

## Visual attention for region of interest coding in JPEG 2000

Andrew P. Bradley<sup>a,\*</sup> and Fred W.M. Stentiford<sup>b</sup>

<sup>a</sup> *Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP),  
School of Information Technology and Electrical Engineering, The University of Queensland,  
St Lucia, Qld 4072, Australia*

<sup>b</sup> *UCL @ Adastral Park, Martlesham Heath, Ipswich, UK*

Received 8 January 2002; accepted 16 May 2003

---

### Abstract

This paper details work undertaken on the application of an algorithm for visual attention (VA) to region of interest (ROI) coding in JPEG 2000 (JP2K). In this way, an “interest ordered” progressive bit-stream is produced where the regions highlighted by the VA algorithm are presented first in bit-stream. The paper briefly outlines the terminology used in JP2K, the packet structure of the bit-stream, and the methods available to achieve ROI coding in JP2K (tiling, coefficient scaling, and code-block selection). The paper then describes how the output of the VA algorithm is post-processed so that an ROI is produced that can be efficiently coded using coefficient scaling in JP2K. Finally, a two alternative forced choice (2AFC) visual trial is undertaken to compare the visual quality of images encoded using the proposed VA ROI algorithm and conventional JP2K. The experimental results show that, while there is no overall preference for the VA ROI encoded images; there is an improvement in perceived image quality at low bit rates (below 0.25 bits per pixel). It is concluded that an overall increase in image quality only occurs when the increase in quality of the ROI more than compensates for the decrease in quality of the image background (i.e., non-ROI).

© 2003 Elsevier Science (USA). All rights reserved.

*Keywords:* Image compression; JPEG 2000; Region of interest; Visual attention; Two alternative forced choice visual trial

---

\* Corresponding author. Fax: +61-7-3365-3684.

*E-mail address:* A.Bradley@cssip.uq.edu.au (A.P. Bradley).

## 1. Introduction

With the introduction of third generation (3G) mobile phones there will be an increasing demand for the efficient transmission of multi-media data, such as speech, audio, text, images, and video. Of these multi-media data types image and especially video data will provide the toughest challenges because of their high bandwidth and user expectations in terms of high quality of service. Therefore, to enable the successful adoption of 3G applications, the transmission of multi-media data must be at high compression ratios and be of a perceptually high quality. One potential method of improving perceived image quality, which has received some previous attention, is to vary the bit-rate, and therefore the quality, spatially across the image. In this way, the most visually important regions of an image are encoded to a higher quality and bits are not *wasted* on visually unimportant areas in the background. This is often referred to as region of interest (ROI) coding (Daly et al., 1998; Privitera and Stark, 2000). In this paper we shall investigate the ROI coding capabilities of JPEG 2000 (JP2K) in an application where the ROI is automatically detected using an algorithm for visual attention (VA). It is the aim of this paper to establish the conditions under which VA ROI coding provides an *overall* improvement in perceived image quality compared to convention JP2K. In particular, the paper reports results of how human observers judge overall image quality when comparing images with constant spatial quality and images where the quality varies spatially.

JPEG 2000 is the latest image and video compression standard developed by the International Organisation for Standardisation/International Telecommunications Union (ISO/ITU-T). JP2K has been designed complement the current JPEG standard (Pennebaker, 1993) by providing improved compression performance and a rich set of new functionalities (Christopoulos et al., 2000). JPEG 2000 Part I, the core-coding algorithm, became an international standard in December 2000 (Taubman and Marcellin, 2001) and further work is ongoing to tailor the standard to specific applications, such as medical imaging and video coding. JP2K provides, in a single bit-stream, a broad set of functionalities, such as: progressive transmission by resolution, quality, component, or location; random access; loss-less to lossy compression; and error tolerance (Christopoulos et al., 2000). The specific functionality investigated in this paper is the ability of JP2K to encode a ROI in an image with more detail than the background. In this paper JP2K ROI coding is used in combination with an algorithm for VA (Stentiford, 2001a,b) to provide a progressive bit-stream where the regions highlighted by the VA algorithm are coded as an ROI and presented first in the bit-stream. This results in an *interest ordered* bit-stream where any valid bit-stream termination results in an image where the ROI is coded to a higher quality than the background. The efficacy of this technique is then evaluated using a visual trial to determine under what conditions it provides an increase in *overall* perceived image quality over conventional JP2K.

## 2. The JPEG 2000 standard

The philosophy behind JP2K is that an image is encoded once (perhaps to the loss-less level) and then it is then up to the decoder (or a transcoder) to extract a sub-set of the bit-stream from which to reconstruct an image of the required spatial resolution and quality. To enable this functionality JP2K has a highly structured bit-stream, with a main header that describes the image and coding parameters used and a series of packet headers that describe the sub-set of wavelet coefficients contained in the packet. Some of the terminology used can be briefly defined as follows:

*Tile.* Consists of a whole image or a rectangular (non-overlapping) sub-image.

*Component.* (Normally) a single colour plane of an image.

*Decomposition level.* A collection of wavelet sub-bands that have the same span with respect to the original samples, i.e., they are from the same resolution.

*Code-block.* A rectangular grouping of wavelet coefficients from the same sub-band and tile-component.

*Precinct.* A sub-division of a tile-component within a resolution.

*Layer.* A collection of encoded bit-planes, or sub-bit planes, from one, or more, code-blocks of a tile-component. Layers have an order for encoding and decoding that must be preserved.

The data contained in each packet in the bit-stream is from one layer of one decomposition level of one component of an image tile. The basic structure of a JP2K packet is illustrated in Fig. 1. For further details of the bit-stream syntax and the JP2K coding and decoding process, refer to Christopoulos et al. (2000) or Taubman and Marcellin (2001).

### 2.1. Region of interest coding in JPEG 2000

Region of interest (ROI) coding is important in applications where certain parts of an image are of a higher importance than the rest of the image. In these cases the

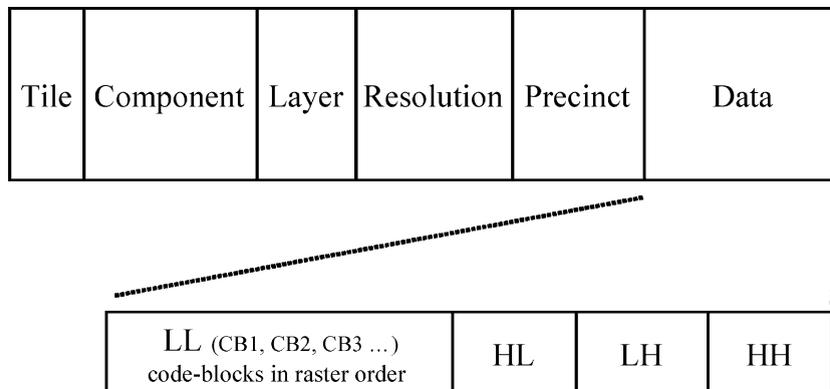


Fig. 1. JP2K basic packet header format.

ROI is decoded with higher quality and/or spatial resolution than the background. Two examples of applications that can be enhanced using ROI coding are:

1. *Client/server applications.* The server initially transmits a low quality/resolution version of an image and the client selects an area of the image as an ROI. The server then transmits only the data needed to refine (i.e., improve the spatial resolution/quality) of the ROI. This is a useful feature to offer when browsing image databases, or in tele-medicine, as it means the client need not download the whole image at the highest resolution (Santa-Cruz et al., 1999).
2. *Facial images.* When browsing a digital photography album, or talking on a video telephone, it is often the case that we are most interested in the face (or lips) of the person in the image. Using an automated face detection algorithm the face can be encoded as an ROI and therefore displayed more accurately than the background (Daly et al., 1998).

There are three mechanisms available in JP2K to encode and decode images with varying spatial quality and therefore implement ROI coding: tiling, code-block selection, and coefficient scaling. These are described and evaluated in more detail in Bradley and Stentiford (2002), however, it is worthwhile to summarise some of the results from these previous rate-distortion experiments.

When using the tiling facility of JP2K to implement ROI coding the bit-rate of each tile is effectively varied so that the “interesting” tiles are coded to a higher quality than “background” tiles. Therefore, in this mode, the ROI is effectively defined to be the sub-set of image tiles that are decoded at the highest bit-rate. This means that tiling requires no ROI shape information at the decoder, as it is simply a re-ordering of the bit-stream. Tiling also reduces the memory requirements of the discrete wavelet transform (DWT), however, this reduces the compression efficiency of the DWT (Bradley and Stentiford, 2002) and can result in visually annoying *block artefacts* at low bit-rates (that can only be reduced with increased decoder complexity).

Code-block selection involves parsing a JP2K bit-stream (encoded to a loss-less or visually loss-less level) and extracting the packets that contain the code-blocks required to decode the ROI. Alternatively, it can be viewed as being equivalent to having used a spatially varying distortion measure in the rate-distortion optimisation during the compression process. Code-block selection has the advantages that no ROI shape information is required at the decoder and there is no compression overhead as only bit-stream re-ordering is required. In addition, the ROI is not embedded in bit-stream and can therefore be extracted from non-ROI bit-streams, which makes it suitable for client/server applications. However, the disadvantages include that the shape of the ROI must be a sub-set of code-blocks, which at higher levels of the DWT relate to image areas of increasing size. Therefore, ROI refinement is slower than coefficient scaling and therefore may not be suitable for low bit rate applications (Bradley and Stentiford, 2002).

Coefficient scaling has the advantages that the ROI can be of arbitrary shape and the *maxshift* algorithm (provided in Part I of the JP2K standard (Christopoulos et al., 2000; Santa-Cruz et al., 1999)) requires no shape information to be transmitted to the decoder. In the *maxshift* algorithm the set of wavelet coefficients that belong to the ROI is first determined by mapping the ROI from the spatial domain to the

wavelet domain (a process that is dependent on the wavelet filters used). Then the magnitude of the largest wavelet coefficient *not* contained in the ROI, i.e., in the background, is found (and put in the bit-stream) and all background coefficients are scaled down by a value just larger than this. In this way, all of the wavelet coefficients that are not part of the ROI now have a magnitude of less than one. Therefore, coefficients belonging to the background can be identified by the decoder and scaled up before the inverse DWT is applied. Disadvantages of the *maxshift* algorithm include a doubling of the dynamic range required during encoding and decoding and an increased coding overhead, due to code-blocks on the background/ROI boundary being coded twice (once for the ROI and once for the background). However, this overhead can be minimised by minimising the number of ROIs and ensuring that they have simple boundaries (ideally on code-block boundaries) (Bradley and Stentiford, 2002). In addition, with the *maxshift* algorithm there is a potential for over-coding of the ROI, as the ROI is coded to the least significant bit (LSB) before any coefficients from the background are coded (Bradley and Stentiford, 2002). However, increasing the step size of the quantiser can reduce this effect, but this results in a reduced maximum quality level for both the ROI and background. Part II of the standard allows for arbitrary bit-shifts which ameliorates this problem, but this does require the shape of the ROI to be transmitted to the decoder (Taubman and Marcellin, 2001).

Selection of the most appropriate JP2K ROI encoding method for a particular application is dependent upon a number of factors: the desired bit-rate; relative ROI/background importance; the shape and size of the ROI; and whether the ROI is fixed or can be selected by the user (Bradley and Stentiford, 2002). For client/server applications it is essential to be able to extract any ROI from an encoded image, in which case code-block selection is the best method to use (Santa-Cruz et al., 1999). However, in the applications we will look at in this paper the ROI is known and fixed. Therefore, it is desirable to have the ROI embedded in the bit-stream using coefficient scaling. In addition, the ROI is assumed to be of primary importance and we desire to receive the ROI as early as possible in the bit-stream, which is equivalent to a low bit-rate application. Therefore, we shall use the *maxshift* algorithm provided in Part I of the standard. In addition, we shall tailor the JP2K ROI coding to our particular requirements by using: small ( $16 \times 16$ ) code-blocks for fast ROI refinement; a 5 level Daubechies 9/7 wavelet transform for high compression (lossy) coding with the lowest level of the DWT defined to be part of the ROI; and an increased quantiser step size (of 0.03125 which is four times the default) to prevent ROI over-coding.

### 3. Visual attention

Visual attention is a subconscious process of the human visual system whereby when we first view a visual scene our focus of attention is drawn to certain objects in that scene. Often these objects *stand out* from the background by being *different* to the background or by being familiar and recognisable. The physiological and

psychological processes of human VA are not fully understood, however, there are a number of phenomena that are known to draw our attention in a scene, e.g., movement, shape, colour, or contrast in relation to the background (Stentiford, 2001b). If we can reliably predict where VA will be first drawn to in an image then it makes sense to code these areas first in a progressive bit-stream or alternatively to code these areas with increased spatial resolution/quality (Stentiford, 2001a). Face detection can be seen as a specific example of VA, although it does rely on higher level visual processing not normally explicitly modelled in VA algorithms.

The VA algorithm used in this work is based upon the dissimilarity between neighbourhoods in an image (Stentiford, 2001b). The algorithm is based upon an evolutionary process in which randomly generated neighbourhood constructs are tested and retained according to their discriminatory performance (Stentiford, 2001a). Pixels that are surrounded by novel structures are distinguished from those that are in neighbourhoods that have many similar counterparts elsewhere in the image. Populations of individual pixels or neighbourhoods are evolved that characterise the regions of interest in an image. In this way, the method is able to avoid the exclusion of whole categories of sub-images by the constraints of specific analysis rules or the use of pre-ordained templates. When an image is input into the algorithm a VA map image is produced that defines the VA associated with each pixel in the input image (on a scale [0,1]).

### 3.1. *Visual attention for region of interest coding*

As was demonstrated in the previous rate distortion experiments (Bradley and Stentiford, 2002), it is important to reduce the overhead associated with the maxshift algorithm in order to ensure maximum coding efficiency. This can be achieved by:

1. Ensuring the ROI is  $<1/4$  of the area of the whole image, which ensures that there is a visible improvement in the ROI quality compared to the background.
2. Reducing the number of regions of interest (to two or less).
3. Ensuring region boundaries are reasonably regular (smooth).

The last two constraints ensure that the overhead associated with maxshift ROI coding is minimised as this minimises the number of code-blocks that contain coefficients from both the ROI and the background.

The VA map produced in Stentiford (2001a,b) can be complex and highlight numerous small regions. In order to reduce the number of regions and also merge adjacent regions, further processing is therefore required. Fig. 2 shows a flow chart of the processing undertaken to reduce the VA map to a single elliptical ROI and Fig. 3 shows the further processing undertaken to find two elliptical regions of interest if the single ROI is found to be unsuitable for effective coding.

The operations illustrated in Fig. 2 are described as follows: First, the VA map is morphologically filtered so as to reduce, or remove, small regions of VA that may be due to noise or isolated edges in the original image. This is done using grey-level morphology, specifically an opening operation (i.e., an erosion followed by a dilation, see Gonzalez and Woods (1993) for details) with an approximately circular structuring element of initial size of  $5 \times 5$  pixels.

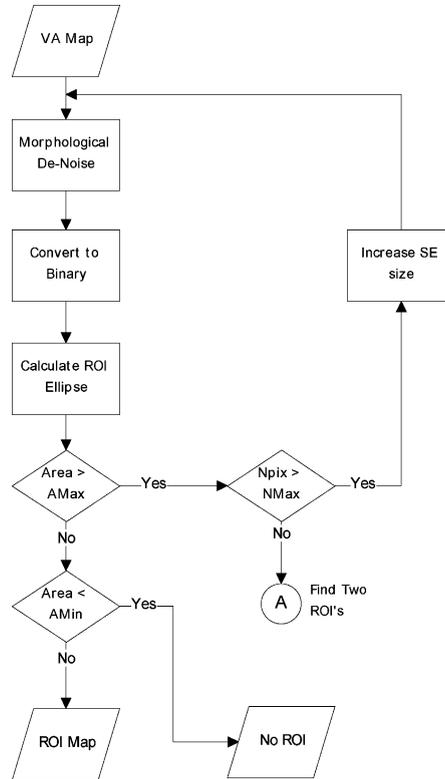


Fig. 2. Flow chart to convert the VA map to a single elliptical ROI.

Second, the VA image is converted to binary by thresholding the cleaned grey-level VA map. Currently a global threshold of 0.7 is used, a level that is thought to be appropriate as it only keeps truly interesting pixels and therefore minimises the size of the resultant ROI. Also note that it was found that performing the morphological filtering followed by thresholding the image to binary was more stable to variations in the free parameters (i.e., size of the structuring element and selected threshold level) than converting the image to binary and then performing binary morphological filtering.

Third, the mean and covariance of the  $(x, y)$  locations of all of the remaining non-zero pixels in the VA map are calculated. An ROI image of the same size as the original image is then generated containing a normalised Gaussian function, centred at the pixel location specified by the mean and with a spread specified by the covariance of the remaining non-zero pixels. This image is then thresholded (at a value of  $p = 0.005$ , so that the elliptical ROI contains virtually all of the remaining interesting pixels) to give a binary image containing an ellipse centred and oriented so as to represent the ROI in the image. Further, processing is then undertaken to ensure that the area of the ROI is not too large ( $>25\%$  of the image area) or too small ( $<1\%$

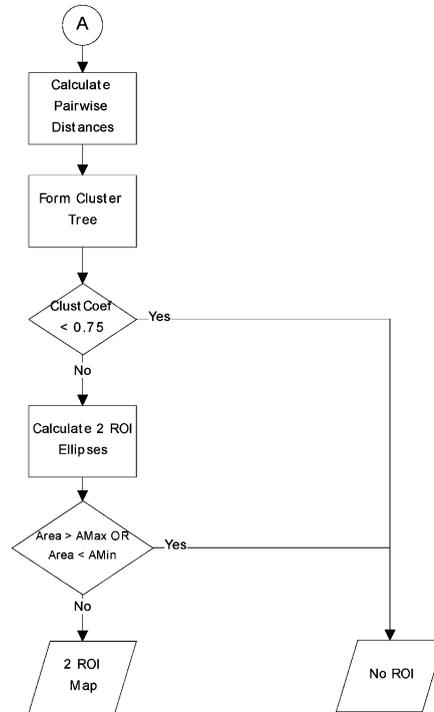


Fig. 3. Flow chart of further processing required for two ROI ellipses.

of the image area). If the ROI is too small it is ignored and no ROI coding is used. If the ROI is too large, the number of non-zero pixels in the VA map is checked to ensure that the computational burden of the following clustering algorithm is acceptable (in the current implementation the maximum number of pixels that can be clustered is fixed at 1500). If there are too many non-zero pixels in the VA map the diameter of the structuring element used in the morphological filtering is increased by one (i.e., initially from 5 to 6 pixels) and the original VA map is filtered and thresholded again. If there are an acceptable number of non-zero pixels in the processed VA map then it is further analysed in an attempt to find two regions of interest with a combined area  $< 25\%$  of the image area. This process is illustrated in Fig. 3 and described in more detail below.

The previous processing attempted to fit a single elliptical ROI to the processed VA map. If successful, this is the most desirable outcome, as the overhead associated with coding a single ROI in the JP2K bit-stream will be minimized. However, if the ROI found is too large to be coded as a single ROI, there is a possibility that it can be efficiently coded as two ROI ellipses. In order to test this hypothesis it is necessary to perform cluster analysis on the remaining non-zero pixels in the processed VA map to see if they can be considered to form two clusters. Note, it is possible to extend this processing to find three or more clusters, however, the overhead associated

with the ROI coding will increase and it will become increasingly difficult to produce an image of superior visual quality.

First, the Euclidean distance, in  $(x, y)$  space, between each pair of non-zero points in the VA map is calculated and stored as a distance matrix. Note, that if there are  $m$  non-zero points in the VA map there are  $m \cdot (m - 1)/2$  pairs of points, i.e., distances in the matrix. Therefore, with the limit ( $N_{\max} = 1500$ ) set in Fig. 2 there will be a maximum of around 1 million distances to analyse, a large, but not unreasonable, number.

Once the distance matrix has been calculated, pairs of non-zero pixels in the VA map are grouped into a binary, hierarchical cluster tree based on the distances between them. In this way, the most similar pairs of pixels are linked together first, being replaced with their mean location so that clusters of clusters can form. The Mathworks Inc. (1999). Once the cluster tree has been formed the cophenetic correlation coefficient, shown in Eq. (1), is calculated. This compares the distances in the cluster tree to the original pair-wise distance data

$$c = \frac{\sum_{i < j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i < j} (Y_{ij} - y)^2 \sum_{i < j} (Z_{ij} - z)^2}}. \quad (1)$$

In Eq. (1)  $Y_{ij}$  is the distance between objects  $i$  and  $j$  in  $Y$  (the distance matrix),  $Z_{ij}$  is the distance between objects  $i$  and  $j$  in  $Z$  (the hierarchical cluster tree), and  $y$  and  $z$  are the averages of  $Y$  and  $Z$ , respectively. A cophenetic coefficient above a threshold value of 0.75 indicates that the cluster tree is a reasonably accurate representation of the distance data. A cophenetic coefficient below this value indicates that the non-zero pixels are probably evenly spread over the image and so do not form appropriate clusters. Therefore, it can be concluded that there are no valid multiple ROI clusters, and the original ROI found (which was too large) is likely to be the only valid cluster. Note, that as we are only attempting to find two clusters if a single cluster produces an ROI that is too large, the specific clustering algorithm used is relatively unimportant and most commonly used clustering algorithms will produce similar results. The binary, hierarchical cluster tree used here was selected because of its conceptual simplicity, wide spread use, and theoretical basis. The operation of the proposed algorithm is further illustrated below with the aid of example image containing two regions of interest.

### 3.2. An illustrated example

Fig. 4 shows an example of the original cycles image and its associated (grey-scale) VA map. Note how the cycles VA map contains a large number of areas with high VA, including those associated with the horizon and bushes in the background. These areas are probably not as visually important as the two cyclists and certainly could not be efficiently coded as a JP2K ROI.

The VA map after the grey-level opening and binary thresholding is shown in Fig. 5. It can be seen that the morphological filter has reduced the VA associated with the background objects as they tend to be small, isolated areas and so are



Fig. 4. Cycles image and VA map.

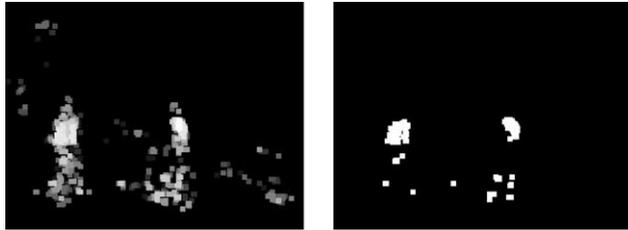


Fig. 5. VA map after morphological opening and thresholding.

not supported by the morphological structuring element. After the thresholding operation the remaining non-zero pixels are primarily associated with the main regions of interest in the image, i.e., the two cyclists.

Fig. 6 shows the ellipse formed from the mean and covariance of all of the non-zero pixels remaining in the image. This ROI has an area  $>25\%$  of the total image area and so clustering is performed on the non-zero pixels. The two main clusters found using the hierarchical binary cluster tree are also shown in Fig. 6.

The ROI ellipses formed using the mean and covariance of each of the two clusters of non-zero pixels are shown in Fig. 7 and have been overlaid on the original image to illustrate the accuracy of the technique in this case. It can be seen that the two ROI ellipses found are a good representation of the cyclists in the original image and they should be effectively coded as two regions of interest by JP2K.

Fig. 8 shows the cycles image, coded at 1 bpp, using both the default JP2K (left) and JP2K with the ROI found using the proposed VA post-processing algorithm

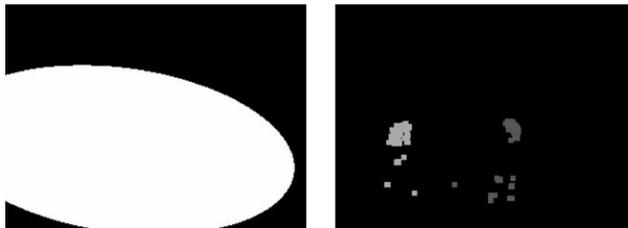


Fig. 6. Single ROI and two clusters of VA non-zero pixels.

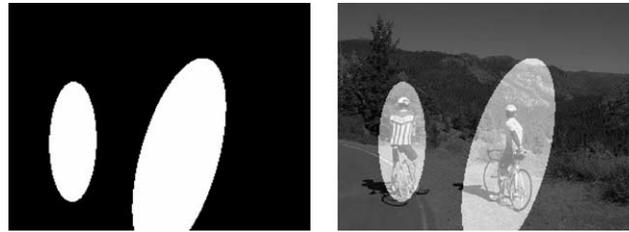


Fig. 7. Two ROI image and superimposed on the original image.



Fig. 8. Comparison of JP2K and VA ROI coding on the cycles image (1 bpp).

(right). It can be seen that there is an improvement in image quality centred on the ROI (the cyclists), although this is achieved at the expense of reduced quality (i.e., blurring) in the background of the image. In order to determine whether the *overall* perceived quality of the VA ROI images is superior to that of those coded using conventional JP2K (at a number of bit-rates) a formal visual trial is undertaken.

#### 4. Visual trial

The visual trial is based upon six images, namely: boat, cycles, beach, helicopter, land, and road sign. These images were chosen to have a reasonably varied content, whilst still containing one or two primary objects that could be considered to be more important (visually interesting) than the background. As one of the potential applications for this technology is 3G multi-media mobile communications the images are low resolution (between  $284 \times 214$  and  $640 \times 480$  pixels) and of commercial quality (i.e., JPEG compressed). The images selected for the visual trial are not intended to be representative of any particular potential application, but were chosen solely to judge the efficacy of VA ROI coding in JP2K. Fig. 9 shows the original boat image, its VA map, and the resultant ROI ellipse. The resultant ROI does not wholly enclose the boat, mainly capturing the mast of the yacht, but successfully ignores the mountain skyline and clouds highlighted by the VA algorithm.

Fig. 10 shows the original beach image, its VA map, and the resultant ROI ellipse. The proposed VA ROI algorithm correctly highlights the people on the beach as the primary ROI, despite the fact that the VA algorithm did not highlight the t-shirt of



Fig. 9. Boat image, VA map, and VA ROI.



Fig. 10. Beach image, VA map, and VA ROI.

the main person (presumably because of its similarity in colour to the background sand).

Figs. 4–7 show the cycles image and associated VA maps and the resultant ROI ellipses. Fig. 11 shows the original helicopter image, its VA map, and the resultant ROI ellipse. It correctly highlights the helicopter as the primary ROI, ignoring the VA associated with the mountain skyline, which has a boundary too complex to be reliably coded as a JP2K ROI.

Fig. 12 shows the original land image, its VA map, and the resultant ROI ellipse. The VA ROI algorithm highlights the waterline as the primary ROI.

Fig. 13 shows the original road sign image, its VA map, and the resultant ROI ellipse. It shows that the proposed algorithm correctly highlights the road sign as the primary ROI, ignoring the isolated VA associated with the bushes in the background.

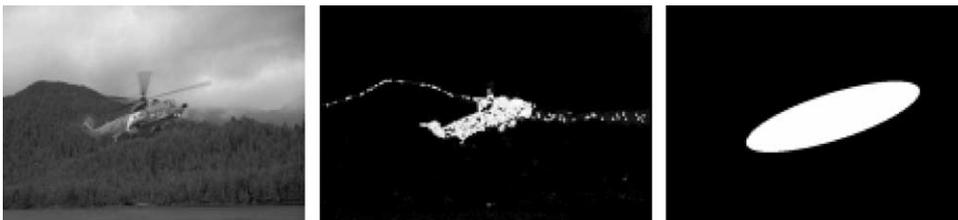


Fig. 11. Helicopter image, VA map, and VA ROI.

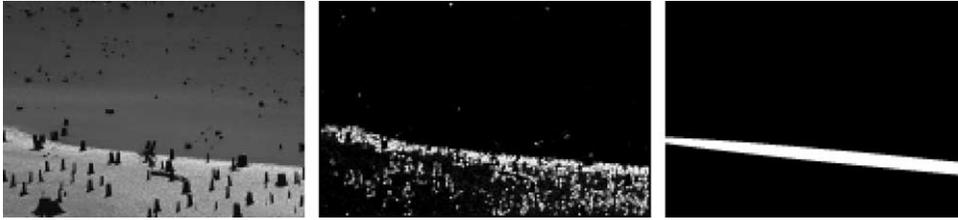


Fig. 12. Land image, VA map, and VA ROI.

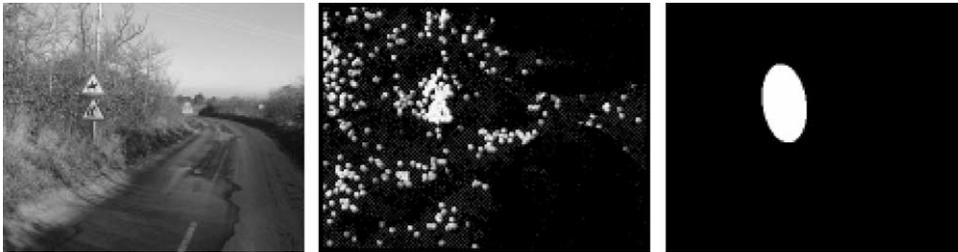


Fig. 13. Road sign image, VA map, and VA ROI.

#### 4.1. Experimental methodology

The purpose of the visual trial is to directly compare images encoded to a specified bit-rate using standard JP2K and JP2K ROI coding where the ROI is determined using the proposed VA ROI algorithm. The comparisons are made at four logarithmically spaced bit-rates (and hence varying image qualities) of 0.125, 0.25, 0.5, and 1 bits per pixel (bpp). A two alternative forced choice (2AFC) methodology was selected because of its simplicity, i.e., the observer views the two images and then selects the one preferred, and so there are no issues with scaling opinion scores between different observers (Eckert and Bradley, 1998). There were ten observers (8 male and 2 female) all with good, or corrected, vision and all observers were non-experts in image compression. The viewing distance was approximately 40 cm (i.e., a normal PC viewing distance) and the image pairs were viewed one at a time in a random order. The observer was free to view the images multiple times before making a decision, however a buzzer sounded after 20 s to indicate that they should make a decision. In addition, a blank mid-grey image was shown between each image (for 2 s) to prevent observers switching between the two images to find insignificant differences. Each image pair was viewed twice, giving  $(6 \times 4 \times 2)$  48 comparisons, which means that each observer took approximately 10 min to view all of the images. Images were viewed on a 12.1 in. Thin Film Transistor (TFT) Display in a darkened room (i.e., daylight with drawn curtains). The test images were displayed on a mid-grey background to a maximum size of  $410 \times 600$  pixels. Prior to the start of the visual trial all observers were given a short period of training on the usage the visual trial software and they were told to select they image they preferred assuming that it had been downloaded over the internet or wireless network.

## 5. Results

Table 1 shows the overall preferences, i.e., independent of (summed over) image and bit-rate, for standard JP2K and VA ROI JP2K coding. Table 1 also shows the standard errors associated with the preferences assuming a Gaussian approximation to the Binomial distribution. From Table 1 it can be seen that standard JP2K is preferred over VA ROI JP2K approximately 65% of the time. This shows that standard JP2K produces good quality images over a wide range of bit-rates and indicates that VA ROI coding may not be suitable as a general-purpose image coding technique.

Table 1  
Overall preferences (independent of image and bit-rate)

Method	Preferences	Standard error
JP2K	311	$\pm 12.3$
VA ROI JP2K	169	$\pm 12.3$

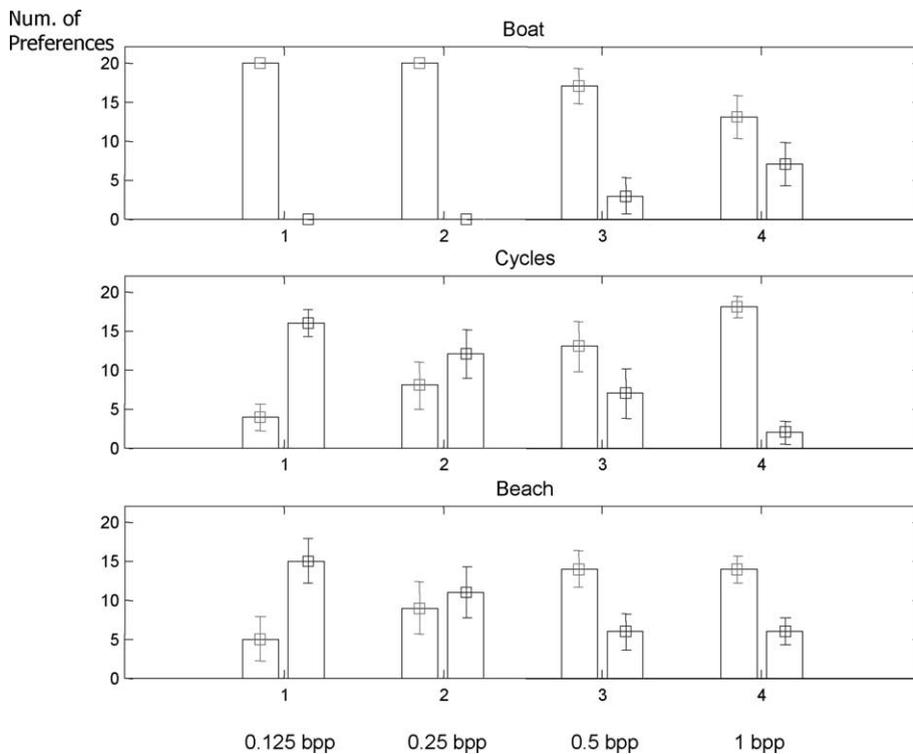


Fig. 14. JP2K (left) and VA ROI JP2K (right) preferences on the boat, cycles, and beach images at 0.125 (1), 0.25 (2), 0.5 (3), and 1 bpp (4).

Therefore, we will have to examine the results in more detail to identify the conditions to which the VA ROI JP2K coder is best suited.

Figs. 14 and 15 show the individual preferences for each image at each bit rate. Standard JP2K is shown with red standard error bars (on the left) whilst VA ROI JP2K is shown with blue standard error bars (on the right). From Figs. 14 and 15 it can be seen that there is a large variation in preferences across each of the images in the test set. The second, and more important, trend that can be observed in Figs. 14 and 15 is an increase in preferences for VA ROI JP2K as the bit-rate decreases. This trend is perhaps better highlighted in the lower graph of Fig. 16, which shows preferences at each bit-rate, independent of image. At the lowest bit-rate tested (0.125 bpp) the preferences for VA ROI JP2K are 68, with a standard error of  $\pm 5.8$ , and 52 ( $\pm 5.8$ ) for standard JP2K. This indicates a clear preference for the proposed method at this bit-rate. Note, the slight decrease in preferences for standard JP2K at 1 bpp, compared to 0.5 bpp, is due to the two methods producing image that look increasingly similar as the bit-rate increases. Therefore, preferences between the two methods will tend to random (50/50) selection.

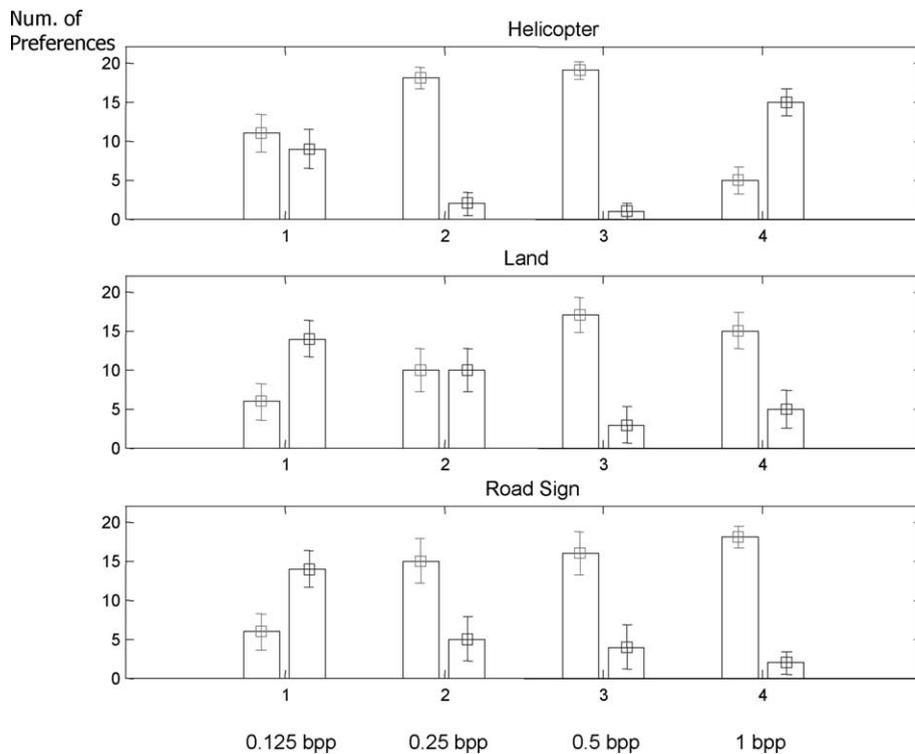


Fig. 15. JP2K (left) and VA ROI JP2K (right) preferences on the helicopter, land, and road sign images at 0.125 (1), 0.25 (2), 0.5 (3), and 1 bpp (4).

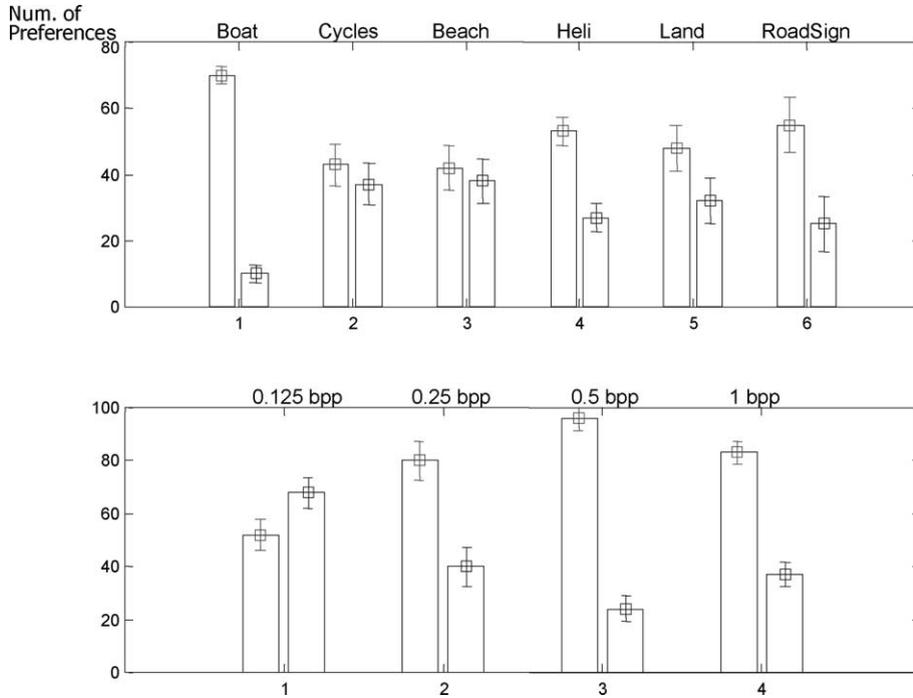


Fig. 16. JP2K (left) and VA ROI JP2K (right) preferences for each image (independent of bit-rate) and preferences at each bit-rate (independent of image).

## 6. Discussion

As illustrated in Figs. 14–16, there are two main sources of variation that can explain the differences in preferences: variation with image and variation with bit-rate (variation with observer being indicated by the standard errors in the results). The variation across the images in the test set shows that for an image that has an ROI and a background of little importance, such as the beach image, the ROI coding works well. However, for an image that has an ROI and also some visually important contextual details in the background, such as the boat image, the ROI coding works less well.

The increase in preferences for the VA ROI coding as bit-rate decreases, illustrated in Fig. 16, is the most significant and consistent effect observed in the visual trial (being apparent in 5 of the 6 images in the test set). As shown in Fig. 17 (for the cycles and road sign images) at low bit-rates ( $<0.25$  bpp), having the ROI encoded first in the bit-stream can significantly improve the visual quality of the ROI compared to standard JP2K. In addition, the background (non-ROI) areas are not of significantly poorer visual quality and are often of preferable quality as they contain less wavelet artefacts. At the low bit-rates the background regions tend to contain only coefficients from the lowest level of the wavelet transform rather than



Fig. 17. JP2K (left) VA ROI JP2K (right) at 0.125 bpp on road sign and cycles.

sporadic coefficients from higher levels of the DWT (as in standard JP2K). This results in a background that is uniformly blurred, which is often preferable to a less blurred background, which also has wavelet artefacts.

At the higher bit-rates ( $>0.25$  bpp), as shown in Fig. 18, the ROI is often not of significantly better visual quality than standard JP2K. This combined with the fact that the background areas are often more blurred and pixelated than standard JP2K results in lower preferences. This effect should come as no surprise as once the ROI is coded to a visually acceptable level it takes a significant number of bit refinements (the least significant bits) to get a visible improvement in image quality. This effect is also compounded by the limited resolution and quality of the original images. However, this effect would have applied equally to both compression techniques and so cannot fully explain the observed results. In addition, because small code-blocks were used at all bit-rates for the VA ROI coding ( $16 \times 16$  compared to  $64 \times 64$ ) there is a reduced compression efficiency which is particularly apparent at the higher bit-rates (Bradley and Stentiford, 2002).

Another important reason for the reduction in preferences at bit-rates  $>0.25$  bpp is due to the inherently *uneven* image quality in the majority of ROI coded images. This results in images that do not appear *natural* as the ROI is in sharp focus, whilst the background is blurred. A more gradual change in image quality between ROI and background would, however, increase the size of the ROI having a negative impact on ROI coding efficiency.

There is also an anecdotal explanation for the reduction in preferences at bit-rates  $>0.25$  bpp by considering the rule of thumb that to observe a significant increase in the visual quality of an image you have to (approximately) double the bit-rate. This means that the ROI has to be coded to twice the bit-rate to observe a significant improvement in perceived visual quality. Therefore, if we assume the ROI is  $1/4$  of the



Fig. 18. JP2K (left) VA ROI JP2K (right) at 1 bpp on road sign and cycles.

image area, then to code an image to the same (target) bit-rate as standard JP2K, the background must be coded to half the target rate so as to allow the ROI to be coded at twice the target rate. For example, if the target bit-rate is 0.25 bpp, then we can either code to this bit-rate using standard JP2K, or code the ROI to 0.5 bpp and the background to 0.125 bpp using JP2K ROI coding (giving an overall bit rate of 0.22 bpp). Therefore, at a given bit-rate, for the ROI to be coded to appear significantly better than standard JP2K the *background must be coded so as to appear significantly worse*. The only exception to this rule observed in the visual trial was at the lowest bit-rate (0.125 bpp) when compression artefacts also play an important role. Therefore, at bit-rates greater than 0.25 bpp the increase (if any) in quality of the ROI does not compensate for the decrease in background quality when observers judge *overall* image quality.

## 7. Conclusions

The proposed VA ROI algorithm has been shown to find the primary region(s) of interest in each of the images in the test set. The algorithm produces an image containing up to two ROI ellipses suitable for maxshift coding in JP2K. If no ROI suitable for JP2K coding is found then it defaults to conventional coding. The proposed algorithm is conceptually simple and, like the VA algorithm upon which it is based, requires no a priori or high-level knowledge. Results from the visual trial indicate an overall preference for standard JP2K independent of image and bit-rate. However, the proposed VA ROI coding method was preferred at the lowest bit-rate tested (0.125 bpp). This indicates that, when observers judge *overall* image quality, it is only at this bit-rate that the visible increase in quality of the ROI *more than* compensates for the decrease in quality of the background.

## Acknowledgments

This work was undertaken whilst Dr. Bradley was a visiting research fellow at BTextact Technologies, Adastral Park. All experiments were performed with JJ2000, version 4.1 (the current version is available from <http://jj2000.epfl.ch>). The authors would also like to thank the anonymous referees for their comments on the earlier versions of this paper.

## References

- Bradley, A.P., Stentiford, F.W.M., 2002. JPEG 2000 and Region of Interest Coding, *Digital Image Computing Techniques and Applications (DICTA)*, Melbourne, Australia, pp. 303–308.
- Christopoulos, C., Skodras, A., Ebrahimi, T., 2000. The JPEG 2000 still image coding system: an overview. *IEEE Transactions on Consumer Electronics* 46 (4), 1103–1127.
- Daly, S., Mathews, K., Ribas-Corbera, J., 1998. Face-based visually-optimized image sequence coding. *International Conference on Image Processing (ICIP 98)* 3, 443–447.
- Eckert, M.P., Bradley, A.P., 1998. Perceptual models applied to still image compression. *Signal Processing* 70 (3), 177–200.
- Gonzalez, R.C., Woods, R.E., 1993. *Digital Image Processing*. Addison-Wesley, Reading, MA.
- Pennebaker, W.B., Mitchell, J.L., 1993. *JPEG: Still Image Compression Standard*. Van Nostrand Reinhold, New York.
- Privitera, C.M., Stark, L.W., 2000. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (9), 970–982.
- Santa-Cruz, D., Ebrahimi, T., Larsson, M., Askelof, J., Christopoulos, C.A., 1999. Region of Interest Coding in JPEG 2000 for Interactive Client/Server Applications, *IEEE 3rd Workshop on Multi-media Signal Processing (MMSP)*, pp. 389–394.
- Stentiford, F.W.M., 2001. An Estimator for Visual Attention Through Competitive Novelty with Application to Image Compression, *Picture Coding Symposium*, Seoul, Korea, pp. 101–104.
- Stentiford, F.W.M., May 2001. An Evolutionary Programming Approach to Simulation of Visual Attention, *Congress on Evolutionary Computation*, Seoul, Korea, pp. 851–858.
- Taubman, D.S., Marcellin, M.W., 2001. *JPEG 2000: Image Compression Fundamentals, Standards, and Practice*. Kluwer Academic Press, Dordrecht.
- The Mathworks Inc., 1999. *Matlab User Guide: Statistics Toolbox, Version 2*.