Elsevier Editorial System(tm) for Signal Processing: Image Communication

Manuscript Draft

Manuscript Number:

Title: Gaze Tracking for Region of Interest Coding in JPEG 2000

Article Type: Full Length Article (FLA)

Section/Category:

Keywords: eye tracking; image coding; importance map; JPEG 2000; region of interest (ROI)

Corresponding Author: Dr Anthony Nguyen,

Corresponding Author's Institution: Queensland University of Technology

First Author: Anthony Nguyen

Order of Authors: Anthony Nguyen; Vinod Chandran; Sridha Sridharan

Manuscript Region of Origin:

Abstract:

Abstract

Current image coding systems such as JPEG are far away from the capability of the human perceptual system in that it does not aim to maximise the reconstruction quality of image contents, especially at low bit-rates. Humans are often concerned with the interpretability of the image and thus improved reconstruction quality in image contents would allow improved recognition performance. This paper addresses this issue by incorporating characteristics of the human perceptual system into an image coding system by making use of the human visual attention spatial and temporal characteristics. This is achieved by using an eye-tracking device at the encoding end. Human visual attention mechanisms would direct the viewer's eye movements around the image to provide a sequence of fixations, which can be analysed, clustered and classified into regions of interest (ROI). These ROIs can be used by the JPEG 2000 image coding standard to encode and prioritise image content and improve recognition performance.

Gaze Tracking for Region of Interest Coding in JPEG 2000

Anthony Nguyen *, Vinod Chandran, Sridha Sridharan

Image and Video Research Laboratory, School of Engineering Systems, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia

Abstract

Current image coding systems such as JPEG are far away from the capability of the human perceptual system in that it does not aim to maximise the reconstruction quality of image contents, especially at low bit-rates. Humans are often concerned with the interpretability of the image and thus improved reconstruction quality in image contents would allow improved recognition performance. This paper addresses this issue by incorporating characteristics of the human perceptual system into an image coding system by making use of the human visual attention spatial and temporal characteristics. This is achieved by using an eye-tracking device at the encoding end. Human visual attention mechanisms would direct the viewer's eye movements around the image to provide a sequence of fixations, which can be analysed, clustered and classified into regions of interest (ROI). These ROIs can be used by the JPEG 2000 image coding standard to encode and prioritise image content and improve recognition performance.

Key words: Eye tracking, image coding, importance map, JPEG 2000, region of interest (ROI)

1 Introduction

The ultimate end users of many imaging systems and processes are humans who are often concerned with the interpretability of the image to achieve maximum content recognition. For example, integral to people requiring the progressive download and display of images over low bandwidth channels such as the Internet and World Wide Web are an image coding system's ability

* Corresponding Author.
Email address: a.nguyen@ieee.org (Anthony Nguyen).

(a) (b)

Fig. 1. Reconstructed 'Lena' image at 0.125 bits per pixel (bpp) for (a) JPEG 2000 and (b) JPEG 2000 based region of interest (ROI) prioritisation method.

to encode objects or regions of interest with higher fidelity and priority such that its reconstruction will enhance or maximise image content recognition. People often have to either wait for a lengthy period of time for particular regions of an image to reconstruct to an appropriate level or wait till the whole image is fully downloaded before it can be interpreted. This limits user satisfaction and hence productivity. The functional interpretability requirement for compressed imagery is also important in other low bandwidth applications such as the delivery of imagery over Mobile telephony and Defence network infrastructures.

Traditional image coding systems, such as JPEG and JPEG 2000 [1], are based on encoding image data using an objective measure of overall image distortion, which treats all impairments as equally important and this may not correlate well with image quality or interpretability [2–5]. An example of a JPEG 2000 encoded 'Lena' image is shown in Fig. 1(a). Notice that the degradation in image quality at this low bit-rate is uniform over the entire field of the image and does not permit regions of visual interest, or more commonly referred to as regions of interest (ROI), to be reconstructed with higher quality for improved interpretability.

On the other hand, when an image coding system aims to compress and prioritise image contents or ROIs, the quality in these regions would be greatly improved to allow faster image content recognition. An example of this is shown in Fig. 1(b), where a JPEG 2000 based ROI coding method was used to prioritise the face region ahead of its surroundings to considerably improve the interpretability of Lena's face. In such a case, the ROI was defined to be the face, which was considered important for interpretability. This ROI can then be used to prioritise the encoding of the image with a certain level of

priority to the ROI. The level of priority can be defined using a quantitative measure of the region's 'importance'. Several JPEG 2000 based ROI coding methods have been standardised (e.g. [1, 6]) and proposed (e.g. [7–13]) to encode and prioritise pre-defined ROIs. An overview of the various algorithms and their advantages and limitations can be found in [5, 12].

Although algorithms exist for encoding ROIs with higher priority, there is still a need to address the question of how to define or select ROIs for such coding. A number of automated techniques using image processing algorithms have been proposed to predict ROIs or regions of high information content in images and video [14–21]. Importance scores (or weights of relative importance) are also commonly assigned to ROIs to generate an importance map. In such a map, a quantitative measure of the degree of importance is assigned to each pixel location in the image space to represent the relative importance of different regions in an image. Some of these approaches have been applied to visually lossless coding [22–24] and ROI coding [25, 26]. Importance maps have also been applied to image coding simulations to improve the interpretability in regions of high importance [21, 27]. The loss of detail during compression is minimised in terms of human perception so that when decoded, degradations in the image does not appear to be visually disturbing in ROIs (i.e. visually salient areas).

Despite these research efforts, there are a few challenges in automatically generating importance maps or models of visual attention, namely: (1) it is difficult to prepare a list of features that will characterise all ROIs, (2) it is a tough problem in weighting and thresholding importance of features without prior knowledge of the image, and (3) adding more and more features can lead to diminishing returns and degraded performance. Therefore, manual involvement in the selection of ROIs can help a lot.

A number of researchers have investigated the concept of manually identifying ROIs by taking advantage of the technology advances that have taken place in eye movement tracking [28–33]. These advances have resulted in cheap, fast, accurate and user friendly gaze tracking systems and thus are no longer intrusive or require cumbersome headgear to be worn. The challenge in these gaze-based ROI identification systems is to be able to appropriately process fast and inherently noisy eye movements.

In particular, Nguyen *et al.* [32, 33] have analysed real-time gaze information for determining ROIs and subsequently applied it to spatially selective image coding using JPEG 2000. In such a system, the image author (or creator of the image code-stream) can use his/her gaze to selectively encode the image. Although, the system proposed in its current form was used for off-line gaze analysis and image coding, the system has potential in real-time image coding applications whereby the passive use of a person's gaze can automatically

and appropriately encode images as desired by the image author. Smart applications such as medical imaging and photographic surveillance, which may require the selective encoding and progressive transmission of regions or targets of interest for expert analysis at remote sites will greatly benefit from advances in gaze-based image coding. In addition, a web-based application may include an image author editing and placing compressed images of a photo album online for remote access by interested parties. In such an application, large photographs can be progressively transmitted to the client with ROIs reconstructing faster than unimportant regions. Hence, images need not be fully decoded for the user to view contents and/or interpret whether he/she wants to proceed further with the image download or not.

This paper will present a review of the gaze-based image coding system proposed in [32, 33] with the aid of an extended range of example images and figures illustrating the ROI identification process. Further investigations into the system's ROI coding performance in objective quality experiments will also be presented.

2 Gaze Tracking for Image Compression

The motivation for the use of human input to an importance mapping algorithm are its visual attention mechanisms and search strategies when presented with a visual stimulus. The brain effortlessly builds up the knowledge of the scene and interprets the scene contents through a series of high resolution ROIs at the point of fixation and a low resolution background contained in the periphery. The fixations can be monitored using eye tracking devices to record where and when a person's gaze is directed to establish the fixation-saccade path (or gaze pattern).

A methodology is proposed to make use of a viewer's gaze pattern to automatically generate a ROI based importance map. The aim of the approach is to track a viewer's gaze and develop an importance mapping algorithm that can analyse and group gaze locations into a representative number of clusters to represent ROIs. An importance measure for these regions can be derived using known properties and characteristics of the gaze pattern. The identification of ROIs from gaze information was applied to a class of imagery comprising of scenes consisting of a limited set of objects of interest.

An example of a JPEG 2000 based image compression system, called *Gaze-J2K*, which makes use of an interaction device provided by an eye tracker is shown in Fig. 2. Here, the gaze point collection stage records information on the location and sequence of fixations that were followed by the user. The structure of the spatial and temporal characteristics of the gaze pattern can

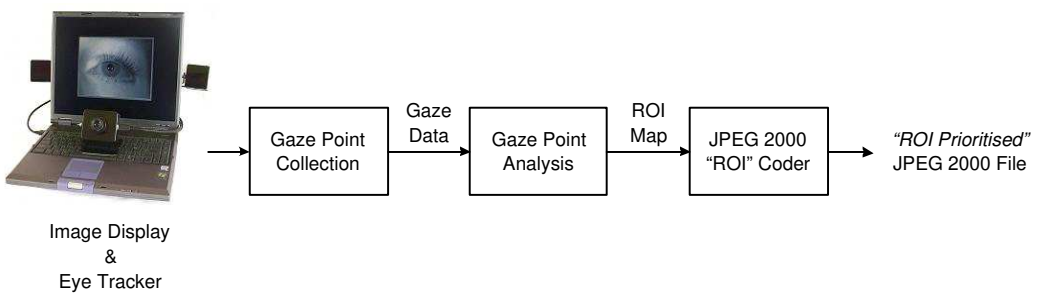


Fig. 2. Gaze-J2K system block diagram.

then be used as parameters and analysed to generate a map of ROIs and its associated measure of importance. These ROIs and importance scores define an importance map and can be input to a JPEG 2000 ROI coder to encode and prioritise the image code-stream according to the importance map specification. A user at the decoding end can subsequently reconstruct the image progressively, say over a low bandwidth network, with the ROIs reconstructing faster than other regions in the image. Each stage of operation is described in further detail in the following subsections.

2.1 Gaze Point Collection

The goal of gaze or eye tracking is to determine the gaze point on the field of view where a user is looking. The device used to record eye movements was an EyeTech video-based corneal reflection eye tracker [34]. The eye tracker consists of two infrared lights (mounted on both sides of the computer monitor) and an image sensing camera (mounted in front of the monitor screen). The method of operation relies on illuminating the user's eye with the infrared lights and through a process of focusing and tracking the position of the infrared light reflections from the eye and the centre of the pupil on the image captured by the camera, the gaze point can be determined. The accuracy of the eye-tracker, as reported by the manufacturer, is within ± 1.0 degree. The gaze-tracker can process 15 to 30 video frames per second (fps) and records the position and time of gaze. The higher the frame update rate, the more fixations that can be recorded by the eye tracker. The system was setup so that gaze data collection can be conducted on an image and screen resolution of 1024×768 pixels. The eye tracker update rate was set to 15 fps.

The collection of gaze data was obtained from 13 viewers, of which 11 were naive to the purpose of the study. The viewers ranged from technical to non-technical backgrounds and with normal or corrected-to-normal vision. Six colour images ('boat', 'cow', 'horse', 'paddock', 'rockclimb', and 'yacht'), each with at least one primary object of interest clustered in a scenic background, as shown in Fig. 3, were displayed on the computer monitor, and each viewer

was instructed to locate and examine the objects in the image. For each image, the task was repeated three times for a duration of 15 seconds each. A shorter duration for gaze recording is possible for gaze analysis but is in general image dependent. Fifteen seconds was determined to be a sufficient amount time to allow viewers to examine the primary object(s) of interest while also having time to scan for other possible objects in the image. It was found that if the duration of gaze recording extended beyond 15 seconds, then viewer's would tend to be distracted from the task at hand and produce less meaningful gaze information.

A selection of six recorded gaze patterns, one from each test image, superimposed on the original image is shown in Fig. 4. The plots do not show any information about the sequence that the gaze points were viewed in. The gaze patterns will be used as examples to show the results for the remainder of the Gaze-J2K process. The figure shows some of the variations in gaze patterns that are possible when recording a person's gaze. Notice that fixation points may be sparsely directed to different regions in an image, but generally with more fixations being directed to objects (or ROIs) in the image. This premise can be used to efficiently develop an algorithm to determine ROIs in an image.

A culling process was performed to discard gaze data sets that were found to be unsuitable for further processing. A total of 229 out of the 234 gaze patterns were retained for the testing of subsequent stages of Gaze-J2K. The discarded gaze data sets were mainly due to the eye tracker failing to locate the viewer's location of fixation for the vast majority of the viewing duration. Drifts that are naturally exhibited with a viewer's fixation and differences in the viewer's interpretation of the eye-tracking task can also result in a diverse range of gaze patterns. These differences can pose some challenges with the gaze point analysis and image compression stage. These issues will be referred to in their respective sections.

2.2 Gaze Point Analysis

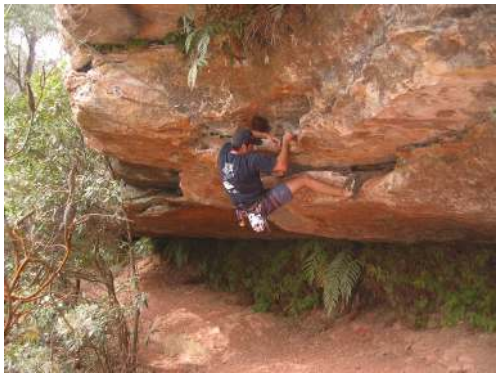
The gaze point analysis procedure reduces the spatial characteristics of the gaze pattern to a limited subset of clusters that would represent ROI candidates. The choice of clustering technique is influenced by a number of factors such as whether the probability densities of the data are known or can be modelled, and the size of the data set. Since the number of gaze location points are limited and its spatial distribution is unknown, an unsupervised clustering technique, such as K-means, can be used. In addition to the clustering procedure, a means to determine the importance of the ROI candidates will also be presented. The following details the development of the ROI clustering and importance mapping stages.

(a) ‘beach’

(b) ‘cow’

(c) ‘horse’

(d) ‘paddock’



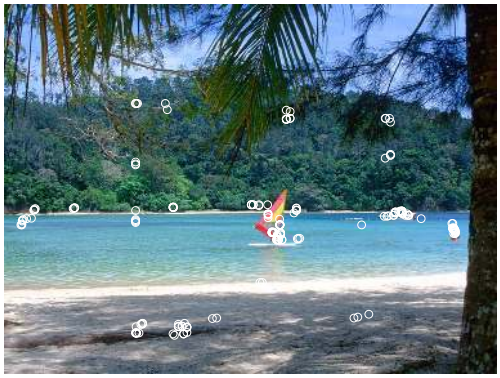
(e) ‘rockclimb’

(f) ‘yacht’

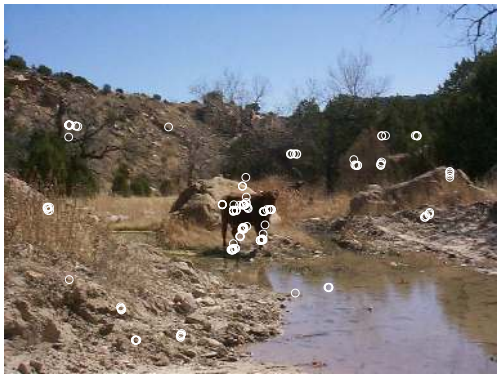
Fig. 3. Gaze-J2K colour test images ($1024 \times 768, 24$ bpp).

2.2.1 ROI Clustering

ROI clustering involves the partitioning of gaze points into mutually exclusive clusters such that the loci of the points belonging to the clusters represent the ROI candidates for the particular gaze pattern. Here, a K-means clustering method is used to iteratively assign data to one of K clusters using the distance from the means of these clusters. The result is a set of clusters that are as compact and well-separated as possible. The K initial values for the cluster



(a) ‘beach’

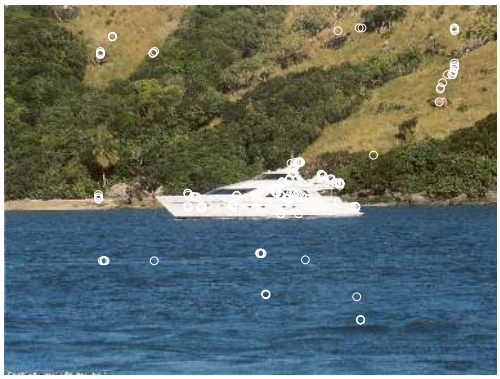


(b) ‘cow’

(c) ‘horse’

(d) ‘paddock’

(e) ‘rockclimb’

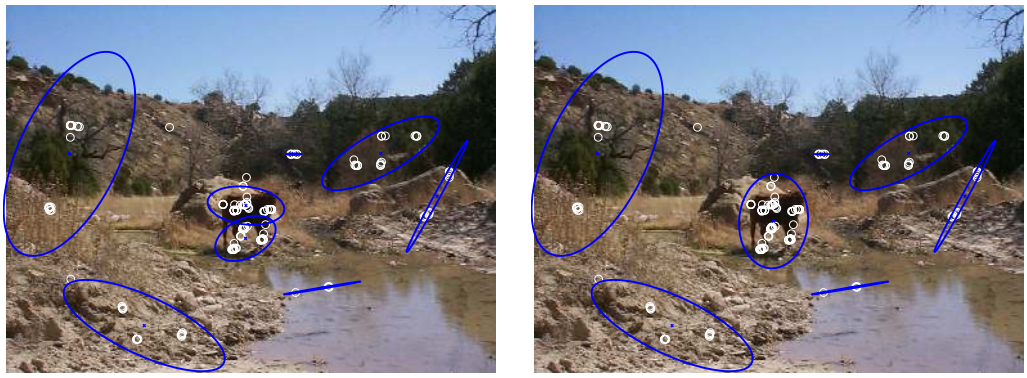


(f) ‘yacht’

Fig. 4. Example gaze pattern data superimposed on the original Gaze-J2K test images. The examples show the variety of gaze patterns that are possible when recording a person’s gaze.

means were chosen randomly from the data set. These initial values can cause K-means to converge to a local minima, where the total sum of distances are a minimum, from which a better solution may exist. To avoid this, K-means was repeated a number of times and if different local minima exists then the case with the lowest total sum of distances, over all repetitions, would be returned.

The value K was automatically determined by increasing the number of clus-



(a) Before Merging. (b) After Merging.

Fig. 5. Example of the merging of ROI cluster candidates for the 'cow' gaze pattern shown in Fig. 4(b); (a) ROI cluster candidates after the K-means clustering procedure, and (b) ROI cluster candidates after the merging of K-means clusters that are close in proximity. Note that the two clusters representing the cow has been merged into one.

ters until K-means found a K which gives maximum cluster separation. A silhouette score can be used to measure how close each point in one cluster is to points in neighbouring clusters [35]. The average of the silhouette scores for each K can be used as a quantitative measure to compare different K's. That is, the larger the silhouette score, the larger the cluster separation. As such, K-means was repeatedly performed by increasing K by 1 at each stage until the silhouette score for the grouping of data for $K+1$ is less than that for K. In such a case, K would give a maximum silhouette score and the data vectors and mean of the clusters that correspond to K would represent the 'best' grouping of the data.

In some cases, a number of clusters may be close together and a procedure is required to merge these clusters. The rule used to merge was if any two cluster means fall within a distance threshold of 10% of the average of the image dimensions, then the two clusters would merge into one. The merging process would often merge multiple clusters belonging to the same object into a single cluster such as that shown in Fig. 5. The number of clusters, K, and the geometric mean of the merged clusters are updated as clusters are merged.

The cluster means and covariances of the data vectors that were assigned to the clusters were used to generate ellipses to represent the loci of ROI candidates. The major and minor radial components of the ellipses were chosen to be 2.58 standard deviations in each direction. In such a case, if the cluster's spatial distribution was Gaussian, then this will represent approximately 99% of data points belonging to the cluster. The total ROI size (area bounded by all the ellipses) was also restricted to less than 25% of the image space. This is to ensure that during the encoding process, the reconstructed quality of ROIs more than compensates for the overhead in encoding the ROI [26, 36]. If the

ROI area did not satisfy this condition, then the process was repeated with K incremented by 1.

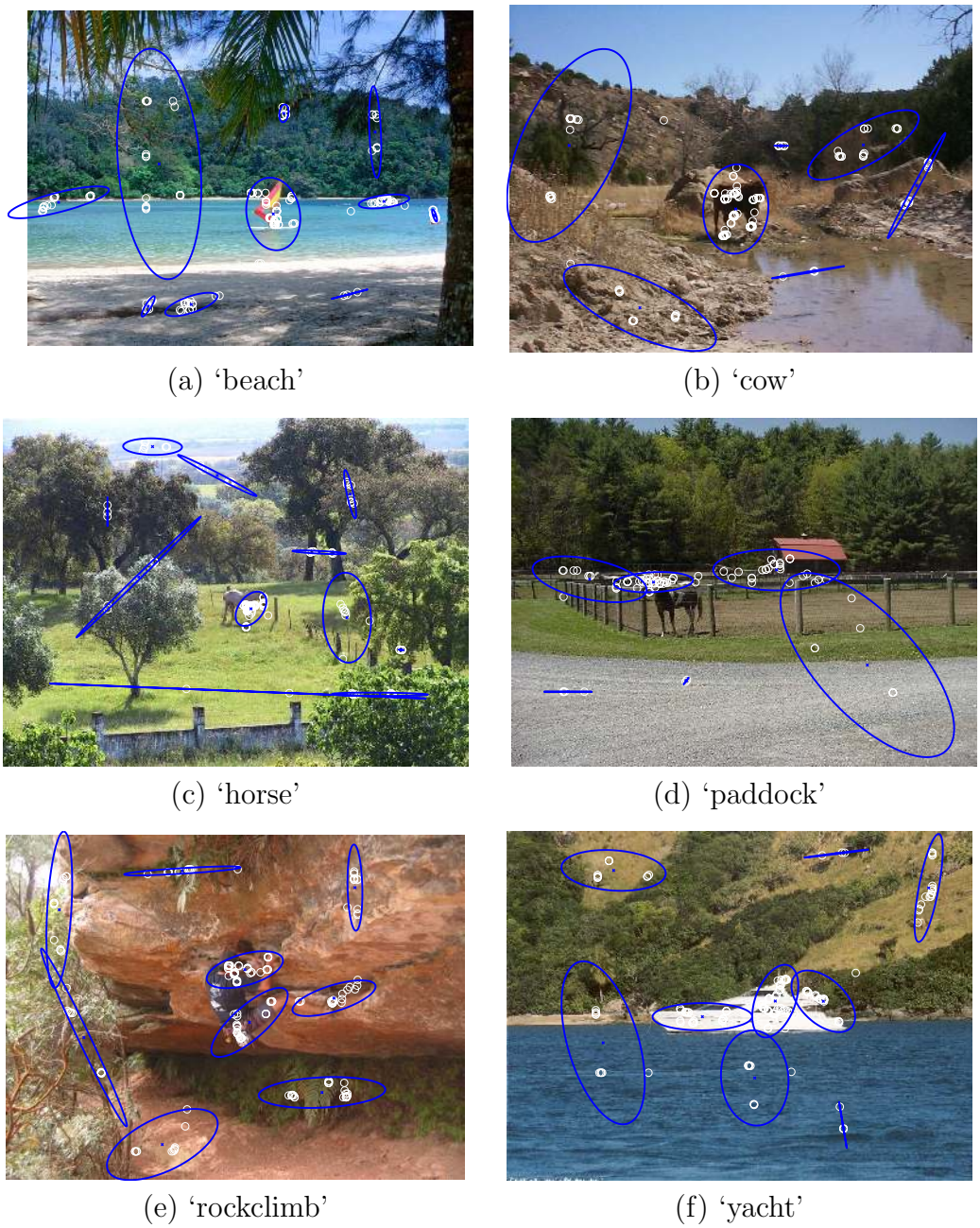
The output of the clustering procedure, representing the ROI candidates, for the gaze patterns in Fig. 4 are shown in Fig. 6. Note that a large number of clusters can result from the process, from which only some are truly representative of objects of interest in an image. Also note that the loci of the clusters are dependent on the gaze pattern. If the ROI in an image is large, such as those in the 'rockclimb' and 'yacht' image, then multiple clusters may result for the single object. There were also cases where viewers only fixated on a particular region of an object, such as the head of the horse, which meant that the clustering procedure will not produce a ROI locus that would encompass the whole object. Other problems that may exist is that some gaze points not belonging to the ROI may be included in the ROI cluster simply because the gaze point was closer to the ROI cluster than any other clusters. These problems can be overcome either by improving the clustering algorithm to reduce the sensitivity of the 'outlier' gaze points and/or by having viewers more experience with the eye tracker hardware and be more aware of the purpose of the task required for the application at hand.

2.2.2 ROI Mapping

Given that the ROI clustering procedure outputs K candidate ROI clusters, a ROI mapping procedure is required to assign an importance measure or score to each of the clusters. This importance score is the degree of importance of the cluster relative to other clusters. Clusters with a high importance score represents regions of high importance, which should be retained and prioritised by an encoder with higher priority than clusters with a low importance score, which should be removed and prioritised along with the image background. The importance score can also be interpreted as the probability of a cluster being an actual ROI.

It is conjectured that the fixation-saccade sequence provided by the gaze patterns would reveal underlying visual attention processes that can be used to develop an importance metric. The number of fixations, x_k, belonging to cluster k is an obvious initial choice to measure a cluster's importance. In the following discussion, three importance metrics derived from the first and second order statistic of x_k is proposed to analyse the ROI cluster candidates. A ROI threshold, T_{ROI}, can be subsequently applied to the importance measures to determine whether or not a ROI cluster candidate is classified as a ROI. The three importance metrics are formulated as follows:

(1) *First Order* metric, $I^{'}(k)$: This metric represents a first order statistical measure of the number of gaze points, x_k, that belong to a given cluster,



(a) ‘beach’ (b) ‘cow’

(c) ‘horse’ (d) ‘paddock’

(e) ‘rockclimb’ (f) ‘yacht’

Fig. 6. ROI cluster candidates for the gaze patterns shown in Fig. 4. The clustering procedure attempts to fit the gaze patterns into well-separated mutually exclusive clusters. These clusters are ROI candidates for the particular gaze patterns and are used for importance analysis to determine ROIs.

k, and is analogous to the duration of gaze within the cluster region, since uniform gaze sampling was recorded. $I^{'}(k)$ is given by $x_k/(K \cdot \overline{x})$ where K is the total number of clusters and $\overline{x}$ is the mean number of gaze points per cluster. The importance mapping procedure effectively maps x_k to a range 0 to 1 and thus represents a cluster’s relative importance. Note

that the importance for each cluster would be $1/K$ when all x_k have the same number of gaze points.

(2) *Second Order* metric, $I^{''}(k)$: This metric represents the second order statistical measure of the number of gaze points, x_k, that belong to a given cluster, k. This measure has the equivalence of squaring the number of cluster gaze points and thus emphasises clusters with a large x_k and penalises those with a small x_k. $I^{''}(k)$ is given by $x_k^2/(K \cdot \overline{x^2})$ where K is the total number of clusters and $\overline{x^2}$ is the mean of the squared number of gaze points per cluster. This metric is conjectured to lead to an enhanced performance over the *First Order* metric since the distribution of the number of gaze points within a cluster is taken into account through the use of the second order statistic. Again note that $I^{''}(k)$ is in a range 0 to 1 and represents a cluster's relative importance.

(3) *Weighted Second Order* metric, $I_w^{''}(k)$: This metric represents a weighted second order statistical measure of the number of gaze points, x_k, that belong to a given cluster, k. It is hypothesised that the number of visits (or revisits) to a given cluster during the course of viewing would provide additional information to determine the cluster's importance. In general, primary objects of interest tend to be visited a number of times during the viewing duration and thus may be reasonably modelled as being proportional to a cluster's importance. The percentage of clusters for a given number of visits is shown in Fig. 7. $I_w^{''}(k)$, is again normalised to the range 0 to 1 and is given by $(w_k \cdot x_k^2)/(K \cdot \overline{wx^2})$ where K is the total number of clusters, w_k is the weight representing the number of visits to cluster k, and $\overline{wx^2}$ is the mean of the weighted square of the number of gaze points per cluster. This metric, however, may not hold or accurately model the importance of clusters with a large number of visits, since it was found that these cases were either image or viewer dependent. To stop these clusters from having a dominant effect on the cluster importance, w_k was capped at 3 (i.e. $w_k = \min(w_k, 3)$); This value was chosen because clusters with a number of visits greater than 3 became less frequent (i.e. on average < 1 cluster per gaze pattern for each integer number of visit greater than 3) and thus would not be statistically reliable.

An illustrated example of the importance mapping process is shown in Fig. 8 for the 'rockclimb' ROI cluster candidates shown in Fig. 6(e). Fig. 8(a) shows the ROI cluster candidates with arbitrary cluster labels ($k = 1 \ldots K$) assigned to each cluster. The corresponding number of cluster gaze points, x_k, and the number of cluster visits, w_k, are tabulated in Fig. 8(c). It can be observed that more fixations were directed to the primary object of interest (i.e. the rock climber), while less were fixated to other parts of the image. In addition, the three clusters belonging to the rock climber were revisit during the duration of the gaze recording. From these two parameters, the importance score for the three importance metrics ($I^{'}(k)$, $I^{''}(k)$, and $I_w^{''}(k)$) were computed and are tabulated in Fig. 8(d). The cluster importance scores that are in bold font

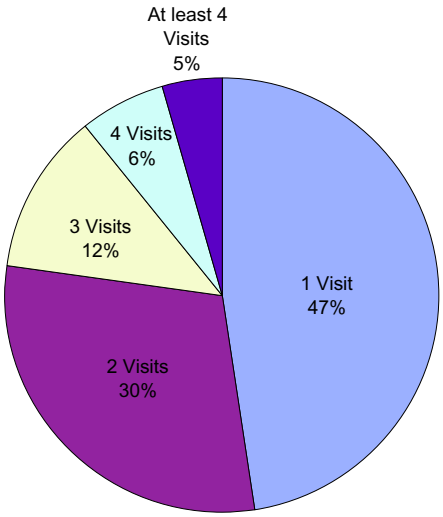


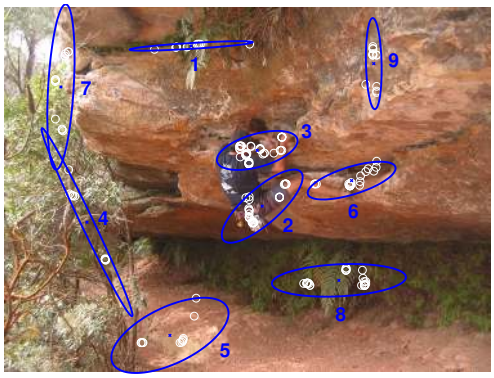
Fig. 7. Proportion of clusters with a given number of cluster visits. Revisits to a cluster is conjectured to provide additional information that can further emphasise a cluster's importance.

represents those clusters retained as an ROI (i.e. importance score > ROI threshold, T_{ROI}, where T_{ROI} was set to 0.15). Note that for the particular 'rockclimb' gaze pattern, the *First* and *Second Order* metric classified a non-ROI cluster (Cluster '8') as a ROI, while the *Weighted Second Order* metric was able to use the number of cluster visit information to its advantage to de-emphasise the non-ROI cluster to only retain the two clusters belonging to the rock climber. The final importance map specification using the *Weighted Second Order* metric is shown in Fig. 8(b).

2.2.3 ROI Importance Map Performance Evaluation

The performance of each of the metric was evaluated in terms of ROI misses and ROI false alarms. ROI misses are those cases where the ROI mapping algorithm did not pick up a primary object of interest in an image as an ROI, while a ROI false alarm is the case where the algorithm considered a cluster as a ROI when it contains no object of interest. Because of the varied range in objects fixated by a viewer and the number of objects that may exist in an image, an 'intuitive' *rule-based ROI definition* was formulated to define some ground truth ROIs. The following rule-based definitions were used:

- 'beach' image - ROI must contain the windsurfer. The boat and people swimming on the right of the image were not considered as ROI false alarms.
- 'cow' image - ROI must contain the head and/or body of the cow.
- 'horse' image - ROI must contain the head and/or body of the horse.
- 'paddock' image - ROI must contain the head and/or body of the horse in the foreground. The horses in the background and the barn are not ROI



(a)

(b)

ROI Cluster Candidate, k	1	2	3	4	5	6	7	8	9	Total
Number of Gaze Points, x_k	11	30	43	11	11	23	12	29	13	183
Number of visits, w_k	1	2	2	1	1	2	1	1	1	12

(c)

ROI Cluster Candidate, k	1	2	3	4	5	6	7	8	9
First Order, $I^{'}(k)$	0.06	**0.16**	**0.23**	0.06	0.06	0.13	0.07	**0.16**	0.07
Second Order, $I^{''}(k)$	0.03	**0.19**	**0.39**	0.03	0.03	0.11	0.03	**0.18**	0.04
Weighted Second Order, $I_w^{''}(k)$	0.01	**0.22**	**0.46**	0.01	0.01	0.13	0.02	0.10	0.02

† ROI cluster candidates in bold are above the ROI threshold, T_{ROI}, of 0.15 and are included in the final importance map for ROI prioritisation.

(d)

Fig. 8. An illustrated example of the Gaze-J2K importance mapping process for the 'rockclimb' ROI cluster candidates in Fig. 6(e). (a) Output of K-means clustering process, (b) Final importance map for the gaze pattern using the *Weighted Second Order*, $I_w^{''}(k)$, metric. Table (c) and (d) shows the number of cluster gaze points, x_k, and visits, w_k, to cluster k, and the importance scores for the *First Order* ($I^{'}(k)$), *Second Order* ($I^{''}(k)$), and *Weighted Second Order* ($I_w^{''}(k)$) ROI importance mapping metrics†. Note that for the particular gaze pattern, the *First* and *Second Order* metrics classified a non-ROI cluster (Cluster '8') as a ROI.

false alarms.

- 'rockclimb' image - ROI must contain the upper body of the rock climber. The rock climber's lower body was not considered to be a ROI false alarm.
- 'yacht' image - ROI must contain at least the centre of the yacht. Other parts of the yacht were not ROI false alarms.

Table 1 shows the performance in terms of ROI misses and ROI false alarms for the three ROI importance mapping methods using the above ROI definitions. The *First Order* metric contained 20 ROI misses and 48 ROI false alarms. The results indicate that, for the given importance score threshold, the 'excess' or additional number of clusters retained as ROIs are mainly contributed by ROI false alarms. However, it is noteworthy that increasing the importance score

Table 1
ROI misses and false alarm results for three ROI importance mapping metrics†.

Importance Metric	Number of ROI Misses	Number of ROI False Alarms
First Order, $I^{'}(k)$	20 (8.7%)	48 (21.0%)
Second Order, $I^{''}(k)$	13 (5.7%)	41 (17.9%)
Weighted Second Order, $I_w^{''}(k)$	7 (3.1%)	45 (19.7%)

† Values in parentheses are percentages of the total number of gaze data.

threshold would increase the number of ROI misses and reduce the ROI false alarms, while decreasing the threshold would significantly increase the number of ROI false alarms and only marginally reduce the number of ROI misses. The *Second Order* metric was found to be a more useful metric and resulted in a much improved ROI performance with 13 ROI misses and 41 ROI false alarms.

The *Weighted Second Order* metric can be to further improve the ROI performance by reducing the ROI misses to 7, while only marginally increasing the number of ROI false alarms compared to the *Second Order* metric. It can be concluded that the number of visits to a given cluster provides additional information that is important to the determination of ROIs from gaze patterns. This metric was hence selected for use in the gaze-based image coding system.

Fig. 9 shows the final importance maps for the gaze patterns and ROI cluster candidates in Figs. 4 and 6, respectively, using the *Weighted Second Order* metric. Notice that the importance mapping algorithm is quite robust in determining ROIs from the diversity of gaze patterns collected.

It should be noted that the majority of ROI misses and ROI false alarms are contributed by only a few viewers. Table 2 provides an indication of the distribution of ROI misses and ROI false alarms across the viewers for the *Weighted Second Order* metric. Notice that only two viewers contributed to the ROI misses, while a varying amount of ROI false alarms were contributed by different viewers. The large number of ROI false alarms indicates that viewers have their own viewing preferences and fixated on other regions in addition to the defined ROIs. As suggested earlier, the ROI performances can be improved substantially if viewers had more experience with the eye tracker and were more informed of the purpose of the task that was required. However, in the context of ROI prioritised image coding, it is assumed that a single compressed image bit-stream will be viewed by many viewers, and thus only the user of the encoder (i.e. the creator of the compressed image bit-stream) is required to be familiar with the eye-tracker to determine the priority of regions in the image of interest. In any case, if the underlying visual attention processes of viewers can be known, improved ROI identification performances (even over a generalised class of imagery) can be gained.

(a) ‘beach’

(b) ‘cow’

(c) ‘horse’

(d) ‘paddock’

(e) ‘rockclimb’

(f) ‘yacht’

Fig. 9. Importance map output from the gaze point analysis stage for the gaze patterns and ROI cluster candidates shown in Figs. 4 and 6, respectively. The importance scores shown in the figure are from the *Weighted Second Order* metric.

2.3 Application to Image Compression

The coding/decoding of images may be influenced by prioritising ROIs in the image code-stream using spatially selective coding methods such as the ROI coding methods standardised and proposed for JPEG 2000 (see Section 1). In particular, the *Importance Prioritised JPEG 2000 (IMP-J2K)* ROI coding

Table 2
ROI misses and false alarms contributed by viewers for the *Weighted Second Order*, $I''_w(k)$, ROI importance mapping metric. ROI misses and false alarms can be seen to be viewer dependent.

Viewer	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Number of ROI Misses	0	0	0	0	0	0	0	0	0	0	2	0	5	7
Number of ROI False Alarm	0	1	1	8	6	1	0	3	4	0	4	7	10	45

method [12] lends itself a number of useful features such as multiple ROI coding using an importance map input specification. These features are highly suitable for encoding and prioritising ROIs using the importance map generated from the Gaze-J2K process. ROIs are emphasised by weighting the Mean Square Error (MSE) distortion measure of code-blocks (blocks of wavelet coefficients) by the square of its importance score and its reconstruction is bounded by the extent of these code-blocks. Although reconstructing ROIs at a code-block level might be a disadvantage in some applications, the ROIs generated from the Gaze-J2K clustering and importance mapping process can use this property to its advantage, since ROIs may not fully encompass the objects in the image. Regions adjacent to ROI clusters will also be considered as ROIs to ensure that the whole object of interest will be prioritised when encoded.

The ROI cluster loci and importance measures as generated by the ROI mapping stage can be input to IMP-J2K for ROI encoding. Regions not belonging to the ROI are assigned an importance score of 0.01. Other ROI coding parameters used are as follows: the lowest resolution sub-bands of a four level wavelet decomposition were assigned an importance score equal to the ROI threshold of 0.15 to provide some degree of background context while prioritising the ROIs, and a code-block size of 32×32 and seven octave quality layered bit-rates spaced between 0.03125 ($= 2^{-5}$) and 2 bpp were used for the code-stream construction. For more information on the ROI coding parameters used, the reader is referred to [12].

Fig. 10 shows the differential PSNR results relative to JPEG 2000 (no ROI prioritisation case) for each of the importance map output shown in Fig. 9. PSNR (Peak Signal-to-Noise Ratio) is a common objective quality (or distortion) metric used to compare the quality of compressed images. Note the gain in ROI performance with IMP-J2K compared to the no ROI prioritisation case for all the test images. Furthermore, when there are 2 ROIs, the PSNR performance were in order of importance. It is also important to note that the value of a ROI's importance score and its corresponding PSNR value at various bit-rates are dependent on a number of factors including the ROI size and location with respect to the code-block boundaries [5, 12].

An example illustration of reconstructed images for the 'rockclimb' image

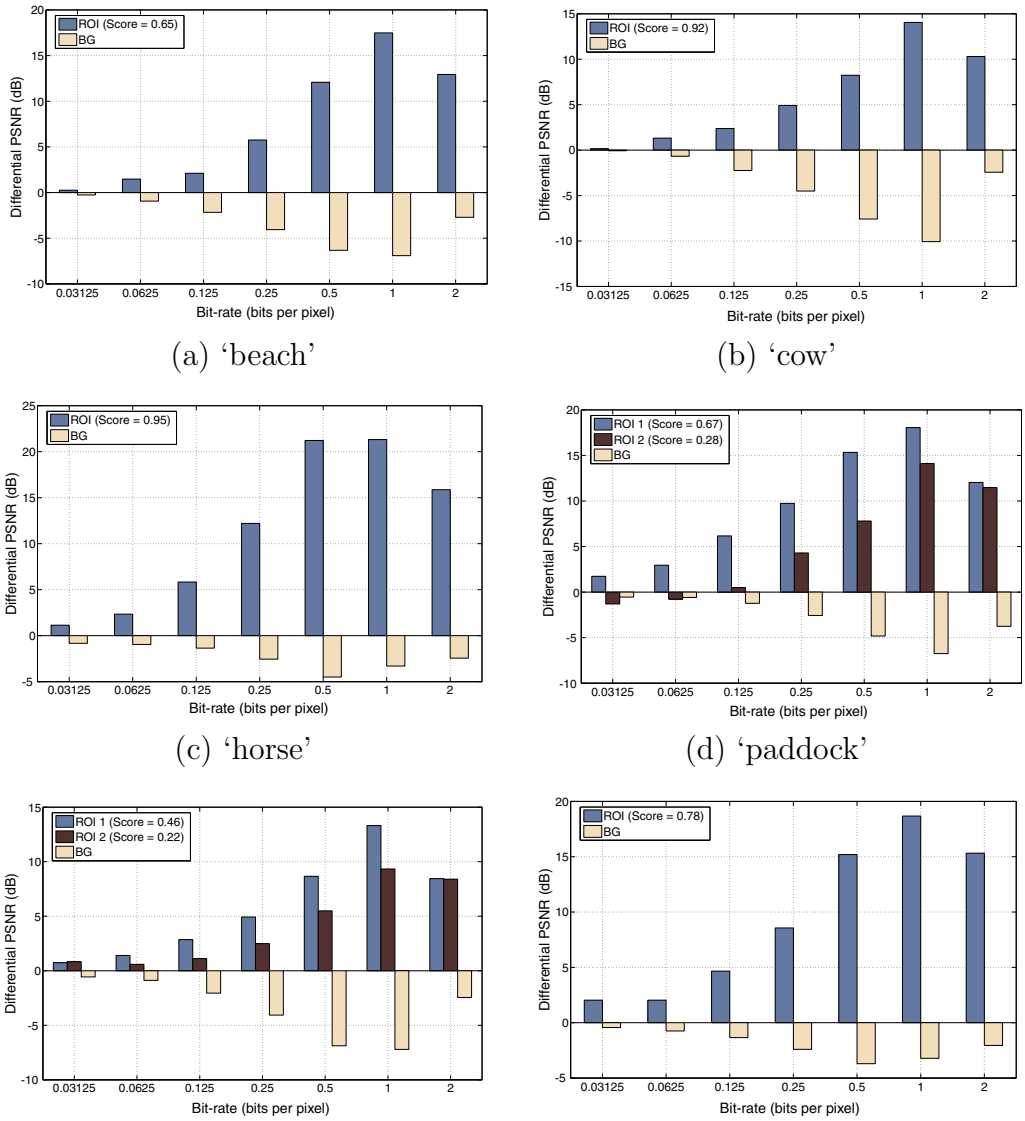
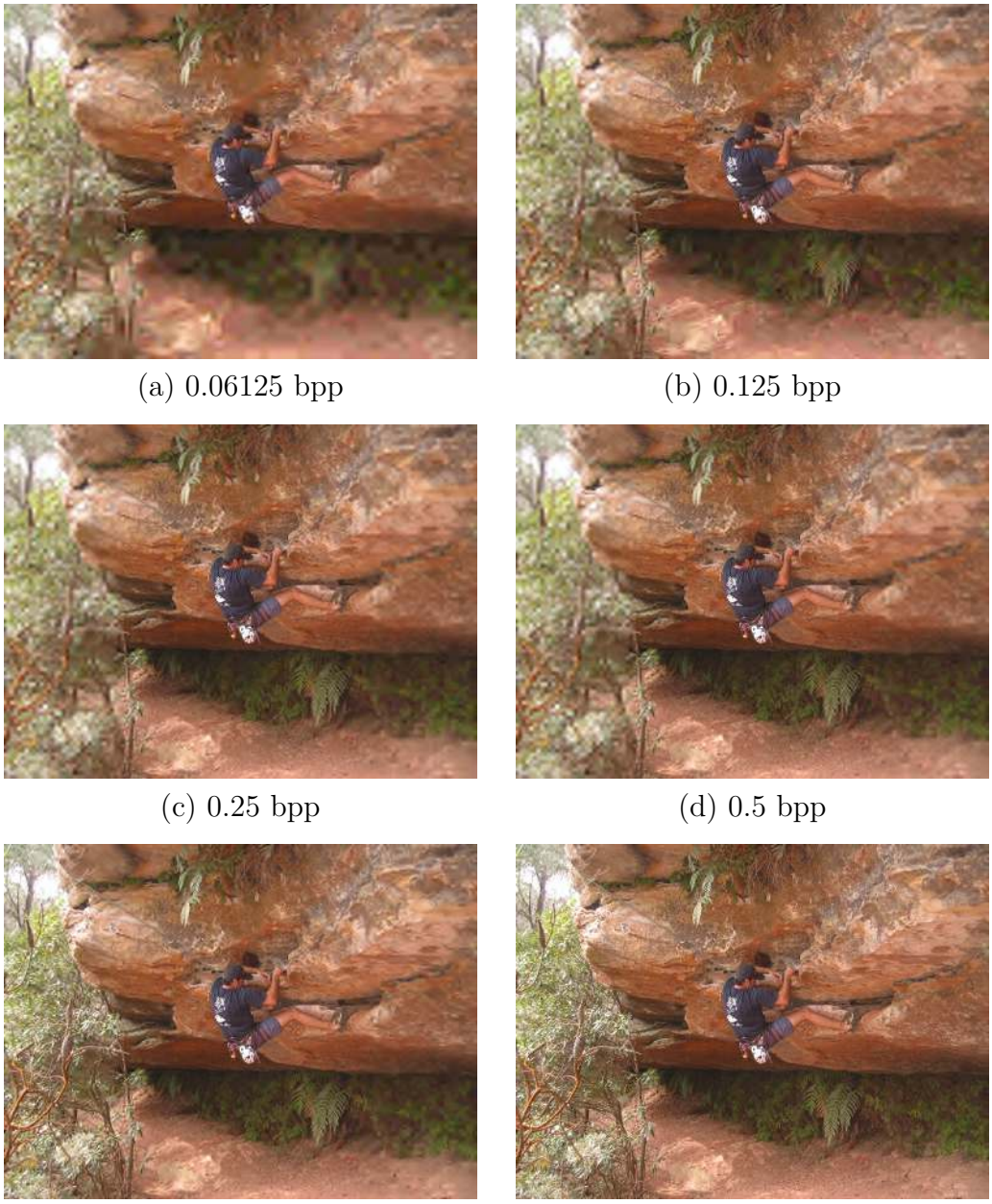


Fig. 10. Differential PSNR results relative to JPEG 2000 (no ROI prioritisation case) for each of the Gaze-J2K importance map output shown in Fig. 9. Note the gain in ROI performance with IMP-J2K compared to the no ROI prioritisation case for all the test images. Furthermore, when 2 ROIs were determined from the gaze patterns, the PSNR performance were in order of importance.

example in Fig. 9 as the ROI encoded image is progressively transmitted and received by a client is shown in Fig. 11. Note that the ROIs (i.e. rockclimber) are reconstructed with better quality and at a higher resolution than the rest of the image, which contains low resolution image information to provide the context for the ROIs. This characteristic is similar to the human visual attention system where fixated regions are processed with a higher resolution than regions in the periphery. The ROI prioritised coding strategy allows the user at the receiver to reconstruct the image as desired by the user at



(a) 0.06125 bpp (b) 0.125 bpp

(c) 0.25 bpp (d) 0.5 bpp

(e) 1.0 bpp (f) 2.0 bpp

Fig. 11. Progressive decoding of ROI prioritised code-stream at 0.0625, 0.125, 0.25, 0.5, 1.0 and 2.0 bits per pixel (bpp) for the 'rockclimb' image in Fig. 8. The ROIs improve most rapidly at reduced bit-rates, while the visually lossless (or near-lossless) reconstruction of the image as a whole is possible at higher bit-rates.

the encoder. The faster reconstruction of ROIs at lower bit-rates provides improved interpretability in ROIs.

Finally, Fig. 12 shows reconstructed images for all other Gaze-J2K test images at 0.125 bpp for the importance map specifications shown in Fig. 9. Again, all ROIs can be observed to exhibit higher reconstruction quality than the image

(a) ‘beach’

(b) ‘cow’

(c) ‘horse’

(d) ‘paddock’

(e) ‘rockclimb’

(f) ‘yacht’

Fig. 12. Decoded Gaze-J2K images at 0.125 bpp (i.e. compression ratio of 192:1) for the importance mapping output shown in Fig. 9. ROIs are decoded at a higher resolution than regions in its periphery. This improves the interpretability in ROIs and thus allows a user of the image to interpret the image contents at lower bit-rates.

background, thus permitting improved image content recognition performance.

3 Conclusion

This paper has presented an importance mapping technique that can be used for image content coding and prioritisation. An eye tracker was used to locate and trace a viewer's eye movements over a visual stimulus, and this was subsequently input into an importance mapping algorithm to automatically identify fixated ROIs. Such an approach incorporates the spatial and temporal characteristics of the human visual attention system, and thus overcomes many of the challenges involved with image processing algorithms in locating or segmenting image contents in cluttered scenery. The importance mapping scheme can also be input to ROI coding schemes such as those defined and proposed for JPEG 2000. Results show that improved interpretability in ROIs at low bit-rates (or high compression ratios) can be achieved. Future advances in gaze-based image coding technologies has potential use in providing real-time solutions for 'smart' image dissemination applications.

References

[1] Information technology – JPEG2000 image coding system – Part 1: Core coding system, ISO/IEC 15444-1 (ITU-T Rec. T.800) (August 2002).

[2] M. P. Eckert, A. P. Bradley, Perceptual quality metrics applied to still image compression, Signal Processing 70 (1998) 177–200.

[3] W. Osberger, Perceptual vision models for picture quality assessment and compression applications, Ph.D. thesis, School of Electrical and Electronic Systems Engineering, Queensland University of Technology, (Brisbane, Australia) (March 1999).

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. on Image Processing 13 (4) (2004) 600–612.

[5] A. Nguyen, Importance prioritised image coding in JPEG 2000, Ph.D. thesis, School of Engineering Systems, Queensland University of Technology, (Brisbane, Australia) (January 2005).

[6] Information technology – JPEG2000 image coding system – Part 2: Extensions, ISO/IEC 15444-2 (ITU-T Rec. T.801) (August 2002).

[7] R. Grosbois, D. Santa-Cruz, T. Ebrahimi, New approach to JPEG 2000 compliant region of interest coding, in: Proc. Applications of Digital Image Processing XXIV, Vol. 4472, San Diego, USA, 2001, pp. 267–275.

[8] Z. Wang, A. C. Bovik, Bitplane-by-bitplane shift (BbBShift) - A suggestion for JPEG2000 region of interest image coding, IEEE Signal Processing Lett. 9 (5) (2002) 160–162.

[9] Z. Wang, S. Banerjee, B. L. Evans, A. C. Bovik, Generalized bitplane-

by-bitplane shift method for JPEG2000 ROI coding, in: Proc. Int'l Conf. on Image Processing, Vol. 3, Rochester, New York, 2002, pp. 81–84.
[10] L. Liu, G. Fan, A new JPEG2000 region-of-interest image coding method: Partial significant bitplanes shift, IEEE Signal Processing Lett. 10 (2) (2003) 35–38.
[11] D. S. Taubman, M. W. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice, Kluwer Academic Publishers, Boston, 2002.
[12] A. Nguyen, V. Chandran, S. Sridharan, Importance prioritisation in JPEG 2000 for improved interpretability, Signal Processing: Image Communication 19 (10) (2004) 1005–1028.
[13] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, JPEG2000 region of interest coding - A hybrid coefficient scaling and code-block distortion modulation method, in: Proc. Australasian Workshop on Signal Processing and Applications, Brisbane, Australia, 2002, pp. 59–62.
[14] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, Vision Research 40 (2000) 1489–1506.
[15] W. Osberger, A. J. Maeder, Automatic identification of perceptually important regions in an image, in: Proc. Int'l Conf. on Pattern Recognition, Brisbane, Australia, 1998, pp. 701–704.
[16] W. Osberger, A. Rohaly, Automatic detection of regions of interest in complex video sequences, in: Proc. Human Vision and Electronic Imaging IV, Vol. 4299, 2001, pp. 361–372.
[17] C. Privitera, L. Stark, Evaluating image processing algorithms that predict regions of interest, Pattern Recognition Lett. 19 (1998) 1037–1043.
[18] F. Stentiford, An evolutionary programming approach to the simulation of visual attention, in: Proc. IEEE Congress on Evolutionary Computation, Seoul, Korea, 2001, pp. 851–858.
[19] A. J. Maeder, Importance maps for adaptive information reduction in visual scenes, in: Proc. Australian and New Zealand Conf. on Intelligent Information Systems, Perth, Australia, 1995, pp. 24–29.
[20] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, Importance assignment to regions in surveillance imagery to aid visual examination and interpretation of compressed images, in: Proc. Int'l Symp. Intelligent Multimedia, Video and Speech Processing, Hong Kong, 2001, pp. 385–388.
[21] R. Prandolini, Coding of surveillance imagery for interpretability using local dimension estimates, in: Proc. Visual Communications and Image Processing, Perth, Australia, 2000, pp. 516–526.
[22] F. Stentiford, An estimator for visual attention through competitive novelty with application to image compression, in: Proc. Picture Coding Symposium, Seoul, Korea, 2001, pp. 101–104.
[23] W. Osberger, A. J. Maeder, N. Bergmann, Perceptually based quantization technique for MPEG encoding, in: Proc. SPIE, Vol. 3299, 1998, pp. 148–159.

[24] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Trans. on Image Processing 13 (10) (2004) 1304–1318.
[25] C. Privitera, L. Stark, Focused JPEG encoding based upon automatic pre-identified regions-of-interest, in: Proc. Human Vision and Electronic Imaging IV, Vol. 3644, San Jose, USA, 1999, pp. 552–558.
[26] A. P. Bradley, F. W. M. Stentiford, Visual attention for region of interest coding in JPEG 2000, J. Visual Communication & Image Representation 14 (3) (2003) 232–250.
[27] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, Importance coding of still imagery based on importance maps of visually interpretable regions, in: Proc. Int'l Conf. on Image Processing, Vol. 3, Thessaloniki, Greece, 2001, pp. 776–779.
[28] D. S. Wooding, Fixation maps: quantifying eye-movement traces, in: Proc. of Eye Tracking Research & Applications, New Orleans, Louisiana, 2002, pp. 31–36.
[29] A. Santella, D. DeCarlo, Robust clustering of eye movement recordings for quantification of visual interest, in: Proc. of Eye Tracking Research & Applications, San Antonio, Texas, 2004, pp. 27–34.
[30] J. I. Khan, O. Komogortsev, A hybrid scheme for perceptual object window design with joint scene analysis and eye-gaze tracking for media encoding based on perceptual attention, in: Proc. Visual Communications and Image Processing, San Jose, California, 2004, pp. 1341–1352.
[31] A. Nguyen, V. Chandran, S. Sridharan, Visual attention based ROI maps from gaze tracking data, in: Proc. Int'l Conf. on Image Processing, Singapore, 2004, pp. 3495–3498.
[32] A. Nguyen, V. Chandran, S. Sridharan, GazeJ2K: A gaze-influenced JPEG 2000 image coder, in: Proc. Workshop on the Internet, Telecommunications and Signal Processing, Adelaide, Australia, 2004, pp. 18–23.
[33] A. Nguyen, V. Chandran, S. Sridharan, Gaze-J2K: Gaze-influenced image coding using eye trackers and JPEG 2000, Journal of Telecommunications and Information Technology, (In Press).
[34] EyeTech Digital Systems, Quick glance eye-gaze tracking system, http://www.eyetechds.com/ (2005).
[35] The Mathworks Inc., Matlab user guide: Statistics toolbox, version 5 (R14) (2004).
[36] A. P. Bradley, F. W. M. Stentiford, JPEG 2000 and region of interest coding, in: Proc. Digital Image Computing Techniques and Applications, Melbourne, Australia, 2002, pp. 303–308.