Image categorisation based on the spatiochromatic information

Zhaohui Wang, Vien Cheung† and Stephen Westland†

State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, China
College of Information Science and Technology, Hainan University, China
†School of Design, University of Leeds, UK
Email: wangzhaohui@hotmail.co.uk

An imaging workflow to categorise images into business graphics and pictorial images based on spatiochromatic information is demonstrated. Studies on three computational characteristics (lightness frequency distribution, average chroma distribution, and Fourier spectrum) show that differences between business graphics and pictorial images can be discriminated mathematically. To enhance the performance of the image categorisation workflow, a multilayer perceptron neural network was designed and experimentally evaluated, and which performed with an accuracy of 95% using a database consisting of 57 randomly selected images.

Received 11 July 2019; revised 05 January 2020; accepted 25 January 2020
Published online: 07 February 2020

Introduction

Due to recent advances in colour imaging science and technology, there has been much progress in the field of colour reproduction. As a result, users have gained more confidence to apply colour images for information storage and communication. Colour images have been used for a wide scope of applications such as education, entertainment, commerce, medicine, science, etc. Ever-higher quality of colour image reproduction is expected to satisfy these wide applications.

A key issue in the process of colour image reproduction is how to cope with the fact that different colour technologies are capable of producing different ranges of colours (known as gamut). Gamut mapping algorithms (GMAs) have been of much interest in colour image reproduction for decades to fully exert the capacity of a certain medium to obtain required colour image reproductions. From the literature of gamut mapping, one can see that the most noticeable trend is that image-dependent methods are preferred over medium-dependent methods. Attempts have been made to take into account image categories and characteristics in the gamut mapping process. It is obvious that an automatic image categorization algorithm could be a favourable pre-processing step for choosing the appropriate GMA for a given type of image in a colour image reproduction workflow.

Within the scope of this study, image categorization is defined to categorise images into business graphics and pictorial images. Business graphics, typically depicting numeric data, are usually symbolized by graphical elements, such as, text, charts, and graphic illustrations to assist in the business activities [1-2]. A pictorial image can be described as a digital representation of a real-world scene: objects, living beings, etc. [3]. Figure 1 gives examples to illustrate these descriptions. Figure 1 (a) presents a typical business graphic with text labels and graphic map; Figure 1 (b) shows a pictorial image of an outdoor scene with buildings, sky and water.

(a)


(b)

Figure 1: Examples of (a) business graphic and (b) pictorial image.

Image categorisation

It is possible for observers to make magnitude estimates in terms of image characteristics, including lightness, sharpness, noisiness or various degrees of colourfulness or naturalness. Some of these judgments can be, at least approximately, predicted computationally, e.g. by a colour appearance model [4]. Some image characteristics are computable, but, which would be unrecognizable by observers, e.g. the phase spectrum of an image, the histogram of a certain colour distribution.

Likewise, business graphics and pictorial images have a number of different image characteristics. By their descriptions, one can categorize an image easily. An image showing any of visual representation of numeric data (e.g. bar graphs), text for a particular information, illustrations (e.g. flow-chart, map) or their combinations will be labelled as a business graphic. On the other hand, a typical pictorial image has many details of natural scenes and hence a variety of rich textures from the presence of, such as, buildings, grass, sky, vegetation, mountain, etc. However, these descriptions prefer the contents of an image rather than computational image characteristics.

In terms of different levels of complexity, a large number of image characteristics have been defined, ranging from low level characteristics, such as edges, lines and colour, to the high level in terms of contents, such as objects or living beings and actions they performed. High level classification can be achieved by analysis of computational low level characteristics geared for the particular classes.

The algorithm is based on the assumption that an image in question depicts one or more characteristics and that each of these characteristics belongs to one of several distinct and exclusive categories.

Current study suggested three hypotheses upon business graphics and pictorial images. 1) The **frequency distribution of lightness values** in a graphical image will on average be sparser than in a pictorial image; 2) the **chroma** of graphical images will on average be greater than in pictorial images; 3) the **Fourier frequency distribution** of graphical images will be more structured than for pictorial images.

Lightness analysis

Since pictorial images and business graphics show obvious differences inherited from their origination, the lightness properties were investigated. Figure 2 illustrates the lightness frequency distribution of a business graphic and a pictorial image (as presented in Figure 1) respectively. The frequency distributions for these two images are apparently different. The pictorial images mostly are captured for presenting the natural scene and contain smoothly varying colours due to light continuously rendering in the natural world. Hence, the frequency distribution presented here for the pictorial image shows a smooth variation with continuous and slight changes between neighbouring points whereas this is not the case for the business graphic.

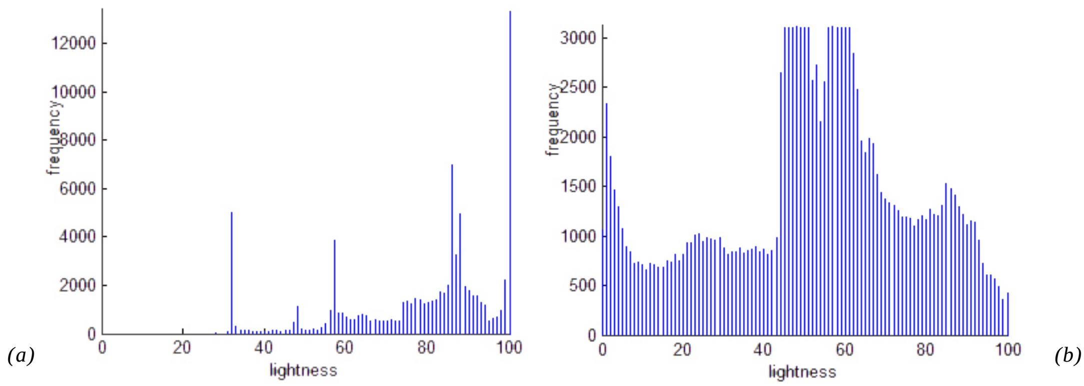


Figure 2: Lightness frequency distributions of (a) a typical business graphic and (b) a typical pictorial image.

A metric, LFD, was developed according to a discreteness statistic, which is sensitive to the sparseness that is evident in Figure 2 (a), thus.

$$LFD = \frac{\sum (f_i - f_{i+1})^2}{\sum f_i^2} \tag{1}$$

where f_i is the frequency of lightness range i. The normalisation provided by the denominator in Equation is necessary so that the LFD metric is not affected by image size. It is expected that, by Equation (1), business graphics product large values and pictorial images result in small values, which can then be separated by an optimised threshold. For example, the business graphic as shown in Figure 1 (a) gives a value of LFD=1.41; and the pictorial image in Figure 1 (b) gives a value of LFD=0.42.

Chroma analysis

Since business graphics typically include quite chromatic colours for the graphic elements: thus the average chroma value of business graphics might be expected to be greater than that of pictorial images. Figure 3 shows the chroma distribution of the business graphic and pictorial image corresponding to Figure 1 (a) and (b) respectively. It is obvious that the chroma of the typical business graphic distributes in a wide area and has large values at high chroma components, on the other hand, that of the typical pictorial image is in a limited range with small chroma values.

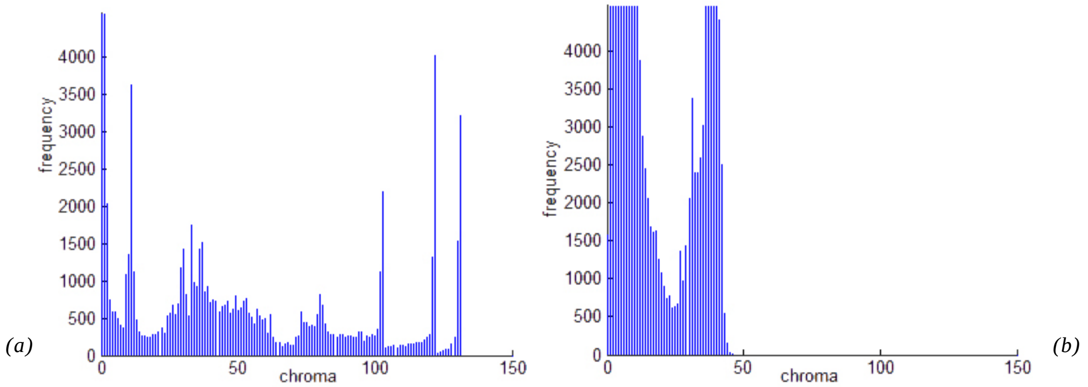


Figure 3: Chroma frequency distributions of (a) a typical business graphic and (b) a typical pictorial image.

Hence, the average chroma (AVGC) value is used to present this difference:

$$AVGC = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} C_{ij} \tag{2}$$

where $M \times N$ is the image size and C_{ij} represents the value of chroma at pixel (i, j). It is expected that business graphics give large AVGC values and pictorial images result in small AVGC values which can then be separated by an optimized threshold. The chroma frequency distributions in Figure 3 yield average chroma values, AVGC=37 for the business graphic, and AVGC=16 for the pictorial image, respectively.

Fourier analysis

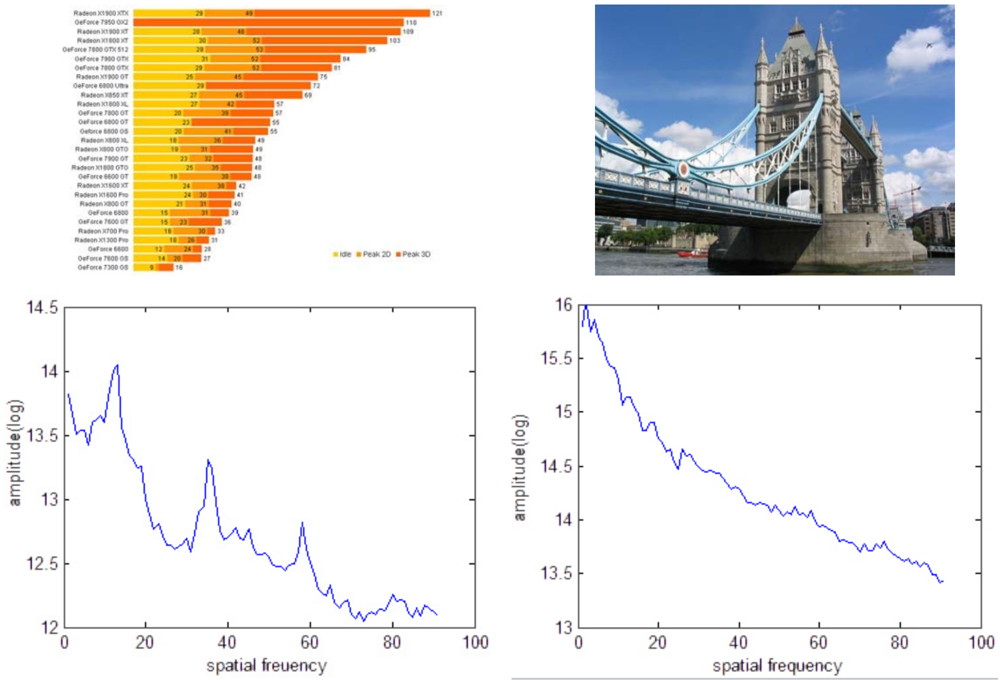
Since business graphics generally contain text, bar or pie charts, line graphs and sharp edges, etc., spike high frequency components in the frequency domain are presented due to the sharp change of neighboring pixels. Figure 4 illustrates the fall-off of amplitude with increasing spatial frequency averaged over all radial angles, which shows the generality of our finding. The amplitude falls off from the centre of each plot towards the edges as the spatial frequency increases. In the case of the graphical spectrum, the spectrum shows more evidence of periodicity and steep slope which stem from the fact that the main spatial feature in a graphical image is the sharp edge and text which give rise to specific and discrete frequencies in Fourier space. On the other hand, the fall-off of the pictorial image is more consistent and smooth.

A formula analogous to Equation 1 can then be used to provide a Fourier-frequency distribution (FFD) metric which is expressed as,

$$FFD = \frac{\sum (A_i - A_{i+1})^2}{\sum A_i^2} \tag{3}$$

where A_i is the amplitude of frequency range i, and the normalisation is used to remove the influence of image size.

According to Equation 3, a typical business graphic as shown in Figure 4 (a) gives a value of FFD=6.8; and the pictorial image in Figure 4 (b) gives a value of FFD=2.6. To classify images according to the FFD metric, it is expected that business graphics give high FFD values and pictorial images give low FFD values, thus a threshold can be defined for separation.



(a) *(b)*

Figure 4: Fourier amplitude (log) of: (a) a business graphic and (b) a pictorial image with respect to spatial frequency.

The workflow and its performance

The above analyses described some typical differences of image characteristics between pictorial images and business graphics. A series of investigations were conducted based on above three image characteristics by using a number of 40 images. These 40 images include 20 business graphics and 20 pictorial images. The business graphics are consisted of 10 graphics recommended by CIE TC8-03 [5] and 10 typical graphics supported by Samsung Advanced Institute of Technology. The 20 pictorial images cover a wide range of real world scenes including buildings, animals, sky, grass, vegetation, water, portraits, and night scenes. All selected images are in a state of sRGB.

Individual performance of each metric

Images were first converted from sRGB space to CIECAM02 Jab space following the guidelines of CIECAM02 [4] by using a Dell 21-inch LCD display characterization model according to ISO3664 [6]. Three metrics (LFD, AVGC, and FFD) were then calculated.

The results of LFD are shown in Figure 5. Business graphics are plotted as type 1 whilst pictorial images are plotted as type 0.

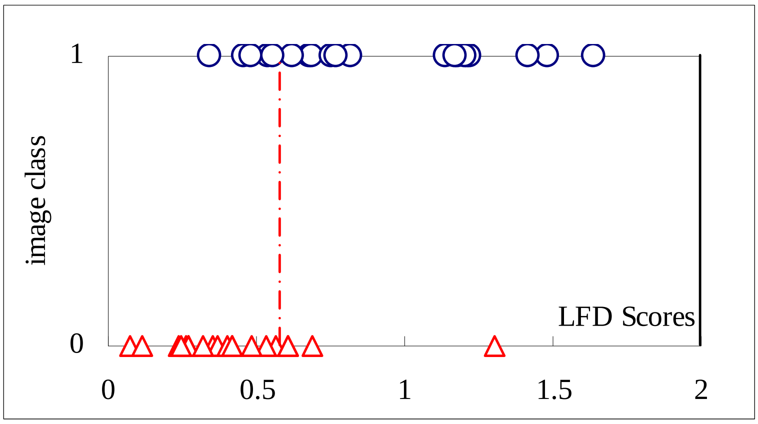


Figure 5: LFD scores for business graphics (type 1, circles) and pictorial images (type 0, triangles).

An optimised threshold value of 0.6 can be obtained by minimising the percentage of wrongly categorised images, which is presented as a dash dot line in Figure 5 for separating business graphics and pictorial images. If the value of LFD is greater or equal to this threshold, the image is deemed to be a graphical image. If the value of LFD is less than the threshold, the image is considered to be a pictorial image. The optimum threshold provides a performance of 80% on the set of 40 images.

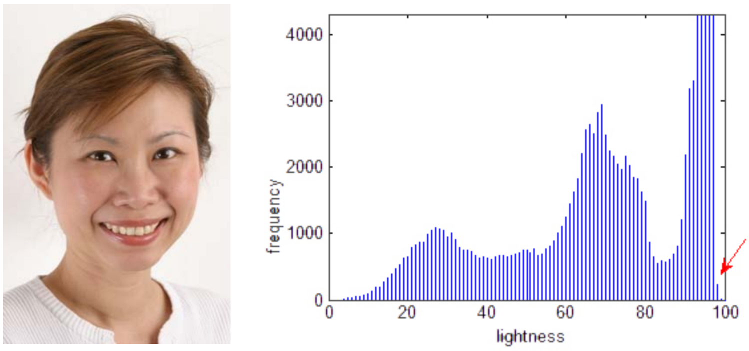


Figure 6: A pictorial image and its lightness frequency distribution.

An example of pictorial image which cannot be successfully discriminated by the LFD metric is a portrait as shown in Figure 6. This image produced a LFD value of 1.3 which was classified wrongly as a business graphic. By its lightness frequency distribution as shown in Figure 6, a steep decrease of frequency is found at high lightness values (indicated by the arrow) which results in a typical frequency variation of business graphics. This might be due to the unsmooth variation of white background and clothing from which a large proportion of the image is formed.

The average chroma scores are plotted against image types in Figure 7. Business graphics are plotted as type 1 whilst pictorial images are plotted as type 0. Again, it is evident that the discrimination is possible using this metric, which business graphics are tending to give larger scores than pictorial images. An optimum threshold of 20 (the dash dot line shown in Figure 7) results in the best performance of 65% which presents 26 images in the set of 40 images succeeding in the categorization. Full of dissaturated components in business graphics or mainly consisted of components with high chroma values in pictorial images are the primary reason for those images which are failed in the categorisation.

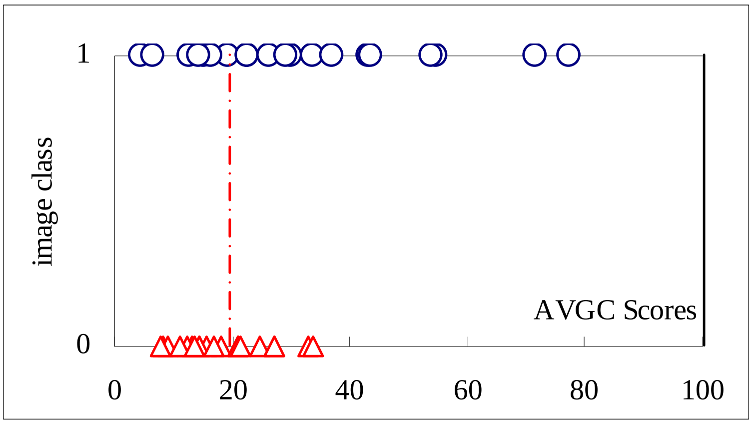


Figure 7: Average chroma scores for business graphics (type 1, circles) and pictorial images (type 0, triangles).

The results of Fourier frequency distribution for the set of 40 images are illustrated in Figure 8 and the optimum threshold at FFD=0.23 yields a performance of 67.5% which presents 27 images classified by the FFD metric correctly. However, there was a risk of incorrect classification about 40% when the FFD value was larger than the threshold, it can either be a pictorial image or business graphic.

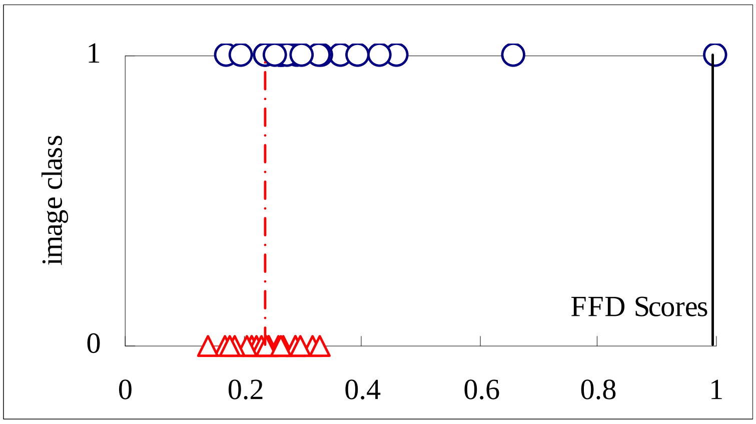


Figure 8: Fourier frequency distribution for business graphics (type 1, circles) and pictorial images (type 0, triangles).

Performances of pair-wise combinations of metrics

The performances above show that no single metric can give a satisfactory result in classifying a quite small image database. Each metric has its own strength and weakness.

Pair-wise combinations of metrics were then investigated in order for there to be any benefit or simple solution. The results are presented in Figure 9 with optimised threshold functions. Business graphics are plotted as type 1 whilst pictorial images are plotted as type 0. The performances of pair-wise combinations of LFD and AVGC, LFD and FFD, and AVGC and FFD, are 90.0%, 87.5%, and 72.5% respectively, which have been improved comparing with performances of individual metric.

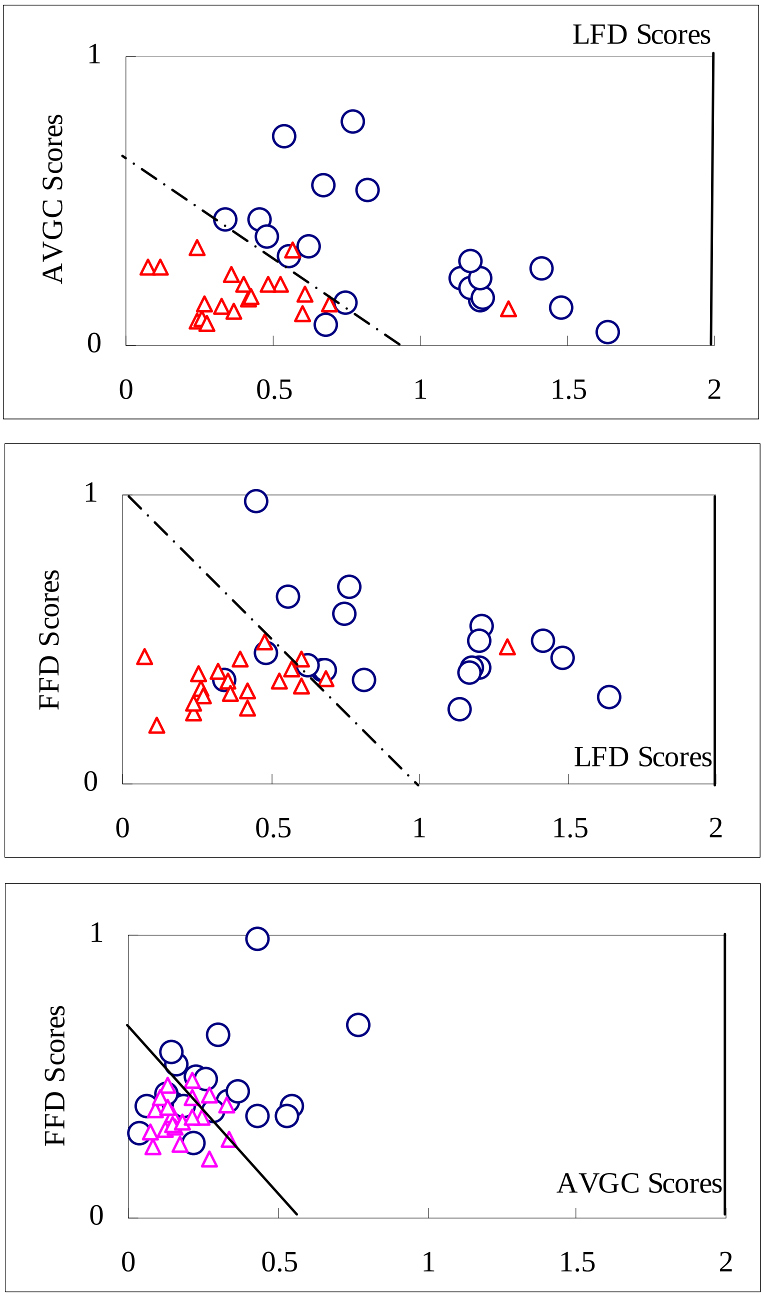


Figure 9: Performances of pair-wise combination of metrics for business graphics (type 1, circle) and pictorial images (type 0, triangle): metrics LFD and AVGC combination (top), metrics LFD and FFD combination (middle) and metrics AVGC and FFD combination (bottom).

Workflow and performance evaluations

In the above section, the performances of individual metric and pair-wise combinations have been discussed. Although a higher performance obtained from the combination of LFD and AVGC metrics, there are still rooms to be improved considering the requirement of post-processing. A further investigation was conducted based on the combination of 3 metrics, which turns out a performance of 92.5% as shown in Figure 10. Considering the testing dataset is somehow small, a workflow is structured by a neural network based on the combination of LFD, AVGC, and FFD metrics to improve the performance.

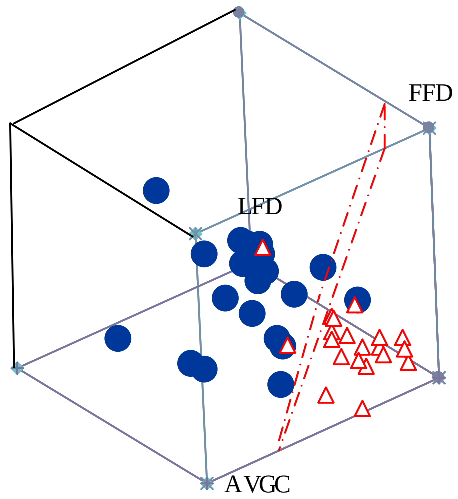


Figure 10: Performance in a 3-D space by combination of LFD-AVGC-FFD (shown with an optimised plain by dot dash line).

A multilayer perceptron (MLP) neural network was designed to improve the overall performance. The network was designed as a two-layer feed-forward network and trained for up to 1000 epochs to an error goal of 0.00001. The network has three inputs, LFD, AVGC and FFD, ranging from 0 to 1, following by the first layer of four tansig neurons and the second layer with one purelin neurons. The neural network was implemented using Matlab (neural network toolbox) by the following steps.

Step 1. Training metrics preparation

$$T = [LFD_{BG} \quad LFD_{PI}; \quad AVGC_{BG} \quad AVGC_{PI}; \quad FFD_{BG} \quad FFD_{PI}]$$

where LFD, AVGC and FFD represent three metrics obtained from training images; the subscribes BG and PI correspond to business graphics and pictorial images respectively. Note the results of three metrics had been normalized in the range of [0 1]. In a perfect situation, it is expected that metrics LFD, AVGC and FFD result in 1 for business graphics and 0 for pictorial images.

Step 2. Feed-forward Neural Network

$$net = newff([0 \quad 1; \quad 0 \quad 1; \quad 0 \quad 1], \quad [4 \quad 1], \quad \{'\tan sig' \quad 'purelin'\})$$

Step 3. Neural Network Training

$$net = train(net, T, t)$$

Based on the training results net, a threshold of 0.5 was selected, which a performance of 100% was reached on the training set of 40 images.

Thus, for each new test image, the network is simulated by:

$$Out = sim(net, \quad [LFD; \quad AVGC; \quad FFD])$$

where LFD, AVGC and FFD represent three metrics obtained from test image; net is obtained from step 3. If the output is equal to or larger than threshold 0.5, the test image is considered to be business graphic, otherwise pictorial image.

To test the overall performance of the workflow, an additional database of 57 images was used which included 26 business graphics and 31 pictorial images. The business graphics covered a wide range of graphical elements, such as bar chart, pie chart, line graph, flow chart, map, text, etc. The pictorial images mainly consisted of animals, portrait of Asian, African and European, indoor and outdoor scenes, night scenes, mountain, vegetation, food, and etc. The workflow results in an error of 5% which meant only 3 of 57 images were not judged correctly. These three images included one business graphic and two pictorial images as shown in Table 1.

Images	LFD scores	AVGC scores	FFD scores
 a	0.8	12	0.22
 b	0.6	35	0.23
 c	1	20	0.40
thresholds	0.6	20	0.23

Table 1: Wrongly classified images.

The analysis of these three images indicates that the incorrectly classified pictorial images have characteristics similar to business graphics to a great extent, such as the hard edge variation of stones in the scene of Stonehenge (c) gives a typical result of business graphical characteristics on FFD analysis. Also the Chinese building (b) captured on a high chroma background having shape of line and cycle results in a wrong judgment as a business graphic. Conversely, the wrongly classified business graphic (a) has low chroma values of background and text, and the main element of motorbike has somehow typical characteristics of pictorial images.

Comparing with the previous investigation, Figure 11 shows the improvement by the proposed neural network. Although the testing image database is limited, the accuracy of 95% of current workflow gives more confidence on the further processing of gamut mapping which will be, in turn, treated differently. One advantage of neural network is more training data provided more accurate on outputs. However, in the case of wrongly classified image, it will be applied unsuitable gamut mapping algorithm 100%. Thus, an accuracy of 100% is desirable for further research.

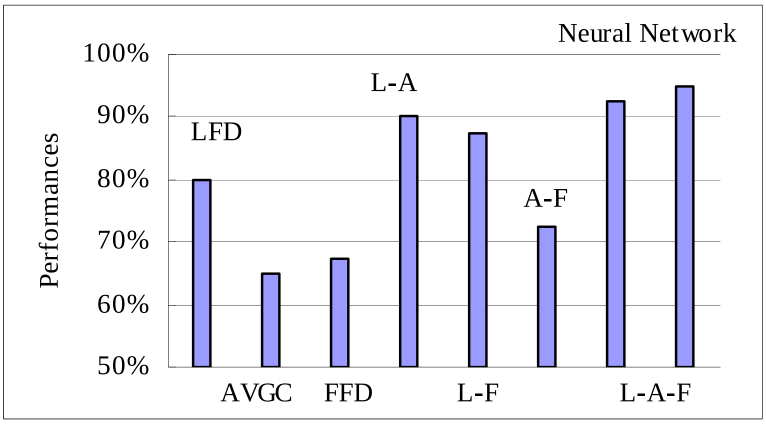


Figure 11: Comparison of performances between different combination of 3 metrics (L-A: LFD&AVGC; L-F: LFD&FFD; A-F: AVGC&FFD; L-A-F: LFD&AVGC&FFD).

Conclusions

Categorising images into business graphics and pictorial images is somehow a subjectively simple task according to their definitions, but a time consuming and hard job for industrial application. Due to the requirement of automatic colour image reproduction workflow, a computational image categorisation procedure has been presented and shown to be successful as a pre-processing step.

There are many existing techniques for measuring different image characteristics, but, according to the task, to date there is no satisfactory method appropriate for the colour image reproduction. The new workflow presented is based on chromatic and spatial information and this is clearly related to the human visual system. The new workflow is capable of categorising images into business graphics and pictorial images. However, more work is required in understanding the questioned image characteristics to bring forth an accuracy of 100% performance.

Acknowledgement

This work was supported by Hainan Provincial Scientific Collaboration Project (KJHZ2015-23), Major Science and Technology Project of Hainan Province (ZDKJ2016015), State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University.

References

1. Dennis AR (1988), The use of business graphics, *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, **19** (2), 17-28.

2. Luther D (1982), Business graphics: What is it?, *Proceedings of the 9th Annual Conference on Computer Graphics and Interactive Techniques,* 275.
3. Jorgensen C (1996), Investigation of pictorial image attributes in descriptive tasks, *Proceedings of SPIE Human Vision and Electronic Imaging*, **2657**, 241-251.
4. Moroney N, Fairchild MD, Hunt RWG, Li C, Luo MR and Newman T (2002), The CIECAM02 color appearance model, *Proceedings of the IS&T/SID 10th Color Imaging Conference*, 23-27, Scottsdale (USA).
5. CIE (2014), Guidelines for the evaluation of gamut mapping algorithms, *CIE Technical Report 156:2004*, Central Bureau of the CIE, Vienna (Austria).
6. ISO (2000), Viewing conditions – graphic technology and photography, *ISO 3664*, ISO Central Secretariat, Geneva, Switzerland).