VISUAL ATTENTION BASED ROI MAPS FROM GAZE TRACKING DATA

Anthony Nguyen, Vinod Chandran and Sridha Sridharan

Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane QLD 4001, Australia
{an.nguyen, v.chandran, s.sridharan}@qut.edu.au

ABSTRACT

The use of visual attention (VA) spatial and temporal characteristics, monitored by a gaze-tracking device, to generate a region of interest (ROI) 'importance' map is proposed. A K-means clustering approach is adopted to group gaze location points into a number of clusters to represent the loci of regions of VA (or ROIs). Several metrics are then derived from the gaze positions and sequences to quantify the relative importance of the K-means clusters. An entropy-weighting strategy is adopted for the combination of these metrics to generate the ROI map. Results show that the ROI map is robust to the number of clusters and different gaze patterns, and can be used in progressive image coding/decoding to enhance the image quality in regions of interest.

1. INTRODUCTION

Visual attention (VA) has become an expanding area of interest in human computer interaction. Studies in VA and eye movements [1, 2, 3] have shown that humans generally only attend to a few regions (or loci) of interests, called ROIs, in an image, which are determined in part by their information content. The use of eye movements to determine ROIs can be promising in enhancing the quality of these regions during progressive image coding/decoding. Here we propose to exploit the personal aspects of the VA processes, by means of monitoring eye movements, to determine the loci and relative importance of ROIs. This loci and importance assigned to the ROIs generate the ROI map.

2. VISUAL ATTENTION AND EYE MOVEMENT

The human visual system relies on the positioning of the eyes (fixations) to parts of an image, which are processed with high resolution. A number of these fixations at specific locations are interspersed by rapid eye movements (saccades) to reposition the eye at the next point of attention. A succession of these fixation-saccade events provide the brain with detailed visual information from which a conceptual image of the visual scene can be constructed by combining the high resolution regions with the large area of low resolution information from the periphery [2, 4].

In order to understand VA processes better, methods have been devised to track gaze location through eye movements. By observing where and when a person's gaze is directed, it is possible to establish the fixation-saccade path followed by the viewer. If the viewer is undertaking a defined task under a given context, such as locating ROIs in an image, then the captured spatial and temporal gaze pattern given by the fixation-saccade events can be post-processed to compute the loci and relative importance of a representative subset of ROIs to generate the ROI map.

Fig. 1. Original 'Rockclimb' image used in experiments.

The general problem of characterising ROIs is that ROIs are determined partly by the application and hence the class of imagery. The brain and visual system is also subject-dependent. Regardless of these factors, there are areas of VA and sequences of transitions between them, durations of attention, spatial extent of these areas, etc, that constitute a response to any particular stimulus. This may differ from subject to subject and image to image, but the structure of the gaze pattern is usable with parameters to determine the loci and importance of ROIs for each case.

3. PROPOSED VISUAL ATTENTION BASED ROI MAP

The experimental methodology adopted consisted of recording gaze data of three different viewers for the image shown in Fig. 1. For each image, the viewer was free to examine the image consciously (i.e. overtly), which was directed by their natural VA processes. For each case, the gaze-tracking experiment was repeated three times for a duration of 10 seconds each.

3.1. Gaze-Tracking Device

The device used to record eye movements was an EyeTech video-based corneal reflection eye tracker. Infrared lights were mounted on both sides of a computer monitor to illuminate the eye and provide reference points for the eye tracker. The method of operation relies on tracking a bright "Purkinje" reflection on the eye from

0-7803-8554-3/04/$20.00 ©2004 IEEE.

the infrared light sources. The reflection used is relative to the location of the pupil centre. The gaze-tracker operated at 15 frames per second (fps) (uniformly sampled in time) and was conducted on an image and screen resolution of 1024×768 pixels.

3.2. Gaze Data Clustering

A clustering procedure is required to reduce the spatial characteristics of gaze patterns into a limited subset of clusters that represent ROIs. The choice of clustering technique is influence 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 a K-means algorithm, may be used.

The K-means clustering method assigns data to one of K clusters using the distance from the means of these clusters. A data vector is assigned to the nearest cluster mean. After all data vectors are classified, the means are updated using the sample means of the data vectors assigned to that cluster. The process is iterated until convergence (i.e. the means do not change significantly when compared against a precision threshold). The K initial values for the cluster means are chosen randomly from the data set. The value of K can be arbitrarily chosen or can be based on examination of the typical gaze tracking data for the application.

The cluster means and covariances of the locations of the data vectors that were assigned to the clusters were used to generate ellipses to represent the boundaries of regions of VA. The major and minor radial components of the ellipse were chosen to be 1.96 standard deviations in each direction. In such a case, if the cluster's spatial distribution was Gaussian, then this will represent approximately 95% of data points belonging to the cluster.

Fig. 2 presents some gaze data and associated ellipses representing the cluster means and covariances for a number of K values. Plots (a)–(c) show the K-means clusters for Scan 1 of Person 1 for $K = 2, 3$, and 4. Plots (d)–(f) represents fixations and clusters for Scan 1 for Person 1, Person 2, and Person 3 respectively, for $K = 5$. The plots in Fig. 2 do not contain any information about the sequence these points were viewed in. The resulting clusters are dependent on the number of clusters, initial cluster means and the gaze data. Despite these dependencies, it can be observed that the main object (i.e. the rock-climber) is more or less covered by the cluster labelled '1'.

3.3. Cluster Importance Metrics

Some measures need to be derived from the spatial locations and temporal order of the gaze data in order to determine the relative importance of ROIs. These measures may be derived from important attentional features of human vision. The result of these importance metrics together with the clustering procedure define the ROI map. The measures employed in the ROI map are as follows:

- *Cluster Count*, $C(k)$, measures the number of gaze points that belong to cluster k. C is analogous to the duration of gaze within the cluster, since uniform gaze sampling was recorded. It represents the total time spent viewing/gazing at that region. The cluster importance is conjectured to be proportional to *Cluster Count* and is given by $C(k)/\sum_{k=1}^{k=K} C(k)$.
- *Cluster Distance*, $D(k)$, measures the average distance of gaze points from their cluster mean. *Cluster Distance* is conjectured to be inversely proportional to the cluster importance, which is given by $(1 - N_D(k))/\sum_{k=1}^{k=K}(1 - N_D(k))$ where $N_D(k)$ is the normalised average distance given by $D(k)/\sum_{k=1}^{k=K} D(k)$.
- *Cluster Variance*, $V(k)$, measures the variance of gaze points from their cluster mean. The cluster importance for this factor is similar to that for *Cluster Distance*, which increases with decreasing variance. The *Cluster Variance* importance metric is given by $(1 - N_V(k))/\sum_{k=1}^{k=K}(1 - N_V(k))$ where $N_V(k)$ is the normalised variance given by $V(k)/\sum_{k=1}^{k=K} V(k)$.
- *Cluster Area*, $A(k)$, measures the area (in pixels) of the ellipse (see Section 3.2) generated from the cluster mean and covariance. A is related to the size of the object or segment that may be of interest to the viewer. The cluster importance for *Cluster Area* increases with decreasing area. The *Cluster Area* importance metric is given by $(1 - N_A(k))/\sum_{k=1}^{k=K}(1 - N_A(k))$ where $N_A(k)$ is the normalised cluster area given by $A(k)/\sum_{k=1}^{k=K} A(k)$.
- *Cluster Time Weighted Visit Count*, $W(k)$, is used to weight the count (or duration) of the nth cluster visited by the inverse of the time of the cluster visit. The importance metric is given by the number of gaze points in cluster k of the nth cluster visited, $C_{n,k}$, divided by the nth cluster visited (i.e. $W_{n,k} = C_{n,k}/n$). For example, if 10 successive gaze points belonged to the first cluster, say 'Cluster 1', then the calculated measure would be $W_{1,1} = 10$ (i.e. $10/1$). Then if the gaze position shifted to another cluster, say 'Cluster 2', which then recorded 30 gaze points in succession, then the measure would give $W_{2,2} = 15$ (i.e. $30/2$), and so on. The importance measure for cluster k, is finally, given by the sum of $W_{n,k}$ that belongs to cluster k divided by the sum of $W_{n,k}$ (i.e. $W(k) = \sum_{n,i=k} W_{n,i}/\sum_{n,i} W_{n,i}$).
- *Cluster Revisit Count*, $R(k)$, measures the number of saccade revisits to a given cluster during the course of viewing. The revisiting of the fixation-saccade path to regions in an image has been found to be a fundamental property of eye movements. The cluster importance for *Cluster Revisit Count* is conjectured to be proportional to the number of cluster revisits. The *Cluster Revisit Count* metric is normalised by the total number of revisits in each cluster (i.e. $R(k)/\sum_{k=1}^{k=K} R(k)$).

Note that the sum of cluster measures for each metric equals 1.0; indicating the cluster's relative importance for the given metric.

3.4. Entropy-Weighted Importance Measures

The metrics used above to characterise the importance of a cluster may be quite correlated and an entropy-weighting strategy is proposed to appropriately de-correlate and weight each metric accordingly. The entropy of a metric, X, is defined as:

$$H(X) = -\sum_{x} P(x) log_2 P(x) \tag{1}$$

where $P(x)$ is the probability that X is in the state x, and $P(x)log_2P(x)$ is defined as 0 if $P(x) = 0$. Here, $P(x)$, is the probability that a given metric belongs to cluster x. That is, $P(x)$, is the same as the relative importance measure for cluster x. These

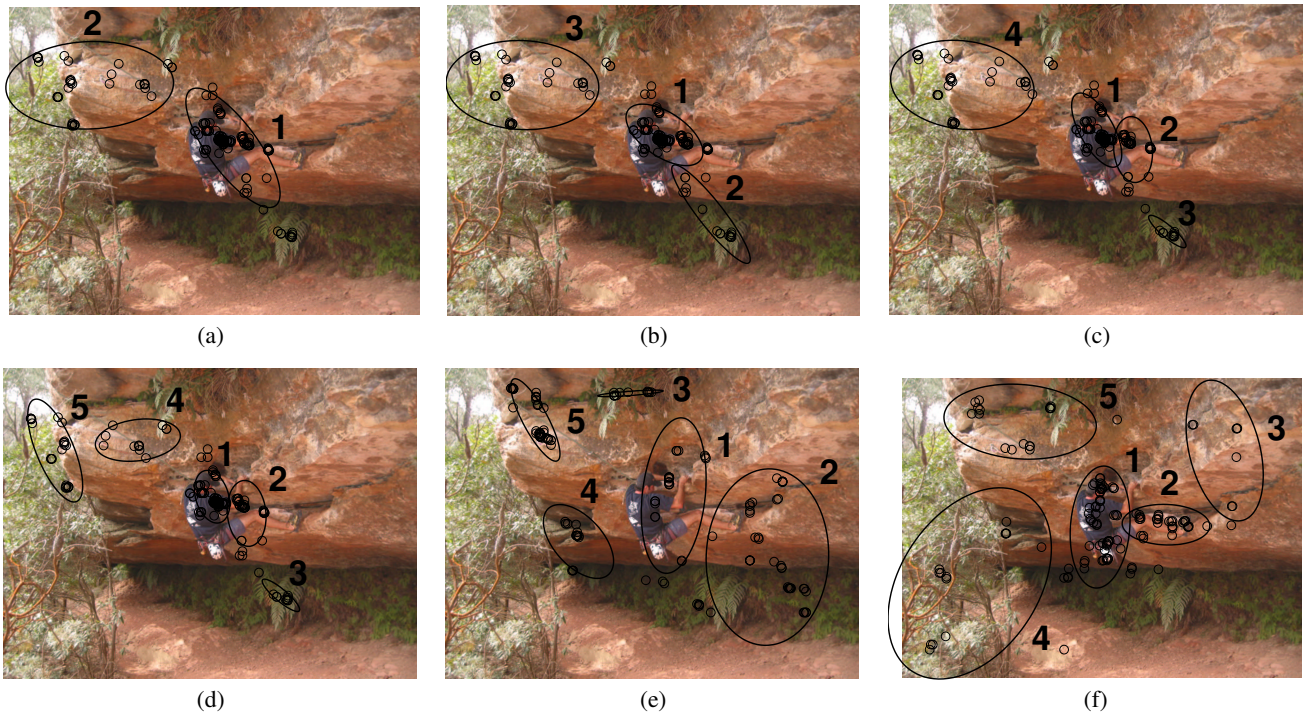

Fig. 2. Gaze locations (circles) and K-means cluster boundaries (ellipses); (a)–(c) shows Scan 1 of Person 1 for $K = 2, 3$, and 4, respectively, (d)–(f) shows Scan 1 of Person 1, 2, and 3, respectively, for $K = 5$. Clusters are labelled in order of importance (1:highest,K:lowest).

entropy measures are used to appropriately weight the importance measures for each metric.

Each metric was categorised into one of three classes to separate independent metrics from dependant ones, namely, (1) *Count*, and *Time Weighted Visit Count*; (2) *Distance*, *Variance*, and *Area*; and (3) *Revisit Count*. These classifications determine the different inter and intra-class weighting procedure for each metric. The entropy-weighted importance measures for a metric is obtained by weighting its cluster relative importance measure by the its inter and intra-class entropy weight, which are given by:

- *Inter-class weight* for a class is given by the average of its intra-class (dependent) entropies divided by the sum of the average entropies for each class.
- *Intra-class weight* for a metric is given by the metric entropy divided by the sum of all intra-class (dependent) entropies for the class that the metric belongs to.

The overall importance measures are obtained by the sum of all entropy-weighted metrics for each cluster. These measures for the clusters shown in Fig. 2 are tabulated in Table 1. The clusters are labelled in order of importance with 'Cluster 1' having the highest importance and 'Cluster K' having the least. For these images, one may conclude that the rock climber is the primary region of importance, which indeed coincides with the results obtained in all cases. Other regions in the background, however, may catch the attention of viewers in a more arbitrary fashion. Although, the actual importance values and their order of importance may depend on the application, viewer, and other subjective factors, the results show promise and are robust to the number of clusters and gaze patterns from the same and different viewers.

Another promising feature is how the entropy-weighted measures are higher than equally-weighted ones for regions that are intuitively more important (e.g. rock climber), and lower for less important regions (with the exception of Fig. 2(a)). An excellent example is Fig. 2(e), where the rock climber was not of most importance from the equally-weighted measures, but was significantly increased to become the most important after entropy-weighting.

4. APPLICATION TO IMAGE COMPRESSION

The coding/decoding of images may be influenced to enhance the image quality in regions of VA (or ROIs). The JPEG 2000 image coding standard provides several ROI coding mechanisms which can prioritise pre-defined ROIs. These methods, however, treat all ROIs with the same degree of importance. To overcome this, an importance prioritised JPEG 2000 (IMP-J2K) image coder [5] was developed to allow multiple ROI coding using variable ROI importance scores. In this method, the ROI was emphasised by weighting the Mean Square Error (MSE) distortion measure of a block of coefficients by the square of its importance score. The reconstruction of the ROIs are bounded by the extent of these blocks. This is advantageous for the given ROI boundaries, since the ROIs may not fully encompass the objects in the image.

The ROI cluster loci and importance measures were input to IMP-J2K for coding, with an additional background (i.e. regions outside the ROI) importance parameter of 0.01. Table 2 demonstrates the improved quality, measured as Peak Signal-to-Noise Ratio (PSNR), of the ROIs over the background (BG) for decoded images at 0.25 bits per pixel (bpp). In some cases, the ROIs were encoded in actual order of importance. The slight inconsistency

Table 1. Entropy and equally[†]-weighted ROI cluster importance measures for Fig. 2.

Figure	K	Cluster				
		1	2	3	4	5
2(a)	2	0.65 (0.70)	0.35 (0.30)	NA (NA)	NA (NA)	NA (NA)
2(b)	3	0.56 (0.52)	0.32 (0.34)	0.12 (0.14)	NA (NA)	NA (NA)
2(c)	4	0.41 (0.34)	0.28 (0.24)	0.23 (0.34)	0.08 (0.08)	NA (NA)
2(d)	5	0.37 (0.33)	0.25 (0.21)	0.20 (0.27)	0.11 (0.11)	0.07 (0.08)
2(e)	5	0.29 (0.22)	0.20 (0.16)	0.19 (0.32)	0.17 (0.15)	0.15 (0.15)*
2(f)	5	0.38 (0.34)	0.29 (0.30)	0.12 (0.15)	0.11 (0.10)	0.10 (0.11)

[†] Shown in parentheses

Table 2. PSNR (dB) of ROI clusters and background (BG) for decoded images at 0.25 bpp[†]. Bracketed values are for prioritisation of 'rock climber' only clusters.

Figure	Cluster					BG
	1	2	3	4	5	
2(a)	33.12 (34.62)	29.02 (NA)	NA (NA)	NA (NA)	NA (NA)	26.13 (24.49)
2(b)	32.84 (35.15)	33.75 (NA)	26.63 (NA)	NA (NA)	NA (NA)	25.92 (25.24)
2(c)	32.10 (35.33)	35.44 (35.24)	32.46 (NA)	26.57 (NA)	NA (NA)	25.93 (24.36)
2(d)	34.85 (35.64)	34.55 (35.16)	32.04 (NA)	27.99 (NA)	23.55 (NA)	25.83 (25.10)
2(e)	32.28 (33.67)	32.75 (NA)	28.74 (NA)	32.53 (NA)	28.85 (NA)	26.84 (28.54)
2(f)	31.58 (33.69)	31.55 (32.84)	29.60 (NA)	26.22 (NA)	28.14 (NA)	29.55 (26.43)

[†] Fine differences, say ± 1 or 2 dB, between ROIs may be perceived as similar quality.

between importance measures and the quality are a result of a number of factors including variations in the ROI's content and the effect of a number of ROI coding attributes such as the number of ROIs, and its size and location with respect to the code-block boundaries. Furthermore, the seeping of importance of higher to lower important ROIs that are close in proximity can correspondingly place more emphasis on the lesser important ROI. In practice, a small number of ROIs occupying a relatively small portion of the image are used as ROIs [6]. If this was the case, for example, only using ROI clusters that belong to the rock climber for prioritisation, then the decoded quality of ROIs can be better controlled and be reconstructed significantly better than the background and also in order of importance (see bracketed values in Table 2).

Fig. 3. Example prioritised ROI coding at 0.25 bpp for Fig. 2(d).

Fig. 3 shows an example decoded image for the case where ROIs, as shown in Fig. 2(d), were used for prioritisation. Note that ROIs, especially the rock climber, are reconstructed with better quality and at a higher resolution than the background. This observation is similar to the operation of human vision as discussed in Section 2 where regions of VA are in high resolution while the periphery is in low resolution.

5. CONCLUSION

A ROI 'importance' map was proposed to exploit the personal aspects of the VA processes, by means of monitoring eye movements, to determine the loci and relative importance of ROIs. The ROI map shows promise and is robust to the number of clusters and different gaze patterns. The metrics and entropy-weighting strategy also provides an adequate impression of the cluster's 'importance'. Future work will apply the ROI map to a greater number of images and viewers, develop a more concrete set of metrics, and consider issues such as the omission of outliers and the determination of an appropriate number of clusters and their validity. The ROI map was also applied to image coding to prioritise ROIs.

6. REFERENCES

[1] D. Norton and L. Stark, "Eye movements and visual perception," *Sci Am*, vol. 224, pp. 34–43, 1971.

[2] C. Privitera and L. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. on PAMI*, vol. 22, no. 9, pp. 970–982, September 2000.

[3] A. L. Yarbus, *Eye movements and vision*, Plenum, New York, 1967.

[4] A. Maeder and C. Fookes, "A visual attention approach to personal identification," in *ANZIIS*, 2003, pp. 55–60.

[5] A. Nguyen, V. Chandran, S. Sridharan, and R. Prandolini, "Importance prioritisation coding in JPEG2000 for interpretability with application to surveillance imagery," in *VCIP*, 8-11 July 2003, vol. 5150, pp. 806–817.

[6] A. Nguyen, V. Chandran, S. Sridharan, and R. Prandolini, "Guidelines to using region of interest coding in JPEG 2000," in *DSPCS*, 8-11 December 2003, pp. 183–188.