45.

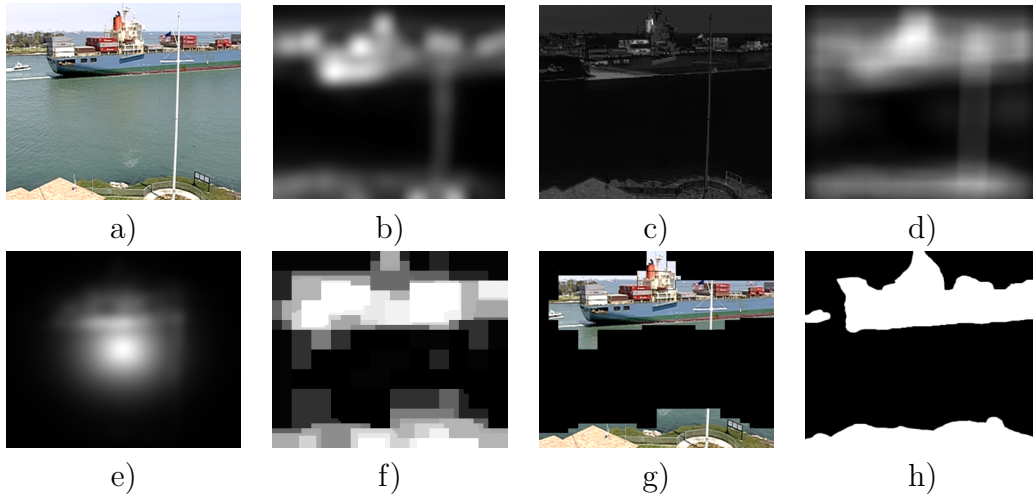


Figure 6.19: a) 'Container' frame b) Itti model [54] c) Ngau model [69] d) Rare model [67] e) Erdem model [68] f) Proposed model g) Thresholded original frame using proposed model h) ground truth frame.

## 6.6 Experimental Results

Experimental results can be divided into 2 sections for the Intra-frame and inter-frame saliency models, in Section 6.6.1 and Section 6.6.2, respectively. The intra-frame model results uses the MSRA database whereas the inter-frame model utilises 15 video sequences, both of which are described in Section 3.4.2. The experimental setup for the consequent subchapters is also described in Section 3.4.2.

### 6.6.1 Intra-Frame Saliency

As the intra saliency model can be implemented on individual frames and recent research has been proposed for Intra-frame HEVC image coding [143], the proposed model can logically be compared against state-of-the-art image saliency domain algorithms. Figure 6.19 compares the proposed method to 4 existing state-of-the-art techniques, with a manually segmented ground truth frame. By subjective visual assessment, in terms of precise saliency estimation, the model can accurately match the performance of the Itti [54], Ngau [69], Rare [67] and Erdem [68] algorithms shown in Figure 6.19b), Figure 6.19c), Figure 6.19d) and Figure 6.19e), respectively. Figure 6.19g) highlights key salient areas within the

Table 6.2: Intra-frame saliency ROC AUC and computational time comparison between proposed and existing models.

	Itti [54]	Ngau [69]	Rare [67]	Erdem [68]	Proposed
ROC AUC for 'Container' frame	0.934	0.775	0.932	0.521	0.960
ROC AUC for 1000 frames	0.875	0.856	0.906	0.878	0.884
Computational time per frame (sec)	0.531	0.329	6.822	16.777	0.159

original frame computed by the proposed methodology.

Subjective assessment alone is not enough to justify the validity of results. ROC curves are shown in Figure 6.20a) and Figure 6.20b) to portray an objective evaluation of each saliency model. Figure 6.20a) displays the ROC curve for the corresponding frame in Figure 6.19, whereas results across the entire MSRA-1000 database are shown in Figure 6.20b). Table 6.2 shows the corresponding ROC AUC and average computational time per frame, for each model. The table values are computed as an average over all 1000 frames in the MSRA database. From the graph in Figure 6.20b) and middle row in Table 6.2 the proposed algorithm exhibits the second highest performance, accurately detecting salient regions, behind the Rare model by only 2.4% difference in ROC AUC. The major drawback of the Rare and Erdem models are the exhaustive iterative procedure requires approximately 43 and 105 times the processing time, respectively, compared with the proposed model. The high complexity constraints deem the models highly unsuitable for any video saliency applications. The Ngau model, despite being capable of rapid scene analysis, performed the worst of all models when estimating salient regions, having a 3.3% lower ROC AUC than the proposed method. The simple algorithm searches for frame regions containing extreme coefficients and is limited to saliency estimation upon very basic scenes. The bottom row in Table 6.2 shows model computational time per frame and the proposed model time requirements are 30%, 48%, 2.2% and 0.9% of the Itti, Ngau, Rare and Erdem models, respectively.

The main advantage of generating the saliency maps from within the compressed domain, is data required for saliency estimation is extracted directly from within

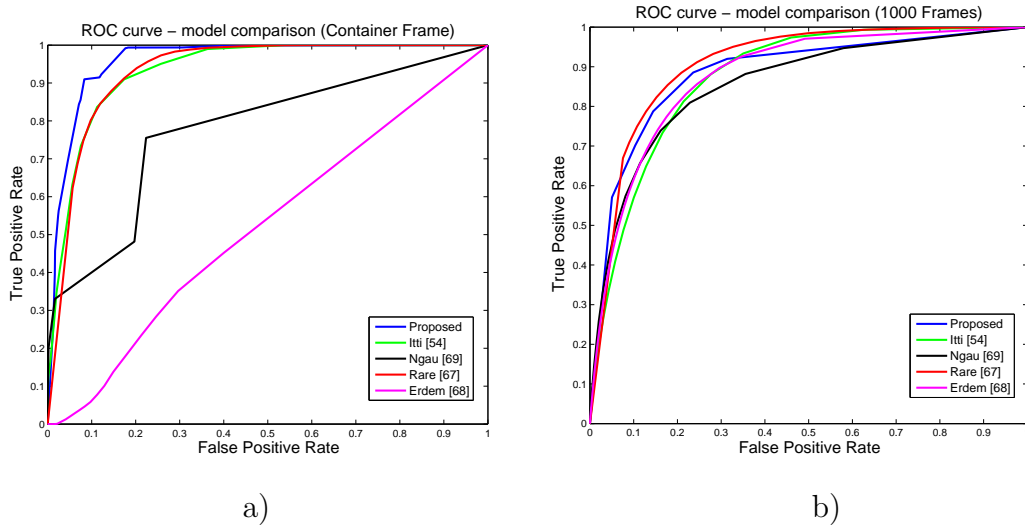


Figure 6.20: ROC curves comparison to existing state of the art methodologies for inter-frame saliency model - a) frame from 'Container' sequence b) entire MSRA-1000 database.

the HEVC codec and re-mapped in a basic learning algorithm. As a result, only minimal added computational costs occur in excess of entropy decoding the HEVC bitstream, so the model can attain an accurate saliency estimation while maintaining a very low overall computational cost. This joint characteristic ensures great suitability for video saliency estimation. The other methods in comparison are pixel domain based, so full HEVC bitstream decoding must be performed before the saliency algorithm is applied.

Test frames demonstrating the proposed model performance are shown in Figure 6.21, with the corresponding ground truth segmentation. A large set of additional results, also from the MSRA-1000 database, are provided in the appendix.

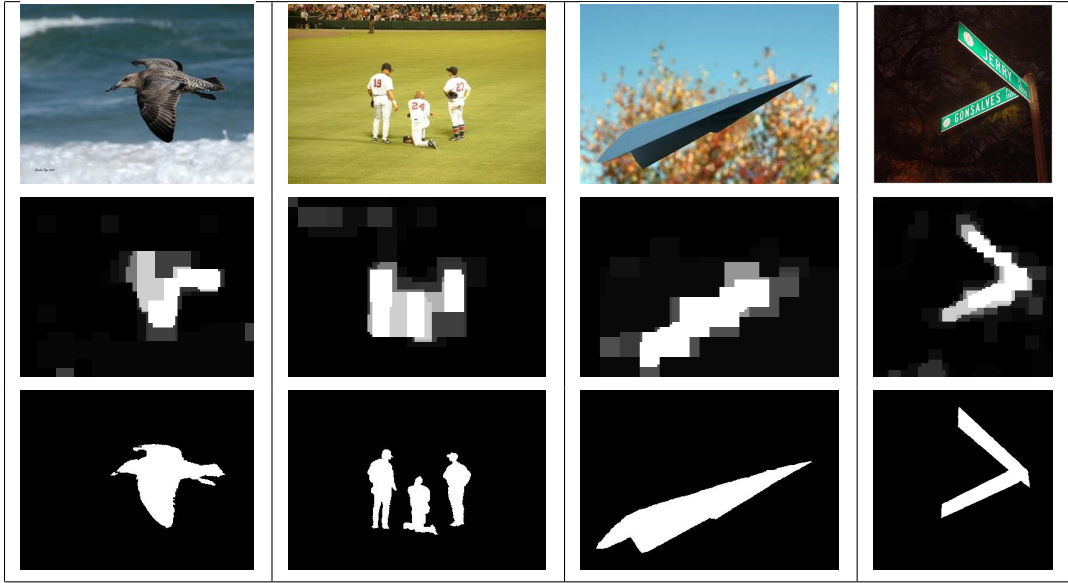


Figure 6.21: Intra-frame saliency results - (Row 1: Original image, Row 2: Saliency regions using the proposed model and Row 3: Ground truth).

### 6.6.2 Inter-Frame Saliency

Figure 6.22 and Figure 6.23 shows saliency maps for the proposed inter-frame method compared with the video saliency algorithms: Itti motion model [74] and Dynamic length model [134]. An ROC curve for the entire saliency video database is shown in Figure 6.24, with corresponding AUC value on the top row in Table 6.3. The results show the proposed model has a 1.1% and 5.7% increase in ROC AUC compared with the Itti and Dynamic models, respectively.

A major benefit provided from compressed domain saliency estimation is the computational complexity, in terms of processing time per frame. The bottom row in Table 6.3 provides computational time per frame for the proposed compressed domain methodology compared with existing approaches. The proposed method has a computational time 20% and 22% of the Itti and Dynamic models, respectively. As with the intra-frame saliency model provided, the proposed methods extremely high computational efficiency is due to the frame analysis, for ROI detection, being provided as a by-product of the HEVC encoder. From a compressed video bitstream, the comparison methods must first fully decode the bitstream as they are based within the spatial domain.

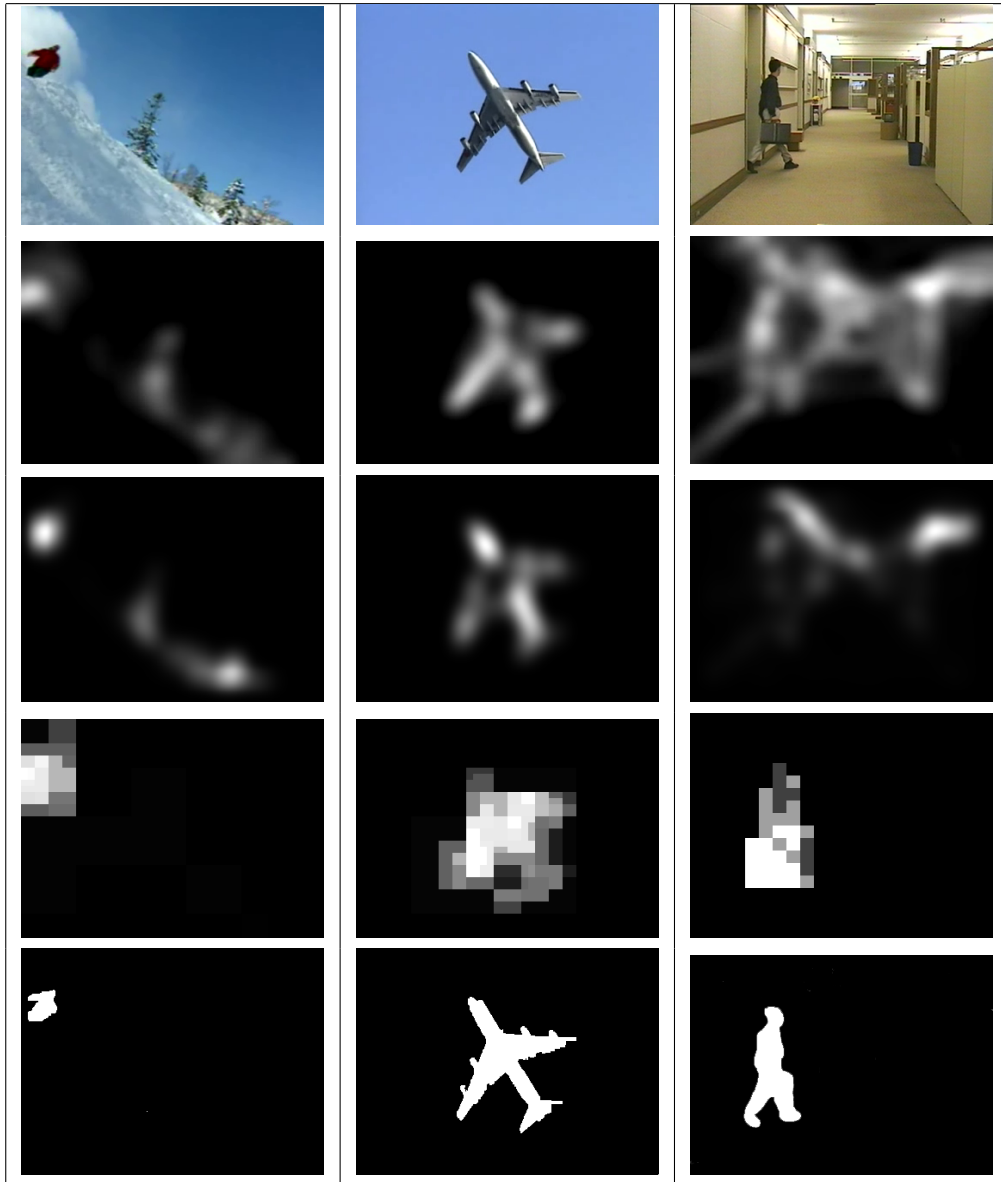


Figure 6.22: Inter-frame Saliency results 1 - (Row 1: Original frame, Row 2: Itti [74], Row 3: Dynamic [134], Row 4: Proposed and Row 5: Ground truth).

The limitations of the proposed model are highlighted in Figure 6.25. A bright pink flower is labelled as the salient object in Figure 6.25a) and Figure 6.25b). However, our saliency estimation in Figure 6.25c) does not compute human visual gaze will fixate entirely upon this region. The luma component only is used for saliency estimation, so neural stimuli sensitive towards bright colours are not necessarily captured within the model. This can be overcome by incorporating the chroma channel within the proposed framework.



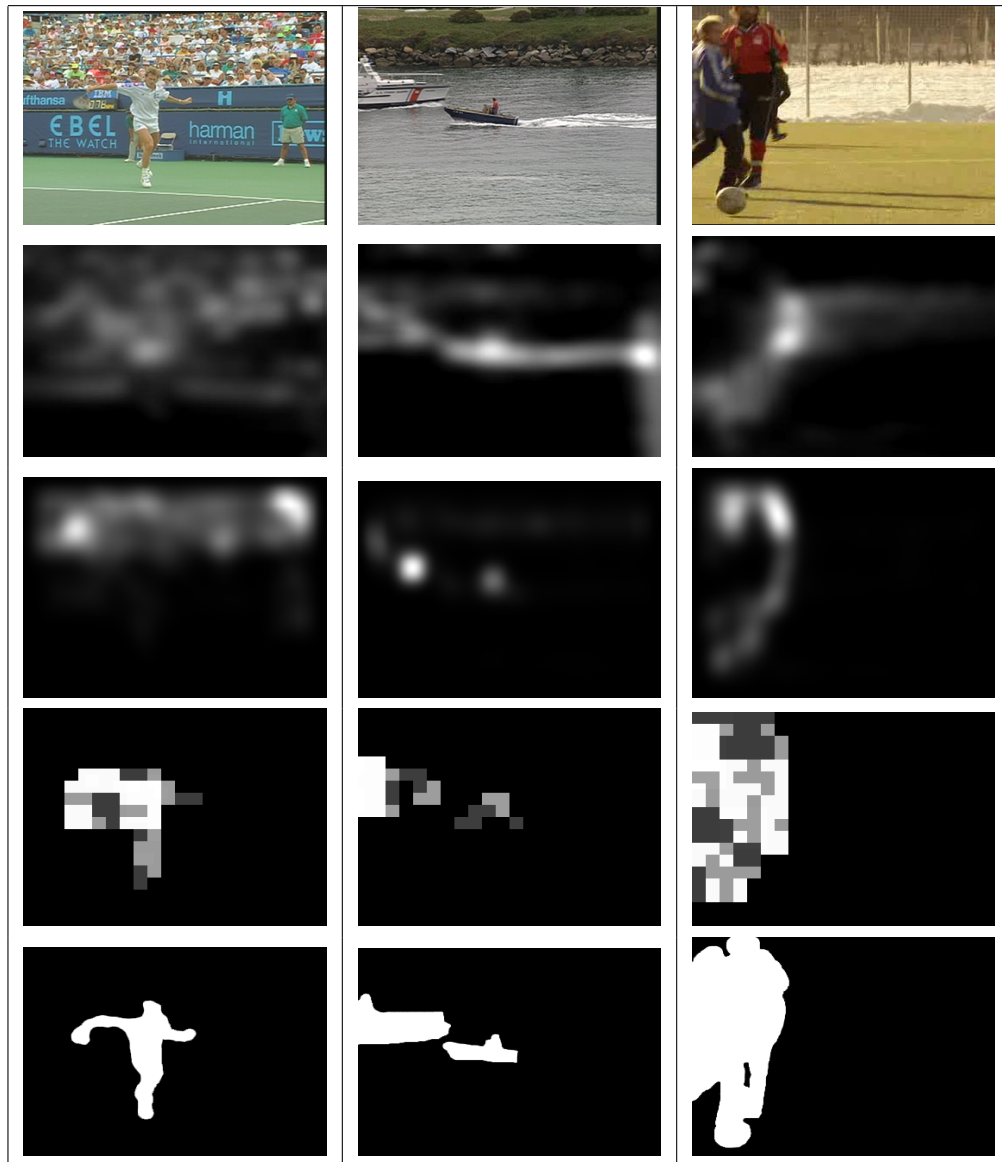


Figure 6.23: Inter-frame Saliency results 2 - (Row 1: Original frame, Row 2: Itti [74], Row 3: Dynamic [134], Row 4: Proposed and Row 5: Ground truth).

As with the proposed model in Section 4.3, video saliency maps are best viewed as an entire sequence, rather than a static image. Video sequences are available for viewing from the website:

<http://svc.group.shef.ac.uk/hevc-va.html>.

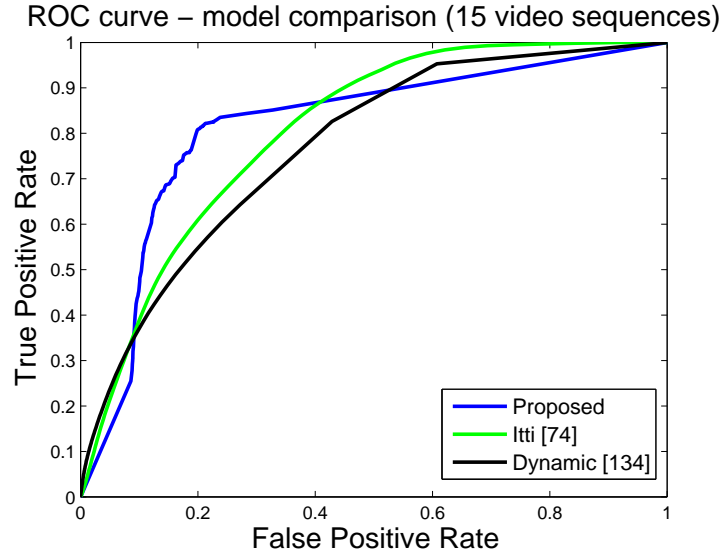


Figure 6.24: ROC curve comparison to existing state of the art methodologies for inter-frame saliency model.

Table 6.3: Inter-frame saliency ROC AUC and computational time comparison between proposed and existing models.

	Itti [74]	Dynamic [134]	Proposed
ROC AUC for 1000 frames	0.804	0.769	0.813
Computational time per frame (sec)	0.494	0.444	0.099

## 6.7 Conclusions

A novel HEVC domain visual saliency algorithm was presented within this chapter by uniquely exploiting features within the coding standard. Using partially decoded HEVC sequences from encoded intra and inter predicted frames, the block partition size, quantised residual magnitude, MV difference data and intra mode differences were combined using conditional probability to constitute a definitive saliency model. Unlike existing methodologies, the HEVC bitstream does not need to be fully decoded to attain an accurate saliency estimation. The proposed model requires only 48% or less computational time, compared against state-of-the-art methodology, while maintaining accurate saliency estimation. The ROC AUC only slightly declines by 2.2% or is higher than existing

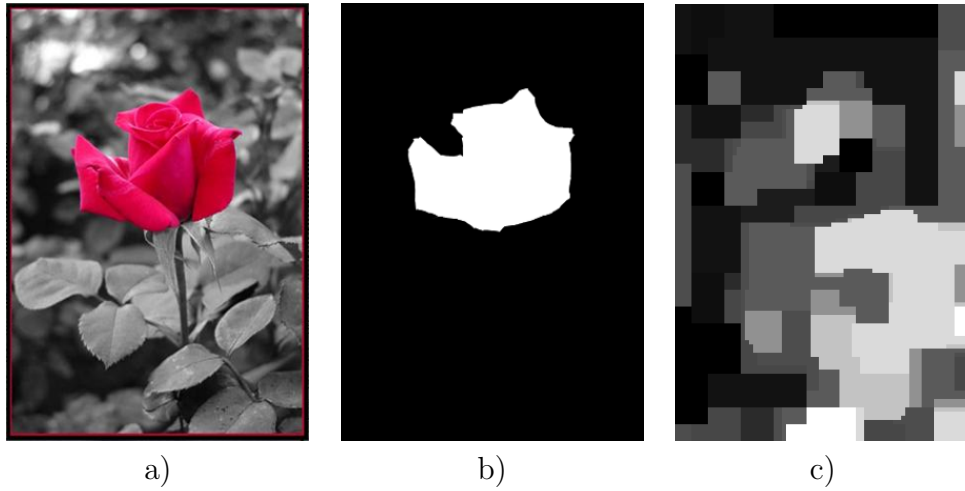


Figure 6.25: Saliency model limitations - a) original frame b) ground truth c) proposed saliency model

methods. The final section of the thesis concentrates on providing an HEVC domain VA-based watermarking framework, dependant upon these derived saliency algorithms.



# Chapter 7

## HEVC domain Watermarking

This chapter of the thesis focuses upon watermarking within the HEVC domain. By implementing the novel compressed domain saliency algorithm in Chapter 6, unique VA-based HEVC watermarking techniques are derived within the video encoder and frame domain. As in Chapter 5, the proposed algorithms embed greater watermark information into coefficients exhibiting low visual salience to provide a robust system, without compromising the visual quality of the media.

Limited encoder-based and compressed domain watermarking strategies have been researched for the H.264/AVC standard [36–40], but new proposals are required to deal with the upcoming HEVC codec. The existing approaches embed information within the transform residual data, to provide robust compressed domain approaches, but none of the research considers VA-based features.

This chapter provides an HEVC domain VA-based watermarking scheme for real time watermarking application. Data is modified within each I-frame and due to the nature of the watermarking scheme, only a minimal increase in the overall bitrate occurs. The algorithm is executed jointly within the video coding process and only partially decoded bitstreams are required to extract an embedded watermark.

The rest of this chapter is arranged as follows: possible HEVC domain watermarking approaches are discussed in Section 7.1. Section 7.2 explains the coef-

efficient selection process within the transform domain and Section 7.3 describes the proposed watermarking scheme, explaining both watermark embedding and detection. The experimental results are described in Section 7.4 and finally, the concluding remarks are shown in Section 7.5.

## 7.1 HEVC Watermarking Approaches

Most video sequences exist within the compressed format to reduce the required file size. There are 3 possible approaches to watermark an HEVC encoded sequence:

- Within the frame domain after the video sequence has been fully decoded;
- A transcoding approach by embedding data within the compressed domain during the bitstream decoding; and
- A tandem coding approach by re-encoding the video sequence and embedding data during the encoding.

Each of the 3 options are described in more detail in Section 7.1.1, Section 7.1.2 and Section 7.1.3.

### 7.1.1 Frame Domain Watermarking

A frame domain VA-based watermarking scheme extends the study provided in Section 5.2, where the saliency map is formulated within the HEVC compressed domain, described in Chapter 6, as opposed to the wavelet domain. Figure 7.1 demonstrates watermarking within the frame domain, where the compressed video is first fully decoded before the watermark is embedded. An obvious advantage of this method is the simplistic approach combining previous techniques implemented within Section 5.2 and Chapter 6.

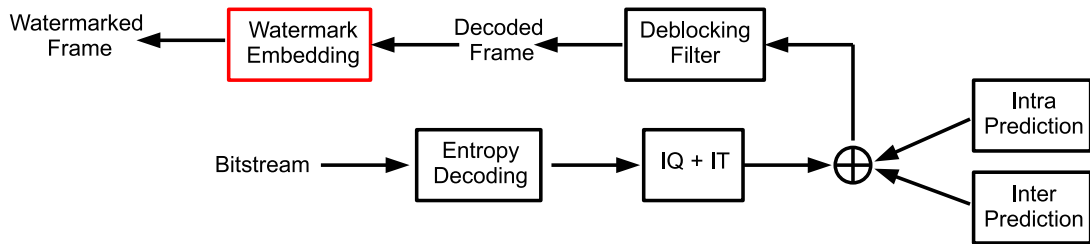


Figure 7.1: Frame Domain Watermark Embedding for an HEVC Encoded Sequence.

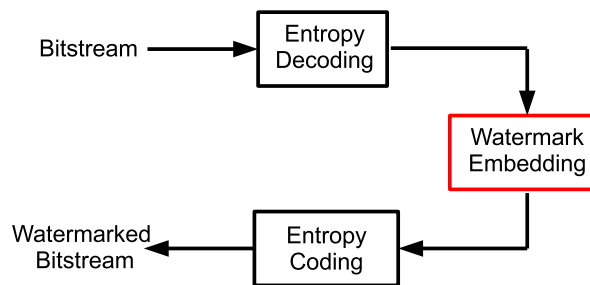


Figure 7.2: Compressed Domain Watermark Embedding for an HEVC Encoded Sequence.

## 7.1.2 Compressed Domain Watermarking

Figure 7.2 shows the process of compressed domain watermarking, employing a transcoding approach. The HEVC encoded sequence is partially decoded and a watermark is embedded within the quantised transform residual data. The modified data is consequently encoded back, forming a watermarked bitstream. However, this solution provides a huge drift problem as decoded blocks are used to predict future blocks within the same frame and any modification will become propagated throughout the remaining predictions. Figure 7.3 demonstrates the watermark drift propagation problem which arises within compressed domain watermarking algorithms. In the figure, 1 coefficient within every 8x8 quantised transform block has been modified. The entire Doctor of Philosophy dissertation of Noorkami [144] is dedicated towards determining a compressed domain drift compensation scheme for H.264/AVC transform coefficient watermarking, therefore a compressed domain watermarking scheme is not provided.



Figure 7.3: Watermark drift problem with compressed domain embedding - a) Original frame b) Watermarked frame.

### 7.1.3 Joint Encoder Watermark

The final option employs a tandem coding approach to watermark an HEVC encoded sequence. Firstly, the bitstream is fully decoded into an output video sequence. The decoded media is consequently re-encoded and watermark data is embedded during this process as shown in Figure 7.4. All of the watermarked quantised residual coefficients are incorporated within any consequent future block predictions, therefore this methodology eludes any watermark drift associated with the approach described in Section 7.1.2. To determine a VA-based embedding coefficient selection, the watermarking scheme is a hybrid with the saliency model provided in Chapter 6. Further details describing the joint encoder watermarking scheme are provided in Section 7.2 and Section 7.3.

## 7.2 Transform Coefficient Watermarking Criteria

To provide a robust watermarking scheme a suitable candidate to embed watermark data must be chosen. As it is very unlikely that the host media will be available at the decoder, the watermarking scheme must be blind. The Intra



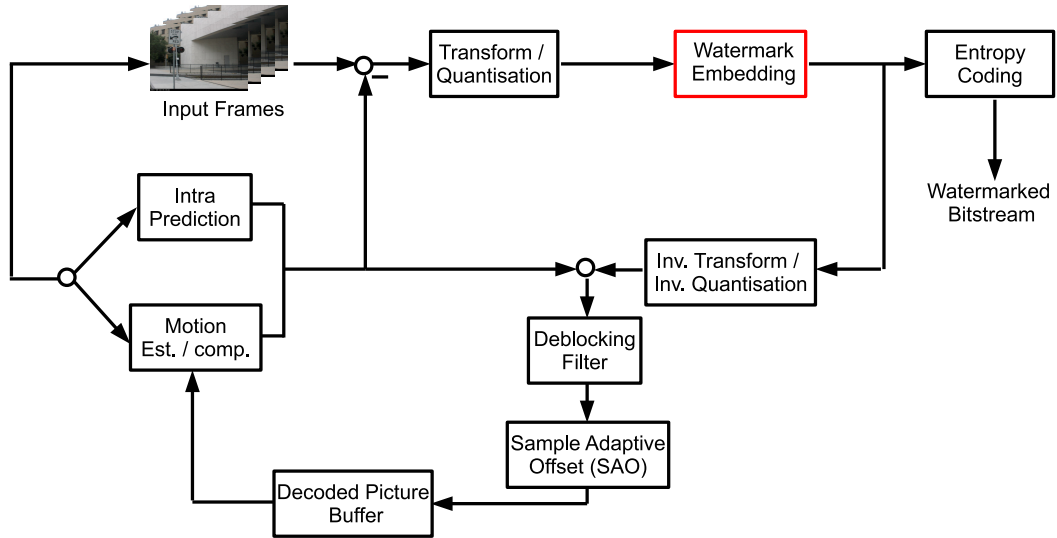


Figure 7.4: Encoder Domain Watermark Embedding for an HEVC Encoded Sequence.

frames are chosen as a suitable candidate for watermark embedding because:

- An adversary is more likely to remove other frames as the I-frame existence is crucial to providing a high quality video signal;
- Each of the possible HEVC coding structures, described in Section 3.3.1, contain I-frames; and
- There is more room for embedding data within I-frames as inter predicted data is very highly compressed.

As quantisation is a lossy operation, it is desirable to embed and detect watermark information prior to the procedure. This ensures any modified data does not have to survive the codec compression. Entropy decoding and encoding is an extremely computationally efficient procedure, consequently, a real time watermarking scheme can be implemented by embedding watermark data close to the entropy coding [37] within the quantised transform coefficients. The DC coefficients, located at the top left position within each transform matrix, represent the average block residual level and are highly sensitive towards any modification. Embedding data within a DC coefficient can cause high fluctuation within the frame brightness, resulting in large visible artifacts [144]. Therefore, transform

AC coefficients are used throughout the watermark embedding procedure as these represent the finer details within the transform. However, large homogeneous regions, free from any texture contain many AC and DC coefficient values close to zero. Any modification within these blocks will be highly perceptible and may influence visual gaze across the media, consequently, these blocks are avoided for embedding. The luma channel is used for embedding as more compression occurs within chroma components. An adversary is less likely to intentionally or unintentionally remove the watermark by changing the video format, i.e YUV 4:4:4 to YUV 4:2:0. Modifying data within the quantised transform residuals can be an extremely risky procedure, especially at high values for QP, as this will have a large effect on the decoded frame. For this reason only 1 coefficient is modified within each transform block, to minimise embedding distortion.

## 7.3 Proposed Watermarking Scheme

The proposed watermarking scheme can be split into 2 sections for the watermark embedding and detection as described in Section 7.3.1 and Section 7.3.2, respectively.

### 7.3.1 Watermark Embedding

There are 4 possible TU 4 sizes the HEVC codec may choose: 4x4, 8x8, 16x16 and 32x32. An integer DCT transform is used for the 8x8, 16x16, 32x32 blocks whereas a Discrete Sine Transform (DST) is used for the 4x4 block size. Firstly, a suitable watermarking coefficient,  $C_T$ , is chosen from the criteria described within Section 7.2. Using an 8x8 transform block,  $B$ , as an example, the horizontal and vertical indexing is given by,  $i$  and  $j$ , respectively. Firstly, only blocks with non-zero DC values are chosen, i.e  $C_T \neq B(0,0)$ . As only 1 coefficient is to be embedded, the coefficients where  $i = j$  provide an unbiased selection towards a particular direction. Figure 7.5 shows the suitable candidates for watermark embedding, highlighted in gray. Coefficients within the lower right half of the transform are heavily quantised and are usually zero, therefore provide a poor

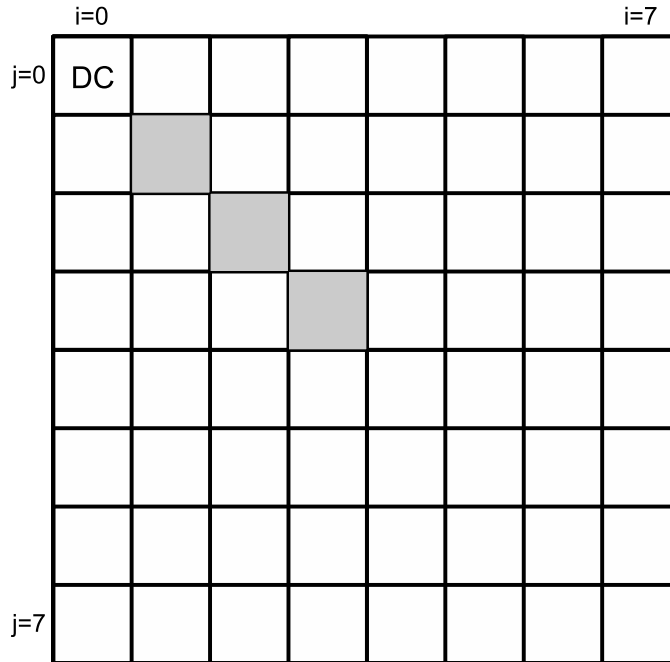


Figure 7.5: Transform domain embedding candidates for an 8x8 block.

choice. The QP value determines the possible  $C_T$  option from the remaining selections. For High QP,  $C_T = B(3, 3)$  is selected, as modifying a highly quantised coefficient will have a large visual impact on the output frame and the higher frequency components will provide less distortion.  $C_T = B(1, 1)$  is selected for low QP. HEVC defines 52 values for QP.  $C_T$  can be determined based upon the QP value, for the watermark candidate selection by Equation (7.1):

$$C_T = \begin{cases} B_{(1,1)} & \text{if } QP = 0 - 16 \\ B_{(2,2)} & \text{if } QP = 17 - 34 \\ B_{(3,3)} & \text{if } QP = 35 - 51. \end{cases} \quad (7.1)$$

The 4x4 block only has 1 possible candidate, whereas similar equations can be derived for the 16x16 and 32x32 transform blocks. As the distributions of coefficient magnitude, within the DCT, are weighted towards the top left coefficient, potential improvements could be made by adapting the selection of  $C_T$  according to the distribution.

The adapted watermarking scheme is based upon the blind watermarking algorithm provided by Noorkami [37] as this highly cited work ensures minimal modification to the visually sensitive quantised transform coefficients. The equa-

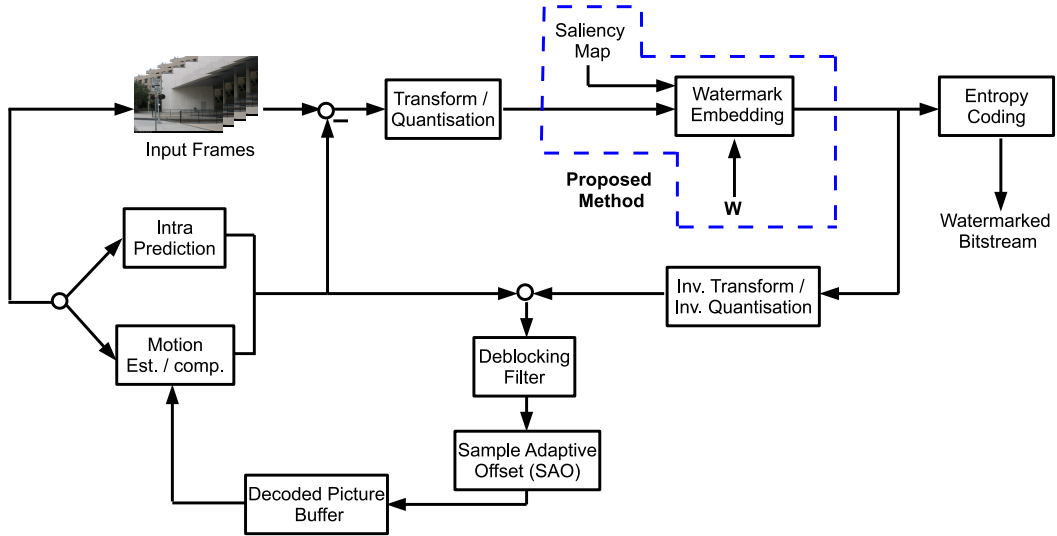


Figure 7.6: Compressed domain VA-based watermark embedding framework within the HEVC encoder.

tion to modify the selected transform coefficient is dependant upon whether the binary watermark,  $W$ , to be embedded is a 1 or 0. Equation (7.2) and Equation (7.3) show both embedding scenarios for  $W = 0$  and  $W = 1$ , respectively, by:

if  $W = 0$ ,

$$C_T = \begin{cases} C_T & \text{if } C_T \% 2 = 0 \\ C_T - 1 & \text{if } C_T \% 2 = 1, \end{cases} \quad (7.2)$$

and if  $W = 1$ ,

$$C_T = \begin{cases} C_T + 1 & \text{if } C_T \% 2 = 0 \\ C_T & \text{if } C_T \% 2 = 1, \end{cases} \quad (7.3)$$

where  $\%$  denotes the modulo operation.

The HEVC encoded sequence is decoded and to employ a VA-based watermarking scheme, a saliency estimation from the algorithm in Section 6.2 is generated as a by product of the decompression. A threshold is applied to the saliency map, as in Section 5.1.1, to determine the visually uninteresting frame regions. The inattentive frame blocks are consequently watermarked using Equation (7.2) and Equation (7.3) during video re-encoding. An overall framework for HEVC joint

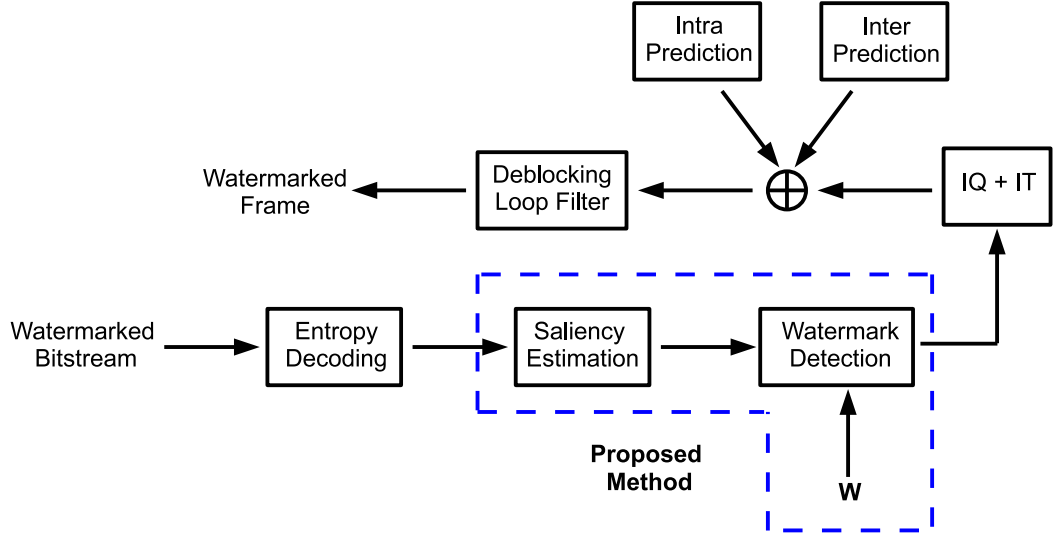


Figure 7.7: Compressed domain VA-based watermark detection framework within the HEVC decoder.

encoder VA-based watermark embedding is shown in Figure 7.6.

### 7.3.2 Watermark Detection

The embedded watermark in Section 7.3.1 can be detected by applying the saliency estimation algorithm from Section 6.2 to determine the visually uninteresting watermarked blocks and applying Equation (7.4):

$$W_0 = \begin{cases} 0 & \text{if } C_T \% 2 = 0 \\ 1 & \text{if } C_T \% 2 = 1, \end{cases} \quad (7.4)$$

where  $W_0$  is the extracted watermark bit. The Hamming distance between  $W$  and  $W_0$  can consequently be calculated to authenticate the media as described by Equation (3.3).

## 7.4 Experimental Results

All of the data test sets used are described in Section 3.4.2. The experimental results comprise of 2 sections for frame domain watermarking, in Section 7.4.1,

Table 7.1: PSNR, SSIM and VQM of 4 video sequences for HEVC-based frame domain watermarking.

	<b>Low strength</b>	<b>Proposed</b>	<b>High strength</b>
PSNR	40.23 ± 1.03	37.76 ± 1.21	34.85 ± 0.90
SSIM	0.99 ± 0.00	0.98 ± 0.00	0.96 ± 0.01
VQM	0.08	0.12	0.22

and joint encoder watermarking, in Section 7.4.2, as described in Section 7.1.1 and Section 7.1.3, respectively.

### 7.4.1 Frame Domain Watermarking Experimental Results

For the frame domain algorithm, an encoded bitstream is firstly fully decoded and a compressed domain saliency estimation is made from Section 6.4. Wavelet domain VA-based watermarking is performed combining the compressed domain saliency maps and the blind watermarking scheme described in Section 5.2. Three frame domain watermarking scenarios are considered, as in Chapter 5:

- 1) a uniform  $\alpha_{min}$  for the entire scene (low strength watermark);
- 2) the proposed watermarking scheme which implements an adaptive VA-based  $\alpha$ , adopting the HEVC compressed domain saliency maps; and
- 3) a constant  $\alpha_{max}$  throughout the embedding procedure (high strength watermark).

Identical test conditions are performed to the experimental setup in Section 5.2.1. Section 7.4.1.1 and Section 7.4.1.2 show the visual quality and watermark robustness results for frame domain watermarking, respectively.

#### 7.4.1.1 Imperceptibility

The PSNR, SSIM and VQM average for 4 watermarked sequences are shown in the top, middle and bottom rows of Table 7.1, respectively. A Similar visual

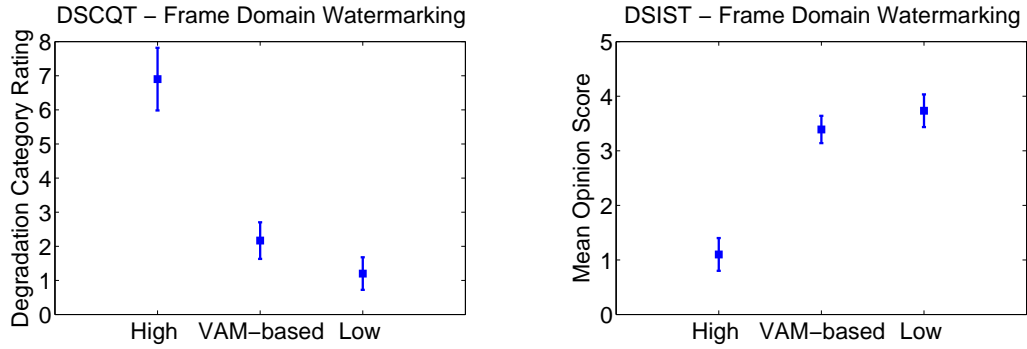


Figure 7.8: Subjective Test for Frame Domain Watermark Embedding.

quality is maintained between the low strength and proposed algorithms, differing by approximately 2dB in PSNR. Results from two subjective tests, DSCQT and DSIST, are shown in Figure 7.8. The low strength and VA-based scenarios portray a similar subjective visual quality, whereas the high strength media quality is deteriorated due to the perceived embedding distortion.

#### 7.4.1.2 Robustness

Robustness against H.264/AVC compression is shown in Figure 7.9 for all 3 watermarking scenarios. From the graph, a 20% improvement in Hamming Distance is possible when comparing the low strength and proposed VA-based algorithm. Despite having an increased robustness, the visual quality remain similar as described in Section 7.4.1.1.

Results and reasoning behind the HEVC-based frame domain watermarking algorithm are highly comparable to Section 5.2.1, as any discrepancies will only lie resultant of dissimilar saliency estimations.

### 7.4.2 Joint Encoder Watermarking Experimental Results

Due to the nature of the blind watermarking scheme, only 2 scenarios are considered:

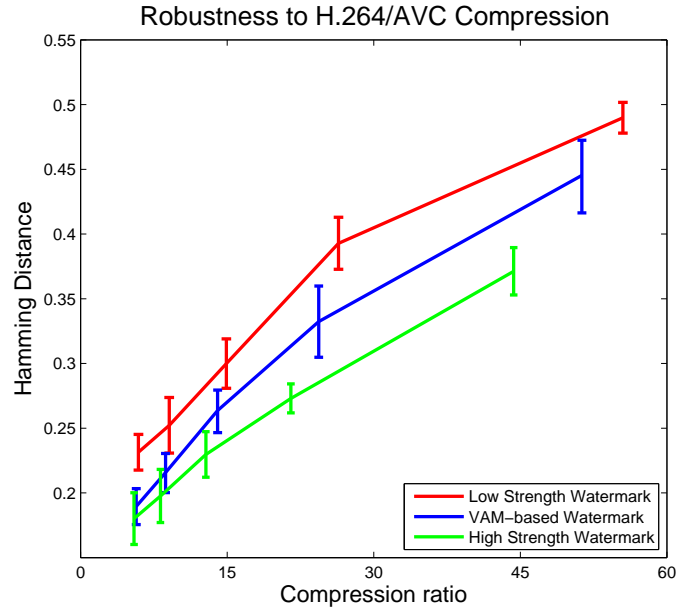


Figure 7.9: Robustness to H.264/AVC compression - average of 4 video sequences.

- 1) the proposed VA-based watermarking scheme only embedding transform blocks portraying low saliency; and
- 2) high strength watermark embedding within every transform block throughout the frame regardless of VA.

As 1 bit is embedded within every transform block for the high strength proposal and VA-based watermarking is performed by only modifying transform blocks exhibiting low visual salience, low strength watermarking cannot be achieved, therefore is omitted. The results subsection is split into 2 sections for the robustness and imperceptibility evaluation of the proposed method.

#### 7.4.2.1 Imperceptibility

The imperceptibility of the proposed watermarking scheme is compared against the high strength algorithm. The top, middle and bottom rows in Table 7.2 show the PSNR, SSIM and VQM, respectively, for the Stefan, Container, Hall and Soccer sequence. Naturally, the proposed method shows a high increase in visual



Table 7.2: PSNR, SSIM and VQM of 'Stefan', 'Container', 'Hall' and 'Soccer' sequences.

	Stefan		Container	
	<b>Proposed</b>	<b>High strength</b>	<b>Proposed</b>	<b>High strength</b>
PSNR	$38.55 \pm 0.56$	$34.62 \pm 0.29$	$38.12 \pm 0.56$	$33.51 \pm 0.29$
SSIM	$0.99 \pm 0.00$	$0.97 \pm 0.00$	$0.99 \pm 0.00$	$0.97 \pm 0.00$
VQM	0.10	0.18	0.09	0.20
	Hall		Soccer	
	<b>Proposed</b>	<b>High strength</b>	<b>Proposed</b>	<b>High strength</b>
PSNR	$42.11 \pm 0.45$	$36.21 \pm 0.24$	$37.44 \pm 0.61$	$32.93 \pm 0.31$
SSIM	$0.99 \pm 0.00$	$0.98 \pm 0.00$	$0.99 \pm 0.00$	$0.97 \pm 0.00$
VQM	0.05	0.17	0.09	0.21

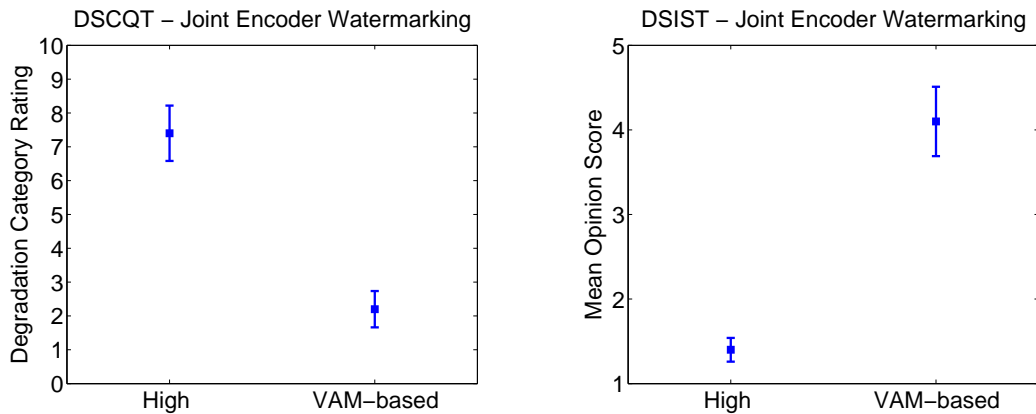


Figure 7.10: Subjective Testing for a Joint Encoder Watermark Embedding System.

quality in all cases as watermark data is only embedded within locations a viewer is less likely to view. This statement is further justified by subjective testing, shown in Figure 7.10, where the average score over all 4 sequences is taken. It is clear from the figure there is a dramatic visual quality difference between the high strength and proposed methodologies, with the VA-based algorithm showing an improved DCR and mean opinion score by 5 and 3 units, respectively, despite only having a PSNR approximately 4dB apart.

Table 7.3: Robustness against HEVC re-encoding using various QP for the 'Stefan', 'Container', 'Hall' and 'Soccer' sequences.

	Stefan		Container	
	<b>Proposed scheme</b>	<b>High strength scheme</b>	<b>Proposed scheme</b>	<b>High strength scheme</b>
QP = 25	0.21 ± 0.02	0.18 ± 0.02	0.23 ± 0.03	0.22 ± 0.02
QP = 27	0.39 ± 0.03	0.35 ± 0.02	0.41 ± 0.02	0.41 ± 0.01
QP = 30	0.45 ± 0.01	0.42 ± 0.01	0.46 ± 0.01	0.45 ± 0.01
	Hall		Soccer	
	<b>Proposed scheme</b>	<b>High strength scheme</b>	<b>Proposed scheme</b>	<b>High strength scheme</b>
QP = 25	0.19 ± 0.03	0.17 ± 0.02	0.22 ± 0.02	0.19 ± 0.02
QP = 27	0.41 ± 0.01	0.40 ± 0.01	0.43 ± 0.01	0.40 ± 0.01
QP = 30	0.47 ± 0.01	0.45 ± 0.01	0.46 ± 0.02	0.45 ± 0.01

#### 7.4.2.2 Robustness

Robustness against HEVC recompression is determined for the 4 watermarked sequences. These sequences are encoded using  $QP = 25$  and the intra frame only prediction mode. Table 7.3 shows the Hamming distance of the proposed watermarking scheme compared with the high strength embedding scenario. The table determines a similar robustness between the high strength and proposed VA-based algorithm, with only a mean decrease of 0.02 in Hamming Distance. Therefore, the high and proposed watermarking schemes are very close in terms of robustness but greater visual distortion appears on the high strength embedded sequences.

As watermark embedding occurs from modifying the transform coefficients by either -1, 0 or +1, the alteration has minimal impact upon the overall bitstream file size. Table 7.4 shows the bitrate increase and average number of watermarked bits per frame for the proposed watermarking scheme under the condition of  $QP = 25$ . The Stefan sequence contains 90 frames where as the Container, Hall and Soccer sequences have 300 frames. The table shows only a minimal increase in bitrate occurs, around 0.02%, resultant of the proposed watermarking scheme.

A major limitation of both the proposed and high strength methodology is the

Table 7.4: Bitrate increase for the proposed watermarking scheme

<b>Video sequence</b>	<b>Watermarked bits per frame</b>	<b>Bit rate increase</b>
Stefan	89	0.03%
Container	81	0.02%
Hall	78	0.02%
Soccer	85	0.02%

limited robustness toward signal processing attacks. The HEVC compression is not an identically reversible procedure, as the codec may choose different ADI prediction modes or partition the frame differently after a signal processing attack. This is a highly common problem within any compressed domain watermarking algorithm [145].

Figure 7.11, Figure 7.12, Figure 7.13, Figure 7.14 and Figure 7.15 show watermarked I-frames for 3 test images and 2 video frames. For each image or video test the figures display the original frame, proposed VA-based watermarking scheme and high strength watermark scenario. The proposed low visual distortion is apparent in each scenario as supported by the imperceptibility evaluation and subjective testing in Table 7.1 and Figure 7.10, respectively. As with the uncompressed domain video watermarking scheme provided in Chapter 5, the VA-based video watermarking algorithm is best subjectively assessed as an entire sequence rather than from static frames within the sequence. Watermarked HEVC domain VA-based videos are available for viewing at the following website address: <http://svc.group.shef.ac.uk/hevc-va-wm.html>.

## 7.5 Conclusions

This chapter utilises the compressed domain saliency algorithm from Chapter 6 to determine a novel watermarking framework. The proposed algorithms embed watermark data into visually uninteresting coefficients to provide both a joint encoder and a frame domain watermarking scheme. For the encoder framework, I-frame low salient quantised transform coefficients are modified by either +1, 0 or -1 to ensure there is only a minimal increase of 0.02% within the over-

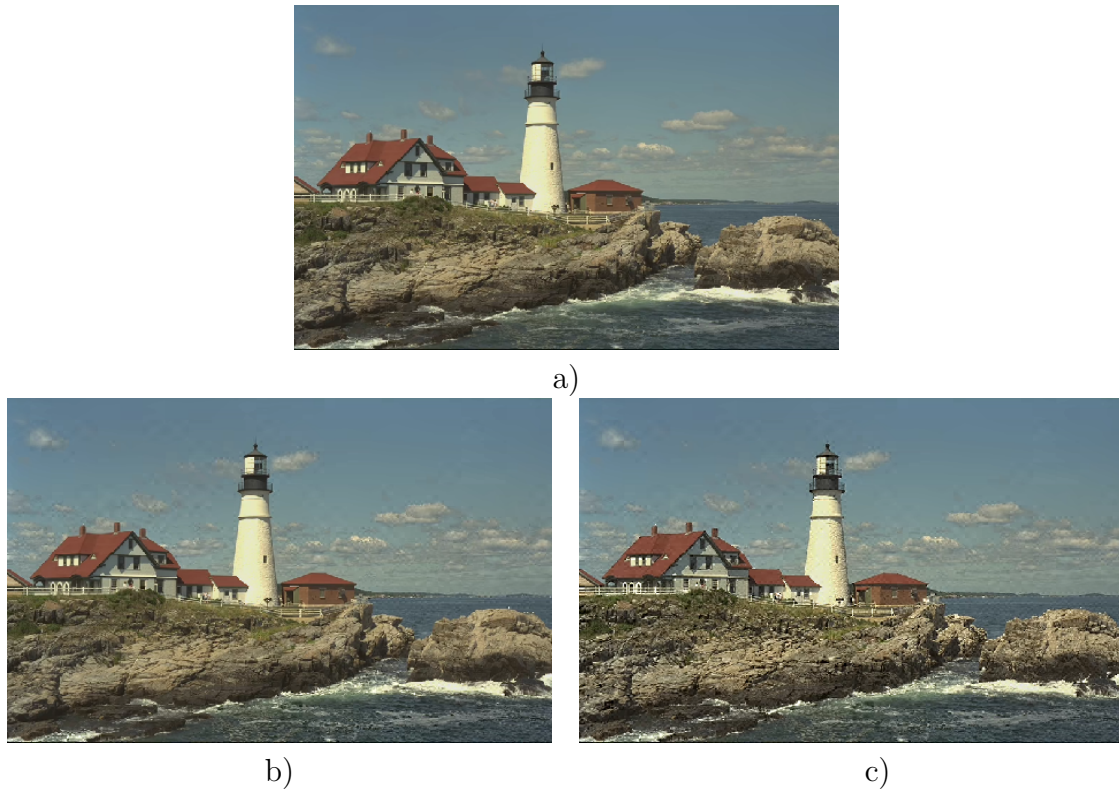


Figure 7.11: Compressed domain watermarking image sequence 1 - a) original image b) proposed watermarking scheme c) high strength watermarked image.

all bit rate. By modifying transform coefficients within regions of low saliency, which also contain a non-zero DC coefficient magnitude, high media visual quality is maintained without causing any conspicuous embedding artifacts. The proposed algorithm has substantial improvements in media quality, verified by a 4dB increase in PSNR and subjective visual testing. The VA-based watermarking scheme provides a maintained robustness, compared with the high strength model, when both watermarking schemes undergo HEVC re-encoding at various QP levels. The frame domain algorithm performance is highly comparable to the VA-based video watermarking study in Section 5.2, providing a 20% increase in robustness against H.264/AVC compression between the low strength and proposed methodology. However, the visual quality between these two scenarios is highly similar, verified by subjective and objective testing. For both VA-based watermarking schemes provided within this chapter, the quality of the model is largely dependant upon the precision of the saliency estimation. Any inaccuracies within the model can severely hinder the overall system performance as data embedded within an eye catching region can be highly noticeable.



a)



b)



c)

Figure 7.12: Compressed domain watermarking image sequence 2 - a) original image b) proposed watermarking scheme c) high strength watermarked image.



a)



b)



c)

Figure 7.13: Compressed domain watermarking image sequence 3 - a) original image b) proposed watermarking scheme c) high strength watermarked image.





a)



b)



c)

Figure 7.14: Compressed domain watermarking video sequence 1 - a) original frame b) proposed watermarking scheme c) high strength watermarked frame.



a)



b)



c)

Figure 7.15: Compressed domain watermarking video sequence 2 - a) original frame b) proposed watermarking scheme c) high strength watermarked frame.

# Chapter 8

## Conclusions and Future Work

In this thesis, numerous VA and watermarking-based research advancements are suggested. This section summarises all the novel propositions and recommends any potential future work.

### 8.1 Conclusions

In Chapter 4, a wavelet based image and video saliency model is provided by merging spatial and temporal features to locate conspicuous regions. Both provided algorithms are highly suitable, but not solely limited towards, applications within the digital watermarking field. The image saliency algorithm detects conspicuous scene regions by combining intensity, colour and orientation contrasts within the wavelet domain. For the video saliency model, salient local object motion provides temporal features maps by estimating and compensating for any global dynamic camera motion. Real-time video saliency estimation is attainable for scenes containing low global camera movement due to the fast and simple algorithm. Subjective evaluation shows the proposed models high performance in terms of accurately estimating salient regions, compared with state-of-the-art existing methodologies. ROC curves provide an objective algorithm evaluation of the proposed model performance which compliments the subjective assessment by an increase of 1.2% and 3.5% in ROC AUC for the image and video models,

respectively. The computational time per frame is also reduced by 50% and 88% for the image and video models, respectively.

A region of interest dictates the most important visible aspects within media, so distortion within these areas will be highly noticeable to any viewer. Chapter 5 incorporates the saliency maps from Chapter 4 within the watermarking framework by embedding greater watermark information within any visually uninteresting scene regions. This provides both non-blind and blind, image and video domain schemes. PSNR, SSIM, VQM, DSCQT and DSIST evaluation shows only minor visual deterioration occurs, resultant of the VA-based watermarking scheme. However, up to a 40% increase in Hamming Distance is possible against robustness to a wide variety of adversary attacks such as: image filtering, JPEG2000 compression and H.264/AVC compression.

Most video sequences exist in an encoded format. In Chapter 6 an HEVC compressed domain saliency framework is provided by uniquely exploiting features within the coding standard. For both intra and inter coded frames, the block partition size, residual energy, intra mode differences and motion vector magnitude are combined using conditional probability to constitute a definitive saliency model. Unlike existing methodologies, the HEVC bitstream does not need to be fully decoded to attain an accurate saliency estimation. The proposed method has a computational time 50% or less, compared with existing methodologies without compromising the accurate saliency estimation, verified by ROC curve evaluation. The saliency model is highly applicable toward low power devices due to the highly efficient algorithm and similarly to the uncompressed domain saliency model, in Chapter 4, the target application is not solely limited toward digital watermarking.

Finally, Chapter 7 examines HEVC domain watermarking approaches and 2 novel blind VA-based algorithms are provided. Firstly, a frame domain watermarking model is presented by combining the saliency model in Chapter 6 with the watermarking framework in Chapter 5. While maintaining the visual quality between the proposed and a constant low strength watermark, a robustness increase against H.264/AVC compression up to 20% is possible. A joint encoder based solution is also provided, where watermark data is embedded within specific quantised AC transform coefficients determined by VA, while only increasing the



overall bitrate by 0.02%. The simulated results show, the proposed VA-based watermarking scheme and constant high strength model maintain a similar robustness against HEVC recompression, however, the visual quality of the proposed algorithm is dramatically higher. This is justified by objective measurements such as PSNR, SSIM and VQM as well as subjective testing.

## 8.2 Key Contributions

Research presented in this thesis produced the following novel key contributions in the field of VA-based watermarking schemes:

- A wavelet domain image and video saliency algorithm is proposed. Both models are highly suitable, but not limited towards, wavelet-based watermarking application.
- A novel VA-based image and video domain watermarking framework is proposed. The watermarking schemes are provided incorporating the novel saliency maps proposed within this thesis.
- Proposed is an HEVC compressed domain saliency algorithm. Saliency maps are generated directly from within the HEVC domain, uniquely exploiting various coding decisions made from within the HEVC codec.
- Proposed is a VA-based joint encoder and frame domain watermarking scheme for HEVC. The compressed domain saliency model provided within this thesis is combined within the watermarking framework, to provide VA-based HEVC watermarking algorithms.

## 8.3 Future Work

There are numerous notable future directions which can be undertaken to extend various aspects of the research within this thesis.

## **VA-based watermark framework refinement by perceptual subband watermark weighting:**

In Chapter 5, an image and video wavelet VA-based watermarking framework is provided by embedding a greater amount of watermark information into the less visually attentive frame portions. This model can be refined by incorporating existing methodologies, based upon perceptual HVS characteristics within the wavelet transform. Barni [138] implements perceptual watermark weighting, based upon the wavelet subband orientation or decomposition level. Incorporation of these HVS based mechanisms within the VA-based watermarking framework design, could potentially enhance the existing models overall performance in terms of joint robustness and imperceptibility, especially within frames containing a wide variety of texture.

## **Provide a framework for HEVC scalable and 3D multiview extensions:**

With current research establishing an HEVC multiview and scalable extension [5] [4], a logical progression for continued studies is to provide a suitable watermarking framework within both 3D stereoscopic and scalable media. By incorporating any potential relationships between depth map information and human VA characteristics, combined within our compressed domain algorithms, watermarking frameworks within the 3D HEVC extension domain can be provided. For the HEVC scalable extension, potential spatial redundancies derived within the scalable codec could be exploited and potentially provide a vital incite into estimating visual saliency across multi-scaled media. This would be of great benefit, providing a universal compressed domain saliency model, regardless of the targeted device.

## **Use a machine learning technique:**

Chapter 6 provides an HEVC domain saliency estimation by combining various extractable features within the codec. By incorporating a machine learning based

technique, the parameters which are most advantageous towards the saliency model can be identified. By implementing a feedback loop, the system can be finely tuned, increasing the accuracy of salient region detection.

### **Saliency model applications:**

This thesis provides saliency maps applicable but not limited towards the watermarking field. Our models are highly adaptable within many other wavelet based applications such as compression, fusion, automatic image resizing and advertising.



## Chapter 9

### Appendix - Additional Results

In this appendix, additional results are shown for the Intra mode Saliency model described in Chapter 6. The results shown are from the MSRA database and show 50 original frames, proposed saliency map and corresponding ground truth frames.









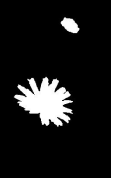


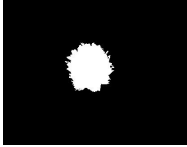



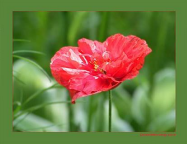











Original Frame	Saliency Map	Ground Truth
		
		
		
		
		
		
		
		
		

Figure 9.1: HEVC Intra-Only mode additional saliency results 1-9 of 50

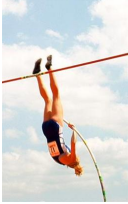


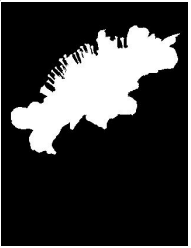



















Original Frame	Saliency Map	Ground Truth
		
		
		
		
		
		
		
		
		

Figure 9.2: HEVC Intra-Only mode additional saliency results 10-18 of 50





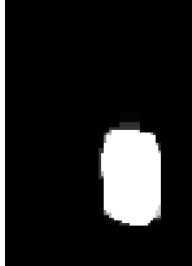
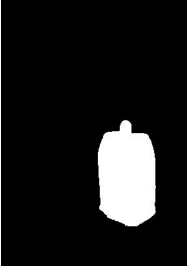












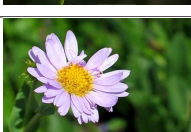

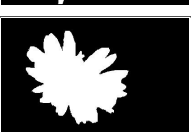






Original Frame	Saliency Map	Ground Truth
		
		
		
		
		
		
		
		
		

Figure 9.3: HEVC Intra-Only mode additional saliency results 19-27 of 50




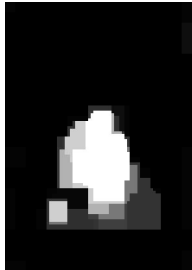























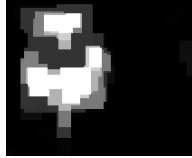

Original Frame	Saliency Map	Ground Truth
		
		
		
		
		
		
		
		
		

Figure 9.4: HEVC Intra-Only mode additional saliency results 28-36 of 50

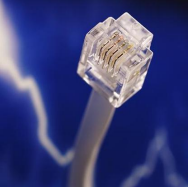





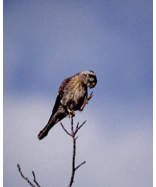
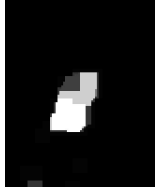



















Original Frame	Saliency Map	Ground Truth
		
		
		
		
		
		
		
		
		

Figure 9.5: HEVC Intra-Only mode additional saliency results 37-45 of 50



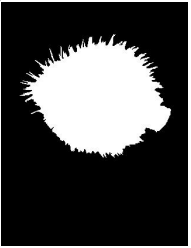
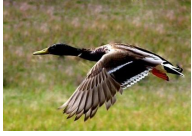











Original Frame	Saliency Map	Ground Truth
		
		
		
		
		

Figure 9.6: HEVC Intra-Only mode additional saliency results 46-50 of 50



# List of Figures

2.1	Video codec watermark possibilities. . . . .	10
2.2	Classical feature based visual attention model structure for imagery or static scenery. . . . .	13
2.3	Classical feature-based visual attention model structure for video. . . . .	15
2.4	Demonstrating the effect of VA within the watermarking framework. a) Original ROI-based watermarked frame b) Extracted watermark c) Patched frame d) Extracted watermark from patched frame. . . . .	17
3.1	Watermark embedding procedure. . . . .	21
3.2	Watermark extraction and authentication procedure. . . . .	22
3.3	General wavelet-based watermarking framework. . . . .	23
3.4	Blind quantisation-based coefficient embedding. . . . .	25
3.5	Visually stimulating features individually defined by a) Intensity contrast b) Colour contrast c) Orientation contrast. . . . .	28
3.6	ROC curve example. . . . .	29

3.7	Demonstrating the effect of VA within the watermarking framework. a) Watermark embedding without incorporating VA b) Watermark embedding incorporating VA. . . . .	30
3.8	Demonstrating the effect of VA-based compression. a) Compressed frame without VA consideration b) Compressed frame in coherence with VAM. . . . .	30
3.9	A collage, automatically formed from multiple salient frame regions.	31
3.10	Automatic frame cropping and resizing a) Original frame b) Automatically resized/cropped in coherence with VAM. . . . .	31
3.11	HEVC comparison with H.264/AVC predecessor. . . . .	32
3.12	HEVC encoder diagram. . . . .	33
3.13	The three possible coding structures within HEVC a) All intra, b) Low delay and c) Random access. . . . .	34
3.14	CU Block partitioning structure with the HEVC codec. . . . .	35
3.15	Possible PU Block partitioning options with the HEVC codec. . .	35
3.16	Possible ADI prediction modes. . . . .	36
3.17	Reference samples $R(x, y)$ used to predict prediction samples $P(x, y)$ for an $N * N$ block. . . . .	37
3.18	Integer and fractional sample positions for luma interpolation. . .	38
3.19	Possible inter merge mode candidates, located a) spatially b) temporally. . . . .	39

3.20	Subjective testing visual quality measurement scales a) DCR continuous measurement scale b)ACR ITU 5-point discrete quality scale. . . . .	41
3.21	Stimulus timing diagram for a) DCR method b) ACR method. . .	41
3.22	MSRA-1000 database a) Test frame from database b) Ground truth ROI frame. . . . .	43
3.23	Video database - 15 thumbnails for each sequence. . . . .	44
4.1	Interpolated LH subbands using a CIF resolution (352x288) image, for each successive wavelet decomposition level. . . . .	47
4.2	Overall Image Saliency Model Block Diagram. . . . .	50
4.3	Image Saliency model state-of-the-art comparison: Row 1 - Original image from MSRA database. Row 2 - Itti model [54]. Row 3 - Rare model [67]. Row 4 - Ngau model [69]. Row 5 - Erdem model [68]. Row 6 - Proposed Method. Row 7 - Ground Truth. . .	51
4.4	Static video frame comparison to state-of-the-art: Row 1 - Original sequence frame. Row 2 - Itti model [54]. Row 3 - Rare model [67]. Row 4 - Ngau model [69]. Row 5 - Erdem model [68]. Row 6 - Proposed Method. Row 7 - Ground Truth. . . . .	52
4.5	ROC curve comparing state-of-the-art image domain saliency algorithms. . . . .	54
4.6	Image saliency model from MSRA-1000 database - column 1: host image - column 2: proposed saliency map - column 3: ground truth map. . . . .	55
4.7	2D+t wavelet decomposition. . . . .	57

4.8	Difference frames after global motion compensation: a) Without local motion b) Containing local motion. . . . .	58
4.9	Motion block Estimation . . . . .	60
4.10	Proposed Saliency Model Block Diagram. . . . .	61
4.11	Temporal Saliency model comparison table: Row 1 - Original frame from sequence. Row 2 - Itti video model. [74] Row 3 - Dynamic model. [134] Row 4 - Proposed Method. Row 5 - Ground Truth. . . . .	63
4.12	ROC curve comparing performance of proposed model. . . . .	64
4.13	Video saliency estimation results sequence 1: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth. . . . .	65
4.14	Video saliency estimation results sequence 2: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth. . . . .	66
4.15	Video saliency estimation results sequence 3: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth. . . . .	66
4.16	Video saliency estimation results sequence 4: Row 1 - Original frame from sequence. Row 2 - Proposed saliency map. Row 3 - Ground truth. . . . .	67
5.1	a) Host image b) VAM saliency map c) Cumulative saliency histogram d) $\alpha$ step graph e) $\alpha$ strength map. . . . .	71
5.2	$\alpha$ strength map examples - A - Original Image B - $\alpha$ strength map. . . . .	72



5.3	Saliency map reconstruction - a) Original host frame b) Watermarked frame embedded using a constant $\alpha_{max}$ c) Host frame saliency map d) Saliency map of watermarked frame e) Original thresholded saliency map. f) Reconstructed saliency map thresholded after blind watermark embedding. . . . .	75
5.4	Visual Attention-based Watermarking Scheme. . . . .	76
5.5	LL low texture subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	78
5.6	LL high texture subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	78
5.7	HL subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	79
5.8	LH subband watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	79
5.9	Subjective Image Watermarking Imperceptibility Testing - top row) non-blind watermarking, bottom row) blind watermarking. . . . .	82
5.10	Robustness to JPEG2000 Compression - non-blind watermarking.	83
5.11	Robustness to JPEG2000 Compression - blind watermarking. . . . .	84
5.12	Overall VA-based Video Watermarking Scheme - Block Diagram.	86
5.13	'Soccer' sequence video watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	87

5.14	'Stefan' sequence video watermarking - a) original image b) low strength watermark image c) VAM-based watermark image d) high strength watermark image. . . . .	88
5.15	Subjective Video Watermarking Imperceptibility Testing - top row) non-blind watermarking, bottom row) blind watermarking. . . . .	90
5.16	Robustness to H.264/AVC compression - average of 4 video sequences: a) non-blind watermarking b) blind watermarking. . . . .	91
6.1	a) Original frame b) PU block structure c) ADI prediction mode number (0-34) d) Average block residual. . . . .	96
6.2	Saliency correlation with a) block size b) intra mode difference c) average block residual. . . . .	97
6.3	Mapped ADI circular difference modes. . . . .	98
6.4	Intra mode circular difference a) Original frame b) ADI intra mode prediction c) Intra mode circular difference. . . . .	100
6.5	a) Histogram of residual magnitudes b) Residual magnitude threshold location c) Thresholded residual magnitude. . . . .	101
6.6	a) Histogram of intra mode differences b) Intra mode difference threshold location c) Thresholded intra mode difference. . . . .	102
6.7	Graphs showing $P(Sal (s \cap R' \cap d'))$ for all 16 possible outcomes. . . . .	103
6.8	Morphological filtering for a 288x416 resolution frame - a) Original frame b) Ground truth frame c) Small kernel size d) Large kernel size e) Suggested kernel size. . . . .	104
6.9	ROC curve comparing morphological kernel size. . . . .	104

6.10	Overall saliency model diagram for an intra encoded frame. . . . .	105
6.11	Inter-Frame prediction features - a) Original frame b) Block structure c) Residual magnitude d) Motion vector difference magnitude.	106
6.12	Correlation between inter mode PU block size and saliency. . . . .	108
6.13	Correlation between inter mode block residual magnitude and saliency.	109
6.14	Motion vector prediction from spatial candidates. . . . .	110
6.15	Correlation between MV difference magnitude and saliency. . . . .	111
6.16	Graphs showing $P(Sal (s' \cap R'_m \cap MVD'))$ for all 16 possible outcomes. . . . .	112
6.17	Overall HEVC saliency model diagram. . . . .	114
6.18	Effect of QP on saliency prediction - a) Block size b) Residual magnitude c) Intra-frame predicted block structure for QP = 15 d) Intra-frame predicted block structure for QP = 30 e) Intra-frame predicted block structure for QP = 45. . . . .	115
6.19	a) 'Container' frame b) Itti model [54] c) Ngau model [69] d) Rare model [67] e) Erdem model [68] f) Proposed model g) Thresholded original frame using proposed model h) ground truth frame. . . . .	116
6.20	ROC curves comparison to existing state of the art methodologies for inter-frame saliency model - a) frame from 'Container' sequence b) entire MSRA-1000 database. . . . .	118
6.21	Intra-frame saliency results - (Row 1: Original image, Row 2: Saliency regions using the proposed model and Row 3: Ground truth). . . . .	119

6.22	Inter-frame Saliency results 1 - (Row 1: Original frame, Row 2: Itti [74], Row 3: Dynamic [134], Row 4: Proposed and Row 5: Ground truth). . . . .	120
6.23	Inter-frame Saliency results 2 - (Row 1: Original frame, Row 2: Itti [74], Row 3: Dynamic [134], Row 4: Proposed and Row 5: Ground truth). . . . .	121
6.24	ROC curve comparison to existing state of the art methodologies for inter-frame saliency model. . . . .	122
6.25	Saliency model limitations - a) original frame b) ground truth c) proposed saliency model . . . . .	123
7.1	Frame Domain Watermark Embedding for an HEVC Encoded Sequence. . . . .	127
7.2	Compressed Domain Watermark Embedding for an HEVC Encoded Sequence. . . . .	127
7.3	Watermark drift problem with compressed domain embedding - a) Original frame b) Watermarked frame. . . . .	128
7.4	Encoder Domain Watermark Embedding for an HEVC Encoded Sequence. . . . .	129
7.5	Transform domain embedding candidates for an 8x8 block. . . . .	131
7.6	Compressed domain VA-based watermark embedding framework within the HEVC encoder. . . . .	132
7.7	Compressed domain VA-based watermark detection framework within the HEVC decoder. . . . .	133
7.8	Subjective Test for Frame Domain Watermark Embedding. . . . .	135

7.9	Robustness to H.264/AVC compression - average of 4 video sequences. . . . .	136
7.10	Subjective Testing for a Joint Encoder Watermark Embedding System. . . . .	137
7.11	Compressed domain watermarking image sequence 1 - a) original image b) proposed watermarking scheme c) high strength watermarked image. . . . .	140
7.12	Compressed domain watermarking image sequence 2 - a) original image b) proposed watermarking scheme c) high strength watermarked image. . . . .	141
7.13	Compressed domain watermarking image sequence 3 - a) original image b) proposed watermarking scheme c) high strength watermarked image. . . . .	141
7.14	Compressed domain watermarking video sequence 1 - a) original frame b) proposed watermarking scheme c) high strength watermarked frame. . . . .	142
7.15	Compressed domain watermarking video sequence 2 - a) original frame b) proposed watermarking scheme c) high strength watermarked frame. . . . .	142
9.1	HEVC Intra-Only mode additional saliency results 1-9 of 50 . . .	150
9.2	HEVC Intra-Only mode additional saliency results 10-18 of 50 . .	151
9.3	HEVC Intra-Only mode additional saliency results 19-27 of 50 . .	152
9.4	HEVC Intra-Only mode additional saliency results 28-36 of 50 . .	153
9.5	HEVC Intra-Only mode additional saliency results 37-45 of 50 . .	154

9.6 HEVC Intra-Only mode additional saliency results 46-50 of 50 . . . 155

# List of Tables

3.1	General watermark properties requirements. . . . .	20
3.2	Applications of digital watermarking. . . . .	21
3.3	Digital watermark attacks 1. . . . .	26
3.4	Digital watermark attacks 2. . . . .	27
3.5	Filter coefficients for luma fractional sample interpolation. . . . .	38
4.1	Computational time comparing state-of-the-art image domain saliency models. . . . .	53
4.2	Computational time comparison of video saliency models. . . . .	62
5.1	PSNR and SSIM values - non-blind watermarking. . . . .	80
5.2	PSNR and SSIM values - blind watermarking. . . . .	80
5.3	Image Filtering Robustness. . . . .	85
5.4	PSNR, SSIM and VQM average of 4 video sequences for blind and non-blind watermarking. . . . .	89

6.1	Generic classification of PU block sizes. . . . .	107
6.2	Intra-frame saliency ROC AUC and computational time comparison between proposed and existing models. . . . .	117
6.3	Inter-frame saliency ROC AUC and computational time comparison between proposed and existing models. . . . .	122
7.1	PSNR, SSIM and VQM of 4 video sequences for HEVC-based frame domain watermarking. . . . .	134
7.2	PSNR, SSIM and VQM of 'Stefan', 'Container', 'Hall' and 'Soccer' sequences. . . . .	137
7.3	Robustness against HEVC re-encoding using various QP for the 'Stefan', 'Container', 'Hall' and 'Soccer' sequences. . . . .	138
7.4	Bitrate increase for the proposed watermarking scheme . . . . .	139



# List of Symbols and Acronyms

## Symbols List 1

Symbol	Description
$\alpha$	Watermark Weight Parameter
$\delta$	Quantisation Step Size
$\xi$	Embedding Process
$\varpi$	Extraction Process
$\tau$	Subband Weightage Parameter
$\psi$	Morphological Filter
$a$	HEVC PU block area
$A$	Interpolated Inter Mode Coefficients
$b$	binary bit
$B$	HEVC Transform Block
$C$	Original Coefficient
$C'$	Watermarked Coefficient
$d$	Intra Mode Circular Difference
$D$	Coefficient Difference
$f$	Frequency
$F$	Temporal Frame
$H$	Hamming
$hl$	HL Subband Feature Map
$HL$	HL Subband coefficients
$HLt$	HL Temporal Subband coefficients
$hh$	HH Subband Feature Map
$HH$	HH Subband Coefficients
$HHt$	HH Temporal Subband Coefficients
$I$	Host Media
$I_0$	Watermarked Media
$In$	Intra Mode ADI Prediction
$\vec{M}$	Motion Vector
$\bar{m}$	Local Maxima Average
$m_d$	Intra Mode ADI Prediction Difference
$M$	Maximum Wavelet Coefficient
$M_g$	Global Maximum
$MVD$	HEVC Motion Vector Difference
$N$	HEVC Block Dimension

## Symbols List 2

Symbol	Description
$P$	HEVC PU predicted Samples
$P(Sal)$	Saliency Probability
$QP$	Quantisation Parameter
$L$	Length of Watermark Sequence
$lh$	LH Subband Feature Map
$LH$	LH Subband Coefficients
$LHt$	LH Temporal Subband Coefficients
$Ln$	Linear Interpolation Operation
$R$	Prediction Residual
$R_m$	Motion Residual
$s$	HEVC PU block partition dimension
$S$	Saliency Map
$T$	Local Threshold
$W$	Host Watermark
$W_0$	Extracted Watermark
$X * Y$	Frame Dimensions

# Acronyms List 1

Acronym	Description
ACR	Absolute Category Rating
ADI	Arbitrary Directional Intra
AMVP	Advanced Motion Vector Prediction
AUC	Area Under Curve
AVC	Advanced Video Coding
CABAC	Context-Adaptive Binary Arithmetic Coding
CU	Coding Unit
D4	Daubechies-4 Wavelet
DCR	Degradation Category Rating
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
DST	Discrete Sine Transform
DSCQT	Double Stimulus Continuous Quality Test
DSIST	Double Stimulus Impairment Scale Test
GOP	Group of Pictures
HD	High Definition
HEVC	High Efficiency Video Coding
HVS	Human Visual System
IDR	Instantaneous Decoding Refresh
IDWT	Inverse Discrete Wavelet Transform
ITU	International Telecommunication Union
JCT-VC	Joint Collaborative Team on Video Coding
JPEG	Joint Photographic Experts Group
JND	Just Noticeable Difference
MCTF	Motion Compensated Temporal Filtering
MSRA	Microsoft Research Asia
MRA	Multi-Resolution Analysis
MV	Motion Vector
POC	Picture Order Count
PU	Prediction Unit
PSNR	Peak Signal to Noise Ratio
QP	Quantisation Parameter

## Acronyms List 2

Acronym	Description
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SSIM	Structural Similarity Index Measure
SURF	Speeded Up Robust Features
TMuC	Test Model under Consideration
TU	Transform Unit
VA	Visual Attention
VAM	Visual Attention Model
VQM	Video Quality Metric
WEBCAM	Watermarking Evaluation Bench for Content Adaptation Modes

# References

- [1] P. N. Tudor, “Mpeg-2 video compression,” *International Journal on Electronics and Communications*, vol. 7, no. 6, pp. 257–264, December 1995. 1
- [2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003. 1
- [3] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, “Overview of the high efficiency video coding (hevc) standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, no. 99, p. 1, 2012. 1, 11, 31, 32, 94, 106
- [4] P. Helle, H. Lakshman, M. Siekmann, J. Stegemann, T. Hinz, H. Schwarz, D. Marpe, and T. Wiegand, “A scalable video coding extension of hevc,” in *Proc. Data Compression Conference (DCC)*, 2013, pp. 201–210. 2, 146
- [5] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, “3d high-efficiency video coding for multi-view video and depth data,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3366–3378, 2013. 2, 146
- [6] D. Bhowmik and C. Abhayaratne, “Morphological wavelet domain image watermarking,” in *European Signal Processing Conference (EUSIPCO)*, 2007, pp. 2539–2543. 8
- [7] —, “A generalised model for distortion performance analysis of wavelet based watermarking,” in *International Workshop on Digital Watermarking (IWDW)*, vol. 5450, 2008, pp. 363–378. 8
- [8] —, “Embedding distortion modeling for non-orthonormal wavelet based watermarking schemes,” in *SPIE Wavelet Applications in Industrial Processing VI*, vol. 7248, 2009, p. 72480K (12 Pages). 8

- [9] X. C. Feng and Y. Yang, "A new watermarking method based on DWT," in *International Conference on Computational Intelligence and Security*, vol. 3802, 2005, pp. 1122–1126. 8, 23
- [10] 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. 8, 24
- [11] L. Xie and G. R. Arce, "Joint wavelet compression and authentication watermarking," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, 1998, pp. 427–431. 8, 23, 24
- [12] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, May 2001. 8, 23, 24
- [13] D. Kundur and D. Hatzinakos, "Toward robust logo watermarking using multiresolution image fusion principles," *IEEE Transactions on Multimedia*, vol. 6, no. 1, pp. 185–198, Feb. 2004. 8, 23
- [14] 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. 8
- [15] 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. 8
- [16] J. R. Kim and Y. S. Moon, "A robust wavelet-based digital watermarking using level-adaptive thresholding," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, 1999, pp. 226–230. 8, 23, 24
- [17] S. Marusic, D. B. H. Tay, G. Deng, and P. Marimuthu, "A study of biorthogonal wavelets in digital watermarking," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 3, Sept. 2003, pp. II–463–6. 8
- [18] 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, no. 1, 2001, pp. 127–131. 8
- [19] F. Huo and X. Gao, "A wavelet based image watermarking scheme," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2006, pp. 2573–2576. 8, 23, 24

- [20] 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. 8
- [21] 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. 8
- [22] T.-S. Chen, J. Chen, and J.-G. Chen, "A simple and efficient watermarking technique based on JPEG2000 codec," in *International Symposium on Multimedia Software Engineering*, 2003, pp. 80–87. 8
- [23] 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. 8
- [24] G. Dorr and J.-L. Dugelay, "A guide tour of video watermarking," *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263–282, 2003. 9
- [25] T. Tokar, T. Kanocz, and D. Levicky, "Digital watermarking of uncompressed video in spatial domain," in *Proc. International Conference Radioelektronika*, 2009, pp. 319–322. 9
- [26] S. N. Merchant, A. Harchandani, S. Dua, H. Donde, and I. Sunesara, "Watermarking of video data using integer-to-integer discrete wavelet transform," in *Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 3, 2003, pp. 939–943. 9
- [27] M. P. Mitrea, T. B. Zaharia, F. J. Preteux, and A. Vlad, "Video watermarking based on spread spectrum and wavelet decomposition," *Wavelet Applications in Industrial Processing II, SPIE*, vol. 5607, no. 1, p. 156164, 2004. 9
- [28] H. Zhang, J. Li, and C. Dong, "Multiple video watermarks based on 3d-dwt and 3d-dct robust to geometrical attacks," in *Proc. International Conference on Automatic Control and Artificial Intelligence (ACAI)*, 2012, pp. 1372–1376. 9
- [29] P. Campisi, "Video watermarking in the 3D-DWT domain using quantization-based methods," in *IEEE International Workshop on Multimedia Signal Processing*, 2005, pp. 1–4. 9

- [30] D. Xu, "A blind video watermarking algorithm based on 3d wavelet transform," in *Proc. International Conference on Computational Intelligence and Security*, 2007, pp. 945–949. 9
- [31] J. Nah, J. Kim, and J. Kim, "International journal on applied mathematics and information sciences," *Wavelet Applications in Industrial Processing II*, vol. 7, no. 6, pp. 2391–2396, 2013. 9
- [32] F. Deguillaume, G. Csurka, J. J. O’Ruanaidh, and T. Pun, "Robust 3d dft video watermarking," in *SPIE Security and Watermarking of Multimedia Contents*, vol. 3657, 1999, pp. 113–124. 9
- [33] H. Zhang, J. Li, and C. Dong, "Multiple video zero-watermarking based on 3d dft to resist geometric attacks," in *Proc. International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012, pp. 1141–1144. 9
- [34] J. H. Lim, D. J. Kim, H. T. Kim, and C. S. Won, "Digital video watermarking using 3d-dct and intracubic correlation," in *SPIE Security and Watermarking of Multimedia Contents III*, vol. 4314, no. 1, 2001, pp. 64–72. 9
- [35] D. Bhowmik and C. Abhayaratne, "Video watermarking using motion compensated 2d+t+2d filtering," in *ACM Workshop on Multimedia and Security*, September 2010, pp. 127–136. 10
- [36] T. Lu, W. Hsu, and P. Chang, "Blind video watermarking for h.264," in *Proc. Canadian conference on Electrical and Computer Engineering*, May 2006, pp. 2353–2356. 11, 125
- [37] M. Noorkami and R. Mersereau, "Compressed-domain video watermarking for h.264," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, september 2005, pp. 890–893. 11, 125, 129, 131
- [38] L. Quan and L. Hong, "Robust video watermarking algorithm based on h.264," in *Proc. International Conference on Wireless Communications, Networking and Mobile Computing, (WiCOM)*, october 2008, pp. 1–3. 11, 125
- [39] T. Chen, S. Liu, H. Yao, and W. Gao, "Robust video watermarking based on dc coefficients of selected blocks," in *Proc. International Conference on Machine Learning and Cybernetics*, vol. 9, 2005, pp. 5273–5278. 11, 125
- [40] Q. Liu and H. Liu, "Robust video watermarking algorithm based on h.264," in *Proc. International Conference on Wireless Communications, Networking and Mobile Computing*, vol. 5200, 2008, pp. 1–3. 11, 125



- [41] W. Pei, Z. Zhendong, and L. Li, “A video watermarking scheme based on motion vectors and mode selection,” in *Proc. International Conference on Computer Science and Software Engineering*, vol. 5, 2008, pp. 233–237. 11
- [42] N. Mohaghegh and O. Fatemi, “H.264 copyright protection with motion vector watermarking,” in *Proc. International Conference on Audio, Language and Image Processing (ICALIP)*, 2008, pp. 1384–1389. 11
- [43] J. Zhang, J. Li, and L. Zhang, “Video watermark technique in motion vector,” in *Proceedings of XIV Brazilian Symposium on Computer Graphics and Image Processing*, October 2001, pp. 179–182. 11
- [44] Z. Liu, H. Liang, X. Niu, and Y. Yang, “A robust video watermarking in motion vectors,” in *Proc. International Conference on Signal Processing*, 2004. 11
- [45] L. Zhang, Y. Zhu, and L. Po, “A novel watermarking scheme with compensation in bit-stream domain for h.264/avc,” in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, march 2010, pp. 1758–1761. 12
- [46] W. Chen, Z. Shahid, T. Sttz, F. Atrousseau, and P. Callet, “Robust drift-free bit-rate preserving h.264 watermarking,” *Multimedia Systems*, pp. 1–15, 2013. 12
- [47] S. Sakazawa, Y. Takishima, and Y. Nakajima, “H.264 native video watermarking method,” in *IEEE International Symposium on Circuits and Systems, (ISCAS)*, 2006, p. 4 pp. 12
- [48] T. Dutta, “Motion compensated compressed domain watermarking,” in *Proc. ACM International Conference on Multimedia*, 2013, pp. 1039–1042. 12
- [49] 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. 12
- [50] 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. 12
- [51] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980. 12, 28, 93, 94, 99
- [52] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar 2001. 12, 14

- [53] Y. Sun and R. Fisher, “Object-based visual attention for computer vision,” *Artificial Intelligence*, vol. 146, no. 1, pp. 77–123, 2003. 12, 93
- [54] 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. 12, 13, 14, 48, 50, 51, 52, 53, 93, 94, 99, 116, 117, 159, 163
- [55] 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. 12, 93, 94
- [56] A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 35, no. 1, pp. 185–207, jan 2013. 12, 13
- [57] L. ZhiQiang, F. Tao, and H. Hong, “A saliency model based on wavelet transform and visual attention,” *Information Sciences*, vol. 53, pp. 738–751, 2010. 12, 13, 14
- [58] 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. 12, 13
- [59] 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. 13
- [60] 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. 13, 30, 84
- [61] 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. 13
- [62] 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. 13
- [63] F. Stentiford, “An estimator for visual attention through competitive novelty with application to image compression,” in *Proc. Picture Coding Symposium*, April 2001, pp. 101–104. 13

- [64] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Advances in Neural Information Processing Systems*, 2007, pp. 545–552. 13
- [65] 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. 13
- [66] G. Kootstra, A. Nederveen, and B. D. Boer, “Paying attention to symmetry,” in *Proc. British Machine Vision Conference (BMVC)*, 2008, pp. 1115–1125. 13
- [67] 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. 13, 14, 51, 52, 53, 94, 116, 117, 159, 163
- [68] 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. 13, 14, 51, 52, 53, 116, 117, 159, 163
- [69] 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. 14, 51, 52, 53, 94, 116, 117, 159, 163
- [70] F. Guraya and F. Cheikh, “Predictive visual saliency model for surveillance video,” in *Proc. IEEE European signal processing conference (EUSPICO 2011)*, 2011, pp. 554–558. 14
- [71] R. Carmi and L. Itti, “Visual causes versus correlates of attentional selection in dynamic scenes,” *Vision Research*, vol. 46, pp. 4333–4345, Dec 2006. 14
- [72] K. Angsar and Z. Li, “Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottomup saliency map in v1,” *Journal of Vision*, vol. 7, no. 7, pp. 1–14, 2007. 14
- [73] O. Meur, P. Callet, and D. Barba, “Predicting visual fixations on video based on low-level visual features,” *Vision Research*, vol. 47, pp. 2483–2498, 2007. 14
- [74] L. Itti and N. Dhavale, “Realistic avatar eye and head animation using a neurobiological model of visual attention,” in *Proc. SPIE International Symposium on Optical Science and Technology*, 2003, pp. 64–78. 14, 28, 61, 62, 63, 119, 120, 121, 122, 160, 164
- [75] Y. Tong, C. Faouzi, G. Fahad, K. Hubert, and T. Alain, “A spatiotemporal saliency model for video surveillance,” *Cognitive Computation*, vol. 3, pp. 241–263, 2011. 14

- [76] L. Itti and K. Christof, “Computational modelling of visual attention,” *Natural Review Neuroscience*, vol. 2, no. 3, pp. 194–203, March 2001. 14
- [77] 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. 15
- [78] M. Dorr, E. Vig, and E. Barth, “Colour saliency on video,” in *Bio-Inspired Models of Network, Information, and Computing Systems*, 2012, vol. 87, pp. 601–606. 15
- [79] C. C. Loy, X. Tao, and G. S. Gong, “Salient motion detection in crowded scenes,” in *International Symposium on Communications Control and Signal Processing, (ISCCSP)*, 2012, pp. 1–4. 15
- [80] 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. 16
- [81] R. Ni and Q. Ruan, “Region of interest watermarking based on fractal dimension,” in *Proc. International Conference on Pattern Recognition*, 2006, pp. 934–937. 16, 17
- [82] 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. 16
- [83] 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
- [84] 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. 16
- [85] 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. 16, 17
- [86] 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. 16, 17

- [87] 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. 16
- [88] A. M. Joshi, A. Darji, and V. Mishra, "Design and implementation of real-time image watermarking," in *Proc. IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2011, pp. 1–5. 20
- [89] J. I. Cox, J. Kilian, F. T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997. 20
- [90] C. Jin and J. Peng, "A robust wavelet-based blind digital watermarking algorithm," *Information Technology Journal*, vol. 5, no. 2, pp. 358–363, 2006. 23, 24
- [91] T. Furon and P. Bas, "Broken arrows," *Journal on Information Security*, vol. 2008, p. 13 pages, 2008. 23
- [92] A. Piper, R. Safavi-Naini, and A. Mertins, "Resolution and quality scalable spread spectrum image watermarking," in *Workshop on Multimedia and Security*, 2005, pp. 79–90. 23
- [93] 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, October 2013. 23, 24
- [94] Q. Gong and H. Shen, "Toward blind logo watermarking in JPEG-compressed images," in *International Conference on Parallel and Distributed Computing (PDCAT)*, 2005, pp. 1058–1062. 24
- [95] 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. 24
- [96] 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
- [97] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, 1998, pp. 2969–2972. 24

- [98] Y. Li, X. Gao, and H. Ji, "A 3d wavelet based spatial-temporal approach for video watermarking," in *Proc. International Conference on Computational Intelligence and Multimedia Applications*, 2003, pp. 260–265. 26, 27
- [99] S. Al-Taweel and P. Sumari, "Robust video watermarking based on 3d-dwt domain," in *Proc. Conference on Convergent Technologies*, 2009, pp. 1–6. 26
- [100] H. Khalilian and I. V. Bajic, "Multiplicative video watermarking with semi-blind maximum likelihood decoding for copyright protection," in *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, 2011, pp. 125–130. 26, 27
- [101] S. A. K. Mostafa, A. S. Tolba, F. M. Abdelkader, and H. M. Elhindy, "Video watermarking scheme based on principal component analysis and wavelet transform," *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 8, pp. 45–52, Aug 2009. 26, 27
- [102] T. Jayamalar and V. Radha, "Survey on digital video watermarking techniques and attacks on watermarks," *International Journal of Engineering and Technology*, vol. 2, no. 12, pp. 6963–6967, 2010. 26, 27
- [103] X. Niu, M. Schmucker, and C. Busch, "Video watermarking resisting to rotation, scaling, and translation," in *SPIE Security Watermarking of Multimedia Contents IV*, 2002, pp. 512–519. 26, 27
- [104] Y. Wang and A. Pearmain, "Blind mpeg-2 video watermarking robust against geometric attacks: a set of approaches in dct domain," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1536–1543, 2006. 26, 27
- [105] H. Jung, Y. Lee, and S. U. Lee, "Rst-resilient video watermarking using scene-based feature extraction," *EURASIP Journal on Advances in Signal Processing*, vol. 14, pp. 2113–2131, Jan. 2004. 26, 27
- [106] <http://www.youtube.com/watch?v=ovEDhFfgdOo>. 26
- [107] J. Haitisma and T. Kalker, "A watermarking scheme for digital cinema," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, 2001, pp. 487–489. 27
- [108] R. Pal, J. Mukherjee, and P. Mitra, "An approach for preparing groundtruth data and evaluating visual saliency models," in *Proc. International Conference on Pattern Recognition and Machine Intelligence*, 2009, pp. 279–284. 28

- [109] J. Swets, *Signal detection theory and ROC analysis in psychology and diagnostics*. Lawrence Erlbaum Associates, 1996. 28
- [110] C. Rother, L. Bordeaux, Y. Hamadi, and A. Blake, “Autocollage,” *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 847–852, july 2006. 30
- [111] F. Stentiford, “Attention based auto image cropping,” in *Workshop on Computational Attention and Applications*, March 21-24 2007, pp. 1–9. 30
- [112] R. Rosenholtz, A. Dorai, and R. Freeman, “Do predictions of visual perception aid design?” *ACM Transactions on Applied Perception*, vol. 8, no. 2, pp. 12:1–12:20, feb 2011. 30
- [113] B.-W. Hong and M. Brady, “A topographic representation for mammogram segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer Berlin Heidelberg, 2003, vol. 2879, pp. 730–737. 30
- [114] N. G. Sadaka and L. J. Karam, “Efficient super-resolution driven by saliency selectivity,” in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 1197–1200. 31
- [115] T. Wiegand and G. J. Sullivan, “The h.264/avc video coding standard,” in *IEEE Signal Processing Magazine*, March 2007, pp. 148–153. 31, 94
- [116] J. Lainema, F. Bossen, W. Han, J. Min, and K. Ugur, “Intra coding of the hevc standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1792 –1801, dec. 2012. 35
- [117] S. Kanumuri, T. Tan, and F. Bossen, “Enhancements to intra coding jctvc-d235,” *JCTVC-D235, Daegu, Korea*, 2011. 36
- [118] A. Alshin, E. Alshina, J. H. Park, and W. J. Han, “Dct based interpolation filter for motion compensation in hevc,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 8499, Oct. 2012. 37
- [119] O. Kwon and C. Lee, “Objective method for assessment of video quality using wavelets,” in *IEEE International Symposium on Industrial Electronics (ISIE 2001)*, vol. 1, 2001, p. 292295. 39
- [120] 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. 40

- [121] A. B. Watson, “Visual optimization of dct quantization matrices for individual images,” in *American Institute of Aeronautics and Astronautics (AIAA)*, vol. 9, 1993, p. 286291. 40
- [122] H. G. Koumaras, “Subjective video quality assessment methods for multimedia applications,” Geneva, Switzerland, Tech. Rep. ITU-R BT.500-11, april 2008. 40, 41
- [123] A. B. Watson, J. Hu, and J. F. M. Iii, “Dvq: A digital video quality metric based on human vision,” *Journal of Electronic Imaging*, vol. 10, pp. 20–29, 2001. 41
- [124] 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. 42
- [125] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, “Saliency-based video segmentation with graph cuts and sequentially updated priors,” in *Proc. IEEE International Conference on Multimedia and Expo*, 2009, pp. 638–641. 43
- [126] H. Wilson, “Psychophysical models of spatial vision and hyper-acuity,” *Spatial Vision*, vol. 10, pp. 64–81, 1991. 45
- [127] 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. 48
- [128] S. Choi and W. W. J., “Motion-compensated 3-d subband coding of video,” *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, 1999. 56
- [129] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. Schaar, J. Cornelis, and P. Schelkens, “In-band motion compensated temporal filtering,” *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 653–673, 2004. 56
- [130] D. Bhowmik, “Robust watermarking techniques for scalable coded image and video,” Ph.D. dissertation, The University of Sheffield, 2011. 56
- [131] R. Jin, Y. Qi, and A. Hauptmann, “A probabilistic model for camera zoom detection,” in *Proc. International Conference on Pattern Recognition*, vol. 3, 2002, pp. 859–862. 60
- [132] H. Bay, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” in *In ECCV*, 2006, pp. 404–417. 60



- [133] D. Mahapatra, S. Winkler, and S. Yen, “Motion saliency outweighs other low-level features while watching videos,” in *Proc. SPIE conference on Human Vision and Electronic Imaging*, vol. 6806, 2008, pp. 1–10. 61, 107
- [134] X. Hou and L. Zhang, “Dynamic visual attention: Searching for coding length increments,” in *Advances in neural information processing systems*, 2008, pp. 681–688. 61, 62, 63, 119, 120, 121, 122, 160, 164
- [135] D. Aggarwal and K. S. Dhindsa, “Effect of embedding watermark on compression of the digital images,” *Computing Research Repository*, vol. 1002, 2010. 73
- [136] S. Pereira, S. Voloshynovskiy, M. Madueno, S. M.-Maillet, and T. Pun, “Second generation benchmarking and application oriented evaluation,” in *International Workshop on Information Hiding*, vol. 2137, 2001, pp. 340–353. 77
- [137] 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. 77
- [138] M. Barni, F. Bartolini, and A. Piva, “Improved wavelet-based watermarking through pixel-wise masking,” *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, may 2001. 77, 146
- [139] J. Kim, C. Yi, and T. Kim, “Roi-centered compression by adaptive quantization for sports video,” *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, may 2010. 94
- [140] Z. Li, T. Fang, and H. Huo, “A saliency model based on wavelet transform and visual attention,” *Information Sciences*, vol. 53, pp. 738–751, 2010. 94
- [141] S. Mitra and G. Sicuranza, “Region-based filtering of images and video sequences: A morphological viewpoint,” in *Nonlinear Image Processing*. Academic Press, 2001, pp. 249–288. 102
- [142] T. Yokoyama, T. Iwasaki, and T. Watanabe, “Motion vector based moving object detection and tracking in the mpeg compressed domain,” in *International Workshop on Content-Based Multimedia Indexing*, 2009, pp. 201–206. 108
- [143] T. Nguyen and D. Marpe, “Performance analysis of hevc-based intra coding for still image compression,” in *Picture Coding Symposium (PCS)*, 2012, pp. 233–236. 116
- [144] M. Noorkami, “Secure and robust compressed domain video watermarking for h.264,” Ph.D. dissertation, Universite de Neuchatel, 2007. 127, 129

- [145] D. Kim, Y. Choi, H. Kim, J. Yoo, H. Choi, and Y. Seo, “The problems in digital watermarking into intra-frames of h.264/avc,” *Image and Vision Computing*, vol. 28, no. 8, pp. 1220–1228, 2010. 139