Inherent Visual Information for Low Quality Image Presentation

Justin Boyle, Anthony Maeder, Wageeh Boles
School of Electrical and Electronic Systems Engineering
Queensland University of Technology, Brisbane, Australia
Email: jr.boyle@qut.edu.au, a.maeder@qut.edu.au, w.boles@qut.edu.au

Abstract
This paper describes our experiments to quantify the inherent information content in images as a means to optimally present images where display pixels are limited. Such low image quality applications include visual prostheses, or "bionic eyes", where implant electrodes are limited in number. Results from subjective tests with 225 normally sighted viewers are compared to predictions made with a metric for information content.

Keywords
Information content, image presentation, low quality images

INTRODUCTION
Ideally image displays should be of highest spatial resolution for adequate human perception. However in cases where size or manufacturing constraints limit the number of display pixels possible, intelligible perception is still desired from these low quality coarse images. One example is the developing area of visual prostheses, or "bionic eyes", where implanted electrodes in contact with nerve cells in the visual pathway are stimulated by electric pulses. Electrode array sizes of current prototypes include 25x25 [1] and 10x10 [2] which result in significant information loss.

We are reviewing image processing methods to efficiently use limited display pixels. Previously we have determined the impact of several novel image processing techniques on object recognition [3]. The concept of importance mapping was found to improve recognition of low quality images. Importance mapping aims to predict where the human eye will fixate in an image, ie. what are the salient areas or regions of interest in an image.

In this paper we propose an improved model which maximizes the information content in the resulting final saliency/importance map. We describe the Importance Map concept and then introduce the concept of Visual Information Content. Our psychophysical experiments to quantify this term are outlined along with the development of our metric for information content in images. Results of subjective visual information are compared to metric predictions. Finally we show that subjective information content is correlated with object recognition and is thus a suitable measure to use to optimise image presentation.

IMPORTANCE
Several region-of-interest algorithms which predict where the human eye fixates on an image are reported in the literature (eg. [4-7]). When compared against subjective tests using eye-tracking machines or similar attention-recording devices, the region-of-interest algorithms correlate highly. An extension of these algorithms is the concept of assigning an importance score or weighting to each area in an image to generate an "importance map" [5,7]. This importance ranking has previously been applied in visually lossless compression, where improved compression ratios have been achieved with high perceived image quality.

Several image features are known to influence attention in the human viewer, including motion, location, contrast, size and shape. Feature maps/images are developed representing each feature and then combined to form an overall importance map. Several combination strategies have been attempted, ranging from linear summation of features (all weighted equally) [4,5] to weights selected in accordance with eye-tracker data [7]. We propose a new method where feature map weights are selected iteratively to maximise the information content in the resulting importance map. Thus there is a need for defining the concept of information content.

QUANTIFYING INFORMATION CONTENT
We have conducted experiments to attempt to quantify the amount of inherent visual information in images. In the experiments images were compared with each other to obtain a ranking from most to least visually informative.

There were 9 image quality classes tested. Original images were 256x256 pixels representing a range of scene types. A decreasing image quality scale was presented using spatial resolutions typical of visual prosthesis designs (25x25, 16x16, 10x10) and reducing the grey levels from full greyscale to binary. It was also of interest to expose the structure of an image by presenting image edges.

A Metric for Information Content
Reduced quality image sets were prepared for each of the images shown in Figure 1. A visual information metric was developed from analyzing the subjective scores of subjects ranking the images.

Figure 1 – Visual Information Metric obtained from 7 images representing a range of scene types.

Participation was on a voluntary basis and comprised 271 Year 11 students and 11 mature age respondents. 57 questionnaires (21%) were rejected due to invalid data. Thus the final sample size was 225, representing sample sizes of 25 for each of the 9 image quality classes.

Participants had no prior knowledge of the images. Booklet instructions stated that a range of high quality and low quality images could be expected, and although the low quality images might just appear as a range of blocks, they may be similar to what a blind person might see with a bionic eye. Viewing conditions for the experiment were not controlled.

Two questionnaire-based methods were used:

<u>1) Seven images presented all on one page</u>

The following instruction was presented with the images:

> *Rank the images shown on each page for visual information. Place a number in each box beside the image.*
>
> *Rank the images for how much visual information they contain:*
>
> *1 = contains most visual information*
>
> *7 = contains least visual information*

<u>2) Paired comparison (binary decision) questionnaire test</u>

Subjects were presented with an image pair and the instruction:

> *WHICH IMAGE APPEARS TO CONTAIN MORE INFORMATION?*
>
> *Which image could you answer the most questions about? (eg. What is the scene? How many objects?)*
>
> *If you had to rely on only one of the images to perform a task which would it be?*

Subject response was measured on a 5-point scale:

Box 1 = left image has much more information than right image

Box 2 = left image has slightly more information than right image

Box 3 = images have same amount of visual information

Box 4 = right image has slightly more information than left image

Box 5 = right image has much more information than left image

Both methods gave similar results for ranking of subjective information content. For example, when considering the ranking for all quality classes (n=225) both methods gave the following near identical ranking order:

Face > Flower > Tree > Buildings > Lighthouse/Capsicum > Balloon.

Subjective rankings have been used to propose a metric to quantify visual information that is stable across all image quality classes (not just the ones used in these tests). 15 image attributes were considered for the visual information metric:

1. file size
2. standard deviation
3. maximum standard deviation in 4 image quadrants
4. variance
5. maximum variance in 4 image quadrants
6. entropy
7. number of edges
8. number of segments
9. fractal dimension
10. 11. 12. image internal similarity measures
13. 14. 15. image symmetry measures

Three measures were used for image internal similarity (exact match across x and y axes) and image symmetry (mirror match across x and y axes):

- exact pixel match - no sub-block analysis (same result operating on big or small block)
- shaded pixel difference between blocks - 5 level sub-block analysis (objects might be in a different position within a block)
- average pixel value - 5 level sub-block analysis

Stepwise regression was used to search for the optimum subset of variables. The procedure was based on sequentially introducing variables into the model one at a time and testing the significance of all variables at each stage. The most stable performance was found to be from a metric consisting of the number of image edges alone.

ie. Information Content = *f*(edges)

This is an interesting result considering Marr's emphasis of zero crossing (edge) detection in producing images of the external world [8]. This includes their role in the formation of a primal sketch to derive shape information from images, and biological mechanisms for detecting oriented zero-crossing segments in retinal ganglion cells.

This metric is now validated against additional data collected in the experiment.

Validating the Information Content Metric

A number of additional aspects/dimensions were explored to provide data to validate the metric and also determine what impact (if any) they had on perceived information content. These issues were assessed by comparing sets of 3 images against each other.

Predictive performance of the information content metric is tested against these results. The additional dimensions explored are shown below in Figure 2. Low quality image sets were developed for the images and subjects were asked to rank the images for the amount of visual information they contain.

1. No. of Objects

3 images of increasing object number.

2. Angle of Object

3 images of a fruit bowl at 90, 45 and 0 degrees.

3. Distance to Object

3 images of a couple on a bicycle with decreasing distance to the couple's faces.

4. Closeness between Image Objects

3 images of different couples with decreasing closeness between the couple.

5. Image Detail

3 images of the same face with different edge detail (phone and second face).

6. Contrast between Objects & Surround

3 images of capsicums with varying contrast

7. Variety of Object Types

3 images comparing different object types (orange, sunglasses, scissors, mug).

Figure 2 – Visual Dimensions used to validate the metric

Results

63 visual information rankings were obtained (7 factors/dimensions x 9 image quality classes). Dominant patterns (ie. the most frequently specified ordering in terms of perceived information content) were identified for each case. The strength of the dominant patterns (ie. the frequency with which that pattern was specified by observers) ranged from 96% (24 of 25 respondents ranking images in that order) to 28% (only 7 of 25 respondents). The number of cases for each ten percentile class is given in the first column of Table 1.

Table 1 – Metric Performance

	Strength and number of cases for dominant viewer patterns (63 in total)	Frequency of image with highest info content being predicted by metric	Frequency of exact ranking being predicted by metric
STRONG	90-100%: 3	100%	100%
	80-89%: 4	75%	75%
	70-79%: 1	100%	100%
Dominant	60-69%: 12	67%	25%
patterns	50-59%: 6	83%	50%
	40-49%: 16	38%	19%
	30:39%: 19	32%	21%
	20-29%: 2	100%	100%
	10-19%: 0	-	-
WEAK	0-9%: 0	-	-

The performance of the Information Content metric in predicting subjective dominant viewer patterns is also shown in Table 1. Out of the 63 test cases examined, three cases had 90% or above consensus from subjects viewing the sample set. For each of these cases, the metric successfully predicted not only which of the 3 images had the highest information content (2nd column above) but also the ranking order chosen by subjects (3rd column above). Metric performance at weaker subject consensus levels are also shown.

It was of interest to further examine strong dominant viewer patterns in the data. Eight of the 63 rankings had 70% or above consensus among viewers. Five of these related to the number of objects in the scene.

Strong viewer preferences are shown in Figure 3.

Number of Objects in Scene

5 image quality classes: 16x16_Binary (88%), 25x25_Binary (92%), 256x256_Edges (84%), 256x256_Binary (96%), 256x256 (96%)

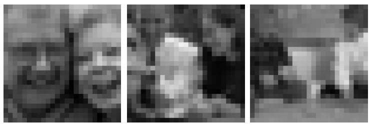
Highest → Lowest

All 5 cases predicted by metric?: Yes

Closeness between image objects

1 image quality class: 25x25greyscale set (80%)

Highest → Lowest

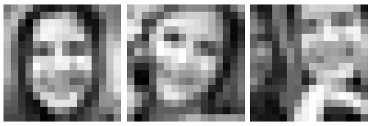


Case predicted by metric?: Yes

Image Detail

1 image quality class: 16x16greyscale set (80%)

Highest → Lowest



Case predicted by metric?: No

(Metric prediction: phone > 2 faces > single face)

Contrast between Objects & Surround

1 image quality class: 256x256_Edge set (72%)

Highest → Lowest

Case predicted by metric?: Yes

Figure 3 – Strong viewer preferences (70% or above consensus among viewers) showing images ranked from highest to lowest perceived information content

Four conclusions can be drawn from Figure 3:

1. the more objects in the scene, the higher the visual information
2. the closer the objects in the scene, the higher the visual information
3. a simple face with no surrounding clutter was most visually informative at low resolution levels
4. strong edges, arising from high intensity contrast, correspond with high perceived information content

The visual information metric predicted 7 of the 8 strong viewer preferences (70% or above consensus level). Viewers of the 16x16 greyscale Image Detail set ranked a simple face as containing most visual information, while the metric ranked the image of the phone and two faces ahead of the single face. The familiarity and strong recognition of the human face at low levels of image quality may cause viewers to select it over others containing unrecogniseable blobs.

The metric was found to work best with binary images, which are expected from at least early prototype designs. (Limited greyscale may be possible by modulating stimulus amplitude, frequency and pulse duration [9]). The number of ranking cases where the metric was able to predict the image with the highest information content is shown in Table 2 below. There are a total of seven ranking cases for each image quality class, corresponding to each visual dimension explored.

Table 2 – The number of correct metric predictions of images with the highest information content

10x10 Binary set - 4/7	10x10 Greyscale set – 1/7
16x16 Binary set - 6/7	16x16 Greyscale set – 1/7
25x25 Binary set - 4/7	25x25 Greyscale set – 3/7
256x256 Binary set - 6/7	256x256 Greyscale set – 3/7
256x256 Edge set - 6/7	

This may be another reason why the metric prediction for the 10x10 greyscale Image Detail set did not agree with the ranking chosen by 80% of viewers. Table 2 shows that for 16x16 greyscale images, the metric was successful in predicting the image with the highest information content in only 1 out of 7 cases. However for 16x16 binary images, the metric prediction was correct for 6 out of 7 cases. It should be remembered that the strength of dominant patterns on which metric performance is assessed range from 96% to 28%. At high levels of viewer consensus, the metric is accurate in predicting images with the highest information content, and is thus considered acceptable for this application.

It is useful to now show that this measure for information content is an adequate pointer to how well an image might be recognised.

CORRELATIONS BETWEEN RECOGNITION RATE AND PERCEIVED INFORMATION CONTENT

We wished to determine if there was any relationship between recognition rates and the amount of visual information as perceived by viewers.

The experiment also included a component where recognition ability was assessed. Subjects were presented with images shown in Figure 1 and the following instruction:

CAN YOU TELL WHAT IS SHOWN IN EACH IMAGE.

> *Write a word under each image to describe the main object or content of the scene.*
>
> *Put a circle around the images that you are confident about.*

There were no clues provided as to the context of the image (ie. an open-ended guess). Relationships between correct object recognition and subjective information content scores were obtained for each image quality class (for example, Figure 4 shows the relationship for 25x25 binary Paired Comparison experiments).

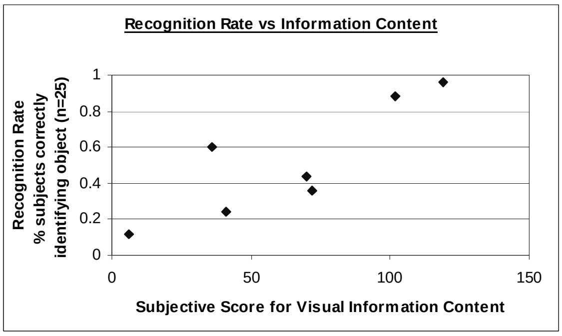


Figure 4 – Example relationship between recognition and information content (25x25 Binary Paired Comparison data)

We then assessed the significance of these relationships. Linear regression models for each quality class were developed for two series of data:

1. where images were presented at one time
2. paired comparison data

The significance of the models and correlation coefficients appears in Table 3 over.

There was some evidence for correlation between ranked information content and recognition rates with significance levels ranging from P=0.05 to P=0.1 for all but the 10x10 greyscale image set. Thus the concept of visual information content can be considered an adequate measure to optimise in importance map generation to enhance recognition.

Table 3 – Correlation coefficients between recognition rate and perceived information content

	Images presented at one time		Paired Comparison	
Image Quality Class	R Correlation Coeff.	Significance $F_{(1,7-1-1)}$ test	R Correlation Coeff.	Significance $F_{(1,7-1-1)}$ test
10x10 Binary	0.76	0.05	0.76	0.05
10x10 G.S	0.54	0.21	0.36	0.43
16x16 Binary	0.69	0.09	0.71	0.07
16x16 G.S	0.70	0.08	0.61	0.15
25x25 Binary	0.90	0.01	0.85	0.01
25x25 G.S	0.75	0.05	0.81	0.03
256x256Edge	0.70	0.08	0.66	0.10
256x256 Bin	0.69	0.09	0.73	0.06

CONCLUSIONS

In the field of low quality vision, there is a need for delivering maximum scene information to a limited number of display electrodes/pixels. In this paper we have proposed a method to enhance recognition using importance maps weighted to maximise the "information content" in the resulting importance map. We have described our experiments to quantify this term. The number of edges in an image was found to be the best statistic out of a 15-variable multiple regression analysis, to correlate with subjective rankings of visual information. The metric was tested on additional data and found to be appropriate in assessing information content. Finally we showed that subjective information content was significantly related to object recognition. We are applying this now to generating improved importance maps which will be compared to other predictive algorithms and eye-tracker data.

REFERENCES

[1] Normann R, Maynard E, Rousche P, Warren D, A neural interface for a cortical vision prosthesis, Vision Research 39(15), pp. 2577-2587, 1999

[2] Suaning G, Lovell N, CMOS Neurostimulation System with 100 Electrodes and Radio Frequency Telemetry, Inaugural Conference of the IEEE EMBS (Vic), Melbourne, pp.37-40, Feb 1999

[3] Boyle J, Maeder A, Boles W, *Image Enhancement for Electronic Visual Prostheses*, Australasian Physical & Engineering Sciences in Medicine Journal 25(2), pp.81-86, 2002

[4] Itti L, Koch C, Feature combination strategies for saliency-based visual attention systems, Journal of Electronic Imaging, 10(1), pp.161-169, 2001

[5] Osberger W, Maeder A, "Automatic identification of perceptually important regions in an image using a model of the human vision system", *14th International Conference on Pattern Recognition, Brisbane, Australia*, pp. 701-704, 1998

[6] Privitera C, Stark L, "Focused JPEG encoding based upon automatic pre-identified regions-of-interest," in *Human Vision and Electronic Imaging IV*, Rogowitz T, Pappas T, Editors, Proceedings of SPIE Vol. 3644, pp. 552-558, 1999

[7] Osberger W, Rohaly A, "Automatic detection of regions of interest in complex video sequences," in *Human Vision and Electronic Imaging VI*, Rogowitz T, Pappas T, Editors, Proceedings of SPIE Vol. 4299, pp.361-372, 2001

[8] Marr D, Vision, W.H. Freeman & Co, New York, pp.54-79, 1982

[9] Suaning G, Lovell N, Schindhelm K, Coroneo M, *The bionic eye (electronic visual prosthesis): A review*, Australian and New Zealand Journal of Ophthalmology 26, 195-202, 1998