DOCTOR OF PHILOSOPHY

Exploring the effectiveness of similarity-based visualisations for colour-based image retrieval

William Plant

2013

Aston University

Exploring the Effectiveness of Similarity-based Visualisations for Colour-based Image Retrieval

William Robert Plant

Doctor of Philosophy

Aston University
Birmingham

October 2012

Aston University
Birmingham

Exploring the Effectiveness of Similarity-based Visualisations for Colour-based Image Retrieval

William Robert Plant

Doctor of Philosophy, 2012

Thesis Summary

In April 2009, Google Images added a filter for narrowing search results by colour. Several other systems for searching image databases by colour were also released around this time. These *colour-based image retrieval* systems enable users to search image databases either by selecting colours from a graphical palette (i.e., query-by-colour), by drawing a representation of the colour layout sought (i.e., query-by-sketch), or both. It was comments left by readers of online articles describing these colour-based image retrieval systems that provided us with the inspiration for this research. We were surprised to learn that the underlying query-based technology used in colour-based image retrieval systems today remains remarkably similar to that of systems developed nearly two decades ago. Discovering this ageing retrieval approach, as well as uncovering a large user demographic requiring image search by colour, made us eager to research more effective approaches for colour-based image retrieval. In this thesis, we detail two user studies designed to compare the effectiveness of systems adopting *similarity-based visualisations*, query-based approaches, or a combination of both, for colour-based image retrieval. In contrast to query-based approaches, similarity-based visualisations display and arrange database images so that images with similar content are located closer together on screen than images with dissimilar content. This removes the need for queries, as users can instead visually explore the database using interactive navigation tools to retrieve images from the database. As we found existing evaluation approaches to be unreliable, we describe how we assessed and compared systems adopting similarity-based visualisations, query-based approaches, or both, meaningfully and systematically using our *Mosaic Test* - a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system under evaluation.

Keywords: Image Databases, Image Retrieval, Colour, Visualisation, Evaluation

Acknowledgements

I would like to acknowledge the contributions of my joint supervisors Dr. Joanna Lumsden and Prof. Ian Nabney to this research. The importance of the advice, time, and untold patience, they have invested into this research over the last three years cannot be overstated and is deeply appreciated. I would also like to thank Dr. Gerald Schaefer for giving me the opportunity to conduct this research and for his support throughout. Finally, I sincerely thank everyone that participated in the two user studies described in this thesis, many of which gave up their time for no financial reward. This research has been funded by the Aston University School of Engineering and Applied Science's EPSRC doctoral training grant.

Contents

1 Introduction **12**
1.1 Thesis Outline . . . 16

2 Background and Related Work **18**
2.1 Extracting and Comparing Colour Content . . . 18
2.1.1 RGB Colour Model . . . 19
2.1.1.1 Representations of the RGB Colour Model . . . 19
2.1.1.1.1 HSV . . . 19
2.1.1.1.2 YCbCr . . . 19
2.1.1.1.3 HMMD . . . 19
2.1.2 Colour-based Feature Vectors . . . 20
2.1.2.1 Colour Histograms . . . 20
2.1.2.2 Auto Colour Correlograms . . . 21
2.1.2.3 MPEG-7 Colour Descriptors . . . 22
2.1.2.3.1 MPEG-7 Scalable Colour Descriptor . . . 22
2.1.2.3.2 MPEG-7 Colour Structure Descriptor . . . 22
2.1.2.3.3 MPEG-7 Colour Layout Descriptor . . . 23
2.2 Query-by-colour Query Style . . . 24
2.3 Alternative Query Styles to Query-by-colour . . . 24
2.3.1 Query-by-example . . . 24
2.3.2 Query-by-sketch . . . 24
2.4 Users of Colour-based Image Retrieval Systems . . . 26
2.5 Similarity-based Visualisations . . . 27
2.5.1 Mapping-based Visualisations . . . 27
2.5.1.1 Dimension Reduction Techniques . . . 27
2.5.1.1.1 Principal Component Analysis . . . 27
2.5.1.1.2 Multi-dimensional Scaling . . . 28
2.5.1.1.2.1 Use of Multi-dimensional Scaling . . . 29
2.5.1.2 Overcoming Image Overlap . . . 30
2.5.1.2.1 Fitting Visualisations to a Regular Grid . . . 30
2.5.1.3 Mapping-based Visualisation Interactive Tools . . . 32
2.5.2 Clustering-based Visualisations . . . 33
2.5.2.1 Uses of Hierarchical Clustering . . . 33
2.5.2.2 Selection of Representative Images . . . 35
2.5.3 Graph-based Visualisations . . . 36
2.5.3.1 Pathfinder Networks . . . 36
2.5.3.2 NN^k Networks . . . 37
2.5.4 Comparison of Similarity-based Visualisation Styles . . . 37
2.6 Colour-based Image Retrieval Evaluations . . . 37
2.6.1 Estimating Colour Proportions . . . 38
2.6.2 Graphical Colour Palette Design . . . 38
2.6.3 Evaluating the Query-by-colour Query Style . . . 39
2.7 Image Retrieval Evaluation Approaches . . . 39
2.7.1 Drawbacks for Colour-based Image Retrieval . . . 41
2.8 Summary and Research Aims . . . 41

2.8.1 Research Aims 43

3 The Mosaic Test **44**
3.1 Introduction 44
3.1.1 Chapter Outline 45
3.2 The Mosaic Test 45
3.2.1 The Mosaic Test Tool 45
3.2.1.1 Displaying the Mosaic Test Tool 46
3.2.1.2 Creating an Image Mosaic 46
3.2.1.3 Enabling Reflection-in-action 47
3.2.1.4 Completing a Mosaic Test 49
3.2.2 Evaluating Colour-based Image Retrieval Systems 50
3.2.2.1 Automatically Assessing Image Mosaic Accuracy 51
3.2.2.1.1 Tran 51
3.2.2.1.2 Nakade and Karule 51
3.2.2.1.3 Zhang et al. 52
3.2.2.1.4 Drawbacks of Existing Mosaic Accuracy Measures 52
3.3 User Study 53
3.3.1 Participants 53
3.3.2 Procedure 54
3.3.2.1 Image Database 54
3.3.2.2 Colour-based Image Retrieval Systems 54
3.3.2.2.1 Query-by-colour System 55
3.3.2.2.2 Zoomable Mapping-based Visualisation System 56
3.3.2.2.3 Pannable-only Mapping-based Visualisation System 56
3.3.2.3 Target Images 57
3.3.2.4 Prior to the Mosaic Tests 57
3.3.2.5 The Mosaic Tests 58
3.3.2.6 After the Mosaic Tests 59
3.4 Results 59
3.4.1 Automatic Mosaic Accuracy Measures 59
3.4.2 Reflection-in-action Observations 60
3.4.3 User Expertise Comparison 61
3.4.3.1 Time Comparison 62
3.4.3.2 Mosaic Accuracy Comparison 63
3.4.3.3 Workload Comparison 64
3.5 Discussion 65
3.5.1 Automatically Measuring Mosaic Accuracy 65
3.5.2 Reflection-in-action 65
3.5.3 User Expertise Comparison 65
3.5.4 Limitations of the Mosaic Test 66
3.6 Conclusion 67

4 Query-by-color vs. Map-based Vis. for Colour-based Retrieval **68**
4.1 Introduction 68
4.1.1 Chapter Outline 69
4.2 Image Database 69
4.3 Query-by-colour System 70
4.4 Zoomable Mapping-based Visualisation System 71
4.5 Pannable-only Mapping-based Visualisation System 72
4.6 Hypothesis 73
4.7 Results 74
4.7.1 Time 74
4.7.2 Image Mosaic Accuracy 76
4.7.3 Workload 77
4.7.3.1 Mental Demand 78
4.7.3.2 Effort 79

4.7.3.3 Frustration . . . 80
4.7.4 User Preference Ratings . . . 81
4.8 Discussion . . . 82
4.8.1 Time . . . 82
4.8.2 Image Mosaic Accuracy . . . 82
4.8.3 Workload . . . 82
4.8.4 Preferred Approach for Colour-based Image Retrieval . . . 83
4.8.5 Zooming Preference . . . 83
4.9 Conclusion . . . 84

5 Comparing Colour-based Image Ret. Query Result Visualisations 86
5.1 Introduction . . . 86
5.1.1 Chapter Outline . . . 87
5.2 User Study . . . 88
5.2.1 Hypotheses . . . 88
5.2.2 Participants . . . 88
5.2.3 Image Database . . . 89
5.2.4 Tested Colour-based Image Retrieval Systems . . . 89
5.2.4.1 Query-by-colour Panel . . . 91
5.2.4.2 Query-by-sketch Panel . . . 92
5.2.4.3 Query Result Visualisation Area . . . 94
5.2.4.3.1 Ranked List . . . 94
5.2.4.3.2 Mapping-based Visualisation . . . 94
5.2.4.3.2.1 Controls . . . 95
5.2.4.3.3 Clustering-based Visualisation . . . 95
5.2.4.3.3.1 Controls . . . 97
5.2.5 Procedure . . . 97
5.2.5.1 Prior to the Mosaic Tests . . . 98
5.2.5.2 The Mosaic Tests . . . 98
5.2.5.3 After the Mosaic Tests . . . 99
5.3 Results . . . 99
5.3.1 Query-by-colour and Query-by-sketch Query Styles . . . 100
5.3.1.1 Time . . . 100
5.3.1.2 Image Mosaic Accuracy . . . 101
5.3.1.3 Workload . . . 102
5.3.1.3.1 Physical Demand . . . 103
5.3.1.3.2 Performance . . . 104
5.3.1.3.3 Frustration . . . 105
5.3.2 Presentation of Query Results . . . 106
5.3.2.1 Time . . . 106
5.3.2.2 Image Mosaic Accuracy . . . 107
5.3.2.3 Workload . . . 108
5.3.2.4 Preference . . . 109
5.4 Discussion . . . 110
5.4.1 Query-by-colour and Query-by-sketch Query Styles . . . 110
5.4.2 Presentation of Query Results . . . 111
5.5 Conclusion . . . 111

6 Conclusions 114
6.1 Summary of Main Findings . . . 115
6.2 Future Work . . . 118
6.2.1 Increasing the Number of Mosaic Test Participants . . . 118
6.2.1.1 Crowd Sourcing . . . 119
6.2.1.2 Mobile Application . . . 119
6.2.2 Incorporating Keyword Search into the Mosaic Test . . . 120
6.2.3 Similarity-based Visualisation Variations . . . 120
6.2.4 The Query-by-colour and Query-by-sketch Query Styles . . . 120

Appendices 132

A Mosaic Test Documents 132

B First Study System Instructions 143

C Image Mosaic Measure Accuracy Rankings 147

D Second Study Query-by-colour Systems Instructions 149

E Second Study Query-by-sketch Systems Instructions 154

F User Study Consent Forms 159

List of Figures

2.1 Two visually dissimilar graphics sharing the same colour histogram [39]. 21
2.2 A graphic showing the `sliding window' approach of the MPEG-7 Colour Structure Descriptor (left) and an image of a telephone box partitioned into 64 (8×8) blocks represented in the MPEG-7 Colour Layout Descriptor (right). 22
2.3 Screenshot of a colour-based image retrieval system which adopts the query-by-colour query style. 23
2.4 Screenshot of the IMGSEEK image retrieval system [8] which adopts the query-by-example query style. 25
2.5 Screenshot of the image retrieval system presented by Jacobs *et al.* [47], adopting the query-by-sketch query style. 25
2.6 Visualisation of the UCID image database [83] generated using multi-dimensional scaling. Many of the images have been overlapped, or completely occluded, by other images in the visualisation. 30
2.7 Visualisation of 100 images generated using multi-dimensional scaling (left), fitted to a regular grid (right) [73]. 31
2.8 Spreading strategies proposed by Rodden *et al.* [73]. Dashed arrows represent the spiral scan, whilst solid arrows depict the path of an image. 31
2.9 Screenshot of Borth *et al's.* Navidgator system [7]. A user has traversed the hierarchy to cluster of video stills captured from the same news broadcast. 34
2.10 The evening strategy of Schaefer and Ruszala [82]. Images (represented as black dots) have been moved from the cells neighbouring the vacant cell to produce a more even distribution of images. 35
2.11 A graph-based visualisation of database images [49]. 36

3.1 Screenshot of the Mosaic Test Tool (outlined in red) and a colour-based image retrieval system (right) under evaluation during a Mosaic Test session. The five cells across the top, and the two cells in the bottom right hand corner, of the image mosaic have been filled using images from the database. 46
3.2 Screenshot of a Mosaic Test participant attempting to fill an image mosaic cell. 47
3.3 Screenshots of the initial Mosaic Test Tool with which users could zoom and pan around the entire target and mosaic images. 48
3.4 Screenshots of the revised Mosaic Test Tool showing only the mosaic cell to be completed and supporting just two zoom levels; *half-size* (Figure 3.4a) and *full-size* (Figure 3.4b). 49
3.5 Two image mosaics of varying accuracy in comparison with the original target image. 51
3.6 Screenshot of the query-by-colour system tested as part of our first user study. 55
3.7 Screenshots of the zoomable and pannable-only mapping-based visualisation systems tested in the first user study 56
3.8 Target images recreated as image mosaics by participants of the study. Each target image is predominantly comprised of three jelly beans of varying colour and orientation. 57
3.9 Screenshot a Mosaic Test practice session. The Mosaic Test Tool (left) is displayed simultaneously with a selection of small selection of suitable images (right) so that participants can practise creating an image mosaic. 58
3.10 Box-plots of the times required by study participants in the `expert' and `non-expert' user groups to complete their image mosaics. 62

3.11 Box-plots of the accuracy of image mosaics (i.e., the L_1 distance between the MPEG-7 Colour Structure descriptors of generated mosaics and their corresponding target images) created by participants in the `expert` and `non-expert` user groups using the three colour-based image retrieval systems. 63
3.12 Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants in the `expert` and `non-expert` user groups whilst using the three colour-based image retrieval systems to create their image mosaics. 64

4.1 Graphical colour palette of the MultiColr system [42], comprising the 120 colours used to index the MIR-FLICKR 25000 image collection [40] with 120 bin colour histograms. 69
4.2 Screenshot of the query-by-colour system. The system has returned the database images which contain the most pixels with the user's selected cyan colour. 70
4.3 Screenshots of the zoomable mapping-based visualisation system before and after a ``zoom in'' operation has been applied. 72
4.4 Screenshot of the pannable-only mapping-based visualisation system. 73
4.5 Box-plots of the time required by participants to complete an image mosaic using each of the three systems tested in our first user study with and without outliers. 75
4.6 Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants of our first user study with each of the tested systems, and their corresponding target images. 76
4.7 Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants of our first study when using the three colour-based image retrieval systems to create their image mosaics. 77
4.8 Box-plots of the mental demand ratings assigned by participants of our first study using the three colour-based image retrieval systems to create their image mosaics. 78
4.9 Box-plots of the effort ratings assigned by participants of our first user study using the three colour-based image retrieval systems to create their image mosaics. 79
4.10 Box-plots of the frustration experienced by participants of our first study when using the three colour-based image retrieval systems to create their image mosaics. 80
4.11 Box-plots of the preference ratings assigned to each of the three colour-based image retrieval systems by participants for creating their image mosaics. 81

5.1 Fundamental interface design of the colour-based image retrieval systems tested in the second user study, comprising a *query panel* (highlighted in green) and a *query result visualisation area* (highlighted in blue). 90
5.2 Screenshot of the query panel of query-by-colour systems tested in the second user study. 91
5.3 Screenshot of the query panel of query-by-sketch systems. The user has filled the bottom 24 cells of their query sketch in green and the upper 40 cells using blue. . . . 92
5.4 Screenshot of query results in a ranked list visualisation. Users browse query results either by scrolling the scroll bar (highlighted in green) or via the mouse wheel. 93
5.5 Screenshots of query results in a mapping-based visualisation. The toolbar with zoom and pan controls is highlighted in green. 94
5.6 Screenshots of query results in a clustering-based visualisation. The toolbar with navigation controls and browsing history is highlighted in green. 96
5.7 The three target images to be recreated as image mosaics by participants of the study. Each target image is predominantly comprised of three jelly beans of varying colour and orientation. 98
5.8 Box-plots of the times required by study participants, using either query-by-colour or query-by-sketch systems, to complete an image mosaic. 100
5.9 Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants using either query-by-colour or query-by-sketch systems, and their corresponding target images. 101
5.10 Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales), experienced by participants using either the query-by-colour or query-by-sketch systems, to create their image mosaics. 102

5.11 Box-plots of the physical demand ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics. 103
5.12 Box-plots of the performance ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics. 104
5.13 Box-plots of the frustration ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics. 105
5.14 Box-plots of the times required by study participants to complete an image mosaics using each of the query result visualisations. 106
5.15 Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants using each of the query result visualisations, and their corresponding target images. 107
5.16 Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants using each of the query result presentations to create their image mosaics. 108
5.17 Box-plots of the preference ratings assigned to each of the query result presentations by participants for creating their image mosaics. 109

A.1 Pre-study questionnaire. 133
A.2 Mosaic Test Tool instructions (*page one*). 134
A.3 Mosaic Test Tool instructions (*page two*). 135
A.4 Mosaic Test Tool instructions (*page three*). 136
A.5 Mosaic Test Tool instructions (*page four*). 137
A.6 Mosaic Test Tool instructions (*page five*). 138
A.7 Mosaic Test Tool `training tasks' check sheet. 139
A.8 NASA-TLX workload assessment [31] background information. 140
A.9 NASA-TLX workload assessment [31]. 141
A.10 System preference assessment. 142

B.1 Zoomable mapping-based visualisation system training instructions. 144
B.2 Pannable-only mapping-based visualisation system training instructions. 145
B.3 Query-by-colour system training instructions. 146

D.1 Query-by-colour query panel instructions. 150
D.2 Ranked list query-by-colour system instructions. 151
D.3 Mapping-based visualisation query-by-colour system instructions. 152
D.4 Clustering-based visualisation query-by-colour system instructions. 153

E.1 Query-by-sketch query panel instructions. 155
E.2 Ranked list query-by-sketch system instructions. 156
E.3 Mapping-based visualisation query-by-sketch system instructions. 157
E.4 Clustering-based visualisation query-by-sketch system instructions. 158

F.1 First study consent form (*page one*). 160
F.2 First study consent form (*page two*). 161
F.3 First study consent form (*page three*). 162
F.4 Second study consent form (*page one*). 163
F.5 Second study consent form (*page two*). 164
F.6 Second study consent form (*page three*). 165

List of Tables

3.1 Spearman's rank correlation coefficients (r_s) between the image mosaic distance rankings made by humans and the 10 most strongly correlated rankings generated by the tested image mosaic accuracy measures. 60

C.1 Spearman's rank correlation coefficients (r_s) between the image mosaic distance rankings made by participants and the rankings generated by the tested image mosaic accuracy measures. 148

Chapter 1

Introduction

In April 2009, Google Images [44] added a filter for narrowing search results by colour. Writing on the company's official blog, Tanguay [94] (a Software Engineer at Google) reported that by clicking one of twelve colours located on the newly added graphical palette, users would be presented with images from their keyword query result set that contained a high proportion of the selected colour. The following year, a developer at Microsoft followed suit - with Hua [100] revealing that Bing Images [17] now had a similar *query-by-colour* facility. Furthermore, Hua also showcased Bing Images' new *query-by-sketch* functionality; enabling users to search for images by drawing a representation of the colour layout sought. Several other systems facilitating *colour-based image retrieval*, such as Chromatik [22] and MultiColr [42], were also released around this time and the web was soon inundated with online articles (e.g., [26, 18, 85]) describing various systems for searching image databases by colour. It was the comments left by readers of these articles that provided us with the inspiration for this research.

We found two prevalent questions posted by readers of these online articles; *how do colour-based image retrieval systems work? who would use them?* As we researched the answer to the first of these questions, we found that much research has been conducted over the past two decades into retrieving images from databases according to content (a research domain often referred to as *content-based image retrieval* [89, 19]). Colour-based image retrieval systems extract the colour content of database images directly from the pixel data and encapsulate a summary of this information using *n-dimensional feature vectors* - that is, a compact numerical representation of colour content is extracted for each image in the database. Colour-based image retrieval systems use these feature vectors to index database images, with most enabling users to query the database by selecting colours from a graphical colour palette (for query-by-colour), by drawing a representation of their requirements (for query-by-sketch), or both. A *query feature vector* is extracted automatically from a user's query and an inter-vector distance between this query feature vector and the feature vectors of all database images is then computed. The database images with a feature vector close (i.e., a short inter-vector distance) to the query feature vector are then returned and displayed to the user.

It surprised us to learn that the underlying query-based technology used for colour-based image retrieval in Google Images and Bing Images was not as innovative as we expected and, is in fact, based on a retrieval model devised almost two decades ago [23]. We found this to be somewhat of a revelation given that image databases have grown exponentially over the last twenty years and can now contain millions of images [68]. We were also surprised at the lack of literature available regarding the user demographic of colour-based image retrieval systems. From the little previous research available, coupled with our consultations with graphic designers, we deduced that individuals operating in creative industries such as graphic, fashion and interior design, search image databases by colour. Tanguay, the Software Engineer at Google, also revealed that ``a very large number'' of users invoke the colour filter in Google Images each day. Discovering this ageing retrieval approach and large user demographic made us eager to research more effective approaches for colour-based image retrieval.

It was whilst investigating how colour-based image retrieval systems work that we found a substantial body of research regarding *similarity-based visualisations* for image retrieval. In contrast to query-based approaches, similarity-based visualisations display and arrange database images so that images with similar content are located closer together than images with dissimilar content [32]. This removes the need for queries, because users can instead visually explore the database using interactive navigation tools to retrieve images of interest from the database. For this research, we focussed on two distinct styles of similarity-based visualisation: *mapping-based* and *clustering-based* visualisations. *Mapping-based* visualisations use dimensionality-reduction techniques to map the often high-dimensional image feature vectors to just 2-dimensions, used as co-ordinates for plotting database images on screen. Users can then explore the resultant database visualisations via zooming and panning tools. *Clustering-based* visualisations create groups of images with similar feature vectors where the group is visualised using a single image (or, in some cases, a small set of images) representative of the images contained within the group. Image groups can contain sub-groups of images, which themselves may also harbour further sub-groups of images, and so on. Users can explore this hierarchical structure by navigating through groups of interest using an interface similar to that of file browsers found in common operating systems (i.e., the representative images of groups act much like folders, with individual images becoming analogous to files). Since we were unable to find any evidence of similarity-based visualisations being evaluated previously for colour-based image retrieval, coupled with our finding that no research existed directly comparing the existing query-by-colour and query-by-sketch query styles for colour-based image retrieval, we formulated the research aim; *to evaluate and compare the effectiveness of query-based approaches (i.e., the query-by-colour and query-by-sketch query styles) and similarity-based visualisations for colour-based image retrieval*.

Further to the above finding, we found two other aspects of similarity-based visualisations that we believed warranted further investigation. First, many systems adopting mapping-based visualisations for image retrieval include a zooming facility for exploring databases. There has, however, been just one previous study into the importance of zooming in these systems - that is, the study of Combs and Bederson [16]. Participants were asked to retrieve target images from a database of just 225 images using two systems; one which permitted zooming and another that did not. As the database used in the study was so small (way short of the magnitude of collections these days), half of the participants failed to perform a single zoom operation when testing the zoomable system, instead reviewing database images at a low resolution (i.e., as small thumbnails). Consequently, Combs and Bederson found no significance difference in the times required by users to retrieve specific images using the two systems. Given that these users would have been viewing such small images, it was perhaps unsurprising to learn that the error count for the zoomable system - that is, the number of images retrieved by users that were not identical to the image they were asked to find - was found to be higher than for the non-zoomable counterpart. As we believed the importance of zooming for image retrieval would correlate directly with the number of images being visualised (i.e., the more images that are displayed, the more effective zooming will become), we decided to make an additional aim for this research; *to assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases*.

The second aspect of similarity-based visualisations that we believed warranted further investigation was their potential for use in presenting the results of colour-based image retrieval queries. Liu *et al.* [55] undertook a user study comparing three image retrieval systems which presented keyword query results either in the form of a ranked list (where images ranked by the system as most similar to the query appear at the head of the list), a mapping-based visualisation, or a clustering-based visualisation. Test users were given a set of 17 query terms and asked to search for images in the test database which best matched each of the pre-determined query terms. It was observed that users could find images fastest using the systems presenting query results in the form of mapping-based and clustering-based visualisations (as opposed to the ranked list presentation). Whilst the search times recorded for the systems adopting mapping and clustering-based visualisations for presenting query results were not found to be significantly different, users did assign higher preference ratings to the mapping-based visualisation system - claiming it to be more intuitive, interesting, and convenient than the ranked list and clustering-based visualisation systems for searching and comparing images. Since Google Images and Bing Images currently adopt a ranked list visualisation for presenting both keyword *and* colour query results, we wanted to evaluate what effect changing the visualisation of results would have on the large number of users searching by colour (as opposed to keywords as per Liu *et al.*'s investigation). Consequently, we added another aim for our research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*.

In order for us to fulfil the research aims stated above, our first task was to overcome the lack of a reliable method for systematically and meaningfully evaluating colour-based image retrieval systems. Our background research revealed that there have been very few reported evaluations of colour-based image retrieval systems. Since the field of image retrieval lacks a standardised approach to evaluation [64], it has been very difficult for individuals, be it researchers or end-users, to reliably assess and compare the effectiveness of colour-based image retrieval systems or approaches. One of the few colour-based image retrieval systems that has been evaluated previously is QBIC (query by image content). For Faloutsos *et al.'s* [23] evaluation of the query-by-colour facility in the QBIC system, test users were supplied with 10 target images and asked to mark each image in a database (of approximately 1,000 images) as either relevant or irrelevant with respect to each supplied target image. The participating users were then asked to submit 10 queries (one per a target image) to the QBIC system that they believed would be sufficient to retrieve the target images from the database. Finally, Faloutsos *et al.* used the number of `relevant' images returned by the system to calculate the effectiveness of QBIC.

There are three fundamental drawbacks to Faloutsos *et al.*'s evaluation approach which disqualify it from being adopted as a standard for the field. Firstly, asking human judges to manually label all images in terms of their relevance to a target image is a very arduous and time consuming task (especially given that the magnitude of image databases has vastly outgrown 1,000 images since Faloutsos *et al.*'s study was conducted). Second, this list of `correct' images - often referred to as a ground-truth [89] - can be highly subjective; one person's perception of image similarity can be entirely different to someone else. The third drawback of this evaluation approach is that it does not accurately reflect the manner in which highly creative users (e.g., graphic, fashion and interior designers) assess the suitability of images for use in their design artefacts. These users typically determine the compatibility of images within the context of the project for which the retrieved image is intended to be used. For example, graphic designers add candidate images to web page designs to assess their appropriateness. Graphic designers will then decide whether to keep their chosen image, or, discard the image and refine their search criteria according to its shortcomings. This iterative image selection process is known formally as *reflection-in-action* [84].

As a solution, we devised the *Mosaic Test*. The Mosaic Test is a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system under evaluation. The time and users' perception of the workload required to complete this creative task, as well as the visual accuracy of their image mosaics (in comparison with the initial target images), are used to assess the effectiveness of the system being tested. Since our Mosaic Test adopts a standardised image database, automatically assesses the relevance of selected images by measuring the accuracy of generated mosaics (thus removing the need for a ground-truth), and enables users to reflect on the suitability of images for use in their mosaics, we believed that the Mosaic Test would provide a reliable mechanism by which to meaningfully and systematically evaluate and compare systems for colour-based image retrieval. To substantiate this claim, we aimed; *to show that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems*.

In the following Section, we outline the structure of this thesis and introduce how we adopted our Mosaic Test for two user studies *exploring the effectiveness of similarity-based visualisations for colour-based Image retrieval*.

1.1 Thesis Outline

In Chapter 2, we review the current state of colour-based image retrieval research. We describe the demographic of users retrieving images on the basis of colour and how the colour content of database images can be extracted and summarised for use in colour-based image retrieval systems. We then provide a comprehensive review of similarity-based visualisations adopted for image retrieval. We also review the small number colour-based image retrieval system evaluations that have been conducted previously, before discussing the unsuitability of existing image retrieval evaluation strategies for assessing colour-based image retrieval systems and approaches. Finally, we summarise our main findings from our background research and reiterate our research aims.

In Chapter 3, we describe how we fulfilled our research aim *to show that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems*. We present findings from our first user study which support our claim that the Mosaic Test overcomes the drawbacks occurring in existing evaluation approaches.

In Chapter 4, we describe the system comparison aspect of our first user study. We describe how 24 participants were recruited for evaluating three colour-based image retrieval systems using our Mosaic Test. The first of these systems was based on the MultiColr colour-based image retrieval system [42] (i.e., it featured the query-by-colour query style). The second system adopted a mapping-based visualisation for colour-based image retrieval (recall that *mapping-based* is a style of similarity-based visualisation) which users could interactively explore using zooming and panning tools. These two systems were included in the study so that we could fulfil part of our research aim *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*; by directly comparing the effectiveness of the query-by-colour query style with a mapping-based visualisation. The third colour-based image retrieval system tested in our first user study featured the same mapping-based visualisation as the second system. For this third system, however, we removed the zoom facility so that users could only pan the visualisation to search for images. Removing this functionality from the third system enabled us to meet another of our research aims; *to assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases*.

In Chapter 5, we detail our second user study to which we recruited 36 participants to compare six colour-based image retrieval systems using the Mosaic Test. Participants in our second user study were divided into two equal groups, the first of which created image mosaics using three systems adopting the query-by-colour query style, whilst the second group generated their mosaics using three systems with the query-by-sketch query style. Grouping test-users in this way enabled us to achieve another aspect of the research aim; *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*; by directly comparing the effectiveness of the query-by-sketch and query-by-colour query styles. Given that many colour-based image retrieval systems typically support either the query-by-colour query style *or* the query-by-sketch query style [47, 42, 22, 44], but rarely both, and that no previous research had been conducted directly comparing the query styles, we wanted to provide designers creating future systems with guidance regarding which of the query styles must be included to best support users undertaking colour-based image retrieval.

Within the two user groups in our second study, the three colour-based image retrieval systems tested by users visualised query results as either a ranked list (i.e., in descending order of closeness to the submitted query), a mapping-based visualisation, or as a clustering-based visualisation. This satisfied the final aim of our research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*.

We outline the contributions to knowledge made by this research and provide a summary of the findings from our two user studies in Chapter 6. As a further contribution, we propose several avenues for future work found via our research which, if conducted, could further advance the field towards more effective colour-based image retrieval.

Chapter 2

Background and Related Work

In this Chapter, we review the current state of colour-based image retrieval research. We begin with Sections 2.1, 2.2 and 2.3, describing how the colour content of database images can be extracted and summarised as feature vectors for use in systems adopting query-based approaches for colour-based image retrieval. In Section 2.4, we identify the demographic of users retrieving images on the basis of colour. In Section 2.5 we review similarity-based visualisations for image retrieval. We then turn our attention to previous colour-based image retrieval research in Section 2.6, reviewing the small number of system evaluations that have been conducted in the field. In Section 2.7, we discuss why current image retrieval evaluation strategies are not suited for assessing colour-based image retrieval systems. Finally, in Section 2.8, we provide a summary of our main findings and reiterate our research aims; to show that the Mosaic Test provides a reliable mechanism by which to meaningfully and systematically evaluate and compare colour-based image retrieval systems; to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval; assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases; and finally, to compare the effectiveness of colour-based image retrieval query result visualisations.

2.1 Extracting and Comparing Colour Content

It is clear that, in order to retrieve images based on colour content, an important issue for colour-based image retrieval system developers to address is how that colour content should be represented: how precise the description of colours needs to be, what threshold should be used to determine the presence of a colour, and whether the location of colour is relevant. The other key question is how the similarity between colour representations should be measured. In this Section, we discuss several previously adopted methods for extracting, summarising and comparing the colour content of images.

2.1.1 RGB Colour Model

A *colour model* describes how colours can be represented numerically, typically using three or four values. The most commonly used colour model used for rendering imagery digitally is the RGB (red, green, blue) colour model. Using the RGB colour model for digital images, each pixel comprises a red, green and blue component which, when added together, produce an overall pixel colour. In the RGB colour model, black is produced by the absence of any red, green and blue, whilst white is generated when each of the red, green and blue channels are at their maximum.

2.1.1.1 Representations of the RGB Colour Model

Whilst the RGB colour model is ideal for rendering imagery on digital devices, it is less suitable for use in computer vision since the distances between colours do not correspond with the way in which humans perceive colour [91]. In short, our colour vision is produced using a combination of cones and rods located on the retina which are responsible for detecting nuances in colour *and* light respectively, and our vision is more sensitive changes in light (i.e., to the black-and-white information) than colour.

Consequently, several alternative representations of points in the RGB colour model have been devised for use in computer vision applications (such as graphical colour selectors, video transmission, and image retrieval) which *do* account for the way humans perceive colour; such as the HSV, YCrCb and HMMD representations. These representations are often referred to as *colour spaces*.

2.1.1.1.1 HSV (hue, saturation and value) is a cylindrical representation of the RGB colour model which non-uniformly follows the manner in which human vision perceives colour variations. This representation has three components; hue (a dominant wavelength producing red, yellow, blue or green), saturation (the purity of the colour, i.e. the standard deviation about the standard wavelength), and value (the amount of brightness or white in that colour) [91]. This representation is described as cylindrical since hue follows the circumference, saturation is the point along the radius, and value is the height, of the cylinder.

2.1.1.1.2 YCbCr represents colour as brightness and two colour difference signals. In YCbCr, the Y is the brightness (luma), Cb is blue minus luma (B-Y) and Cr is red minus luma (R-Y). YCbCr was originally designed to transmit analogue RGB video signals more efficiently. The YCbCr representation is used in the MPEG-7 colour structure descriptor [59] (described later in Section 2.1.2.3.3).

2.1.1.1.3 HMMD (hue, max, min, diff) is, as per HSV, a cylindrical representation of the RGB colour model. In HMMD, hue is as per the HSV representation, whilst the max and min components are the maximum and minimum values among the R, G, and B channels, respectively. The diff component is defined as the difference between the max and min. The HMMD representation is used in the MPEG-7 colour structure descriptor [59] (described later in Section 2.1.2.3.2).

2.1.2 Colour-based Feature Vectors

We have previously described how *feature vectors*, sometimes referred to as a *descriptors*, are extracted from images to form a compact numerical summary of an image's content. In this Section we we introduce feature vectors designed specifically for summarising the *colour content* of images.

2.1.2.1 Colour Histograms

Colour histograms are the most widely adopted feature vector for summarising the colours contained within database images. The colour histogram H for a given image I can be formally defined as;

$$H_I = [B_1, B_2, ..., B_j, ..., B_n] \tag{2.1}$$

where n is the number of distinct colours in the colour space and B_j is the histogram bin containing the number of pixels in image I that are of colour j [93]. These bin values are typically normalised by dividing the corresponding pixel count by the total number of pixels in image I. Since pixels in a typical 24-bit RGB image can be any one of 16,777,216 unique colours (256 red $\times$ 256 green $\times$ 256 blue shades), the processing and storage of colour histograms containing 16,777,216 bins, for all images in a database, would clearly be unmanageable. *Colour quantisation* [93] is therefore undertaken to reduce the total number of colours in a colour space, thus reducing the number of bins required for histograms to boost storage and processing efficiency. For content-based image retrieval, colour quantisation is typically undertaken to reduce colour spaces to 64,128 or 256 colours [59].

$$L_1(A, B) = \sum_{i=1}^{n} |A_i - B_i| \tag{2.2}$$

$$L_2(A, B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \tag{2.3}$$

A variety of distance metrics have been proposed for comparing the colour histograms of images within a database. A shorter distance between two colour histograms indicates a higher degree of similarity between two images. The most common distance measures are the L_1 (city-block) and L_2 (Euclidean) distance metrics [86, 92], defined formally in Equations (2.2) and (2.3) respectively. The L_1 and L_2 distance metrics compare the *corresponding* bins of colour histograms - i.e., the histogram bin containing the pixel count for the i^{th} colour in histogram A is only compared to the bin for the i^{th} colour in histogram B. The L_1 distance between two colour histograms is the absolute sum of the differences between corresponding histogram bins. The L_2 metric considers the histograms as points (or Cartesian co-ordinates) in an n-dimensional space (where n is the histogram length) and calculates the straight-line distance between them.

$$d_{quad}(A, B) = \sum_{i}^{n} \sum_{j}^{n} M_{ij}(A_i - B_i)(A_j - B_j) \tag{2.4}$$

Other distance metrics, such as the quadratic distance metric shown in Equation 2.4, have also been adopted for colour-based image retrieval [23]. Unlike the L_1 and L_2 distance metrics, the quadratic distance compares the similarity between *different* colour bins through use of a distance matrix M, in which the entry M_{ij} represents the distance between colours i and j in the colour space. This distance metric compares each bin in histogram A with all bins in histogram B, as opposed to just corresponding colour bins (as per the L_1 and L_2 distance metrics). The advantage of the quadratic distance metric is that two images containing similar but not exact colours (e.g. a dark red and a lighter shade of red), achieve a smaller distance measure than two images containing very dissimilar colours (e.g. blue and red). With the L_1 and L_2 distance metrics, however, two images containing similar but not exact colours (e.g. a dark and light red) and two images containing very dissimilar colours (e.g. blue and red) will produce an equal distance measure. A disadvantage of the quadratic distance metric, however, is that it is much more computationally expensive than the L_1 and L_2 distance metrics.

Colour histograms are not suitable for use in query-by-sketch systems since they do not account for the location of colour within images. For example, as the two graphics shown in Figure 2.1 have equal quantities of red and blue, they will share identical colour histograms. Clearly, however, the contained colours appear in vastly different locations. Colour histograms are instead more suitable for use when the location of colour is not known or important; systems implementing the query-by-colour query style, for example.

Aston University

Illustration removed for copyright restrictions

Figure 2.1: Two visually dissimilar graphics sharing the same colour histogram [39].

2.1.2.2 Auto Colour Correlograms

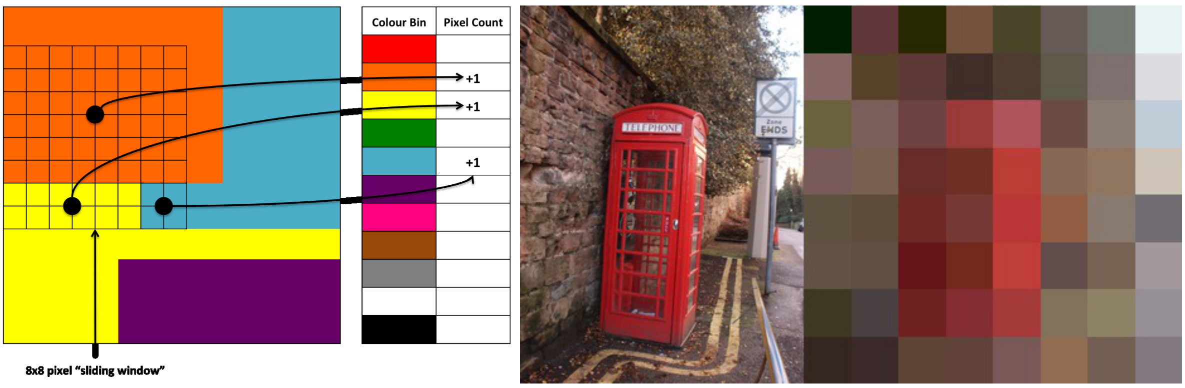
Colour Correlograms were devised by Huang *et al.* [39, 38], and can be seen as a table T indexed by colour pairs, where the k^{th} entry in table cell T_{ij} specifies the probability of finding a pixel of colour i at a distance k from a pixel of colour j in the image. Huang *et al.* also defined a simplified version of their correlogram, known as an Auto Colour Correlogram (ACC), which captures the spatial correlation between identical colours only (reducing the computation required for extraction and comparison). ACCs can be compared using either of the L_1 and L_2 distance metrics. Unlike colour histograms, using ACCs for colour-based image retrieval ensures that images returned via a search contain blocks of the queried colours. Since these coloured blocks could be located anywhere within the image, however, like colour histograms ACCs are only suitable for use with the query-by-colour query style.

2.1.2.3 MPEG-7 Colour Descriptors

MPEG (Moving Picture Experts Group) are a working group developing standards for coded representations of digital audio, video, and related data. MPEG-7, the group's standard for encoding multimedia content, contains several descriptors for representing the colour content within digital images; such as the MPEG-7 Scalable Colour, Colour Structure, and Colour Layout descriptors [59].

2.1.2.3.1 MPEG-7 Scalable Colour Descriptor (MPEG-7 SCD) is a colour histogram encoded using a Haar transform to increase efficiency during retrieval. The histogram is generated by quantising the image into 256 HSV colours, with 16 differing bins for hue, and 4 bins each for saturation and value.

2.1.2.3.2 MPEG-7 Colour Structure Descriptor is also a colour histogram. For the MPEG-7 Colour Structure Descriptor (MPEG-7 CST), however, an 8×8 pixel sliding window moves across the pixels in an image in the HMMD colour space [87] (quantised to 256 colours). With each shift of the structuring element, if a pixel with colour i occurs within the block, the total number of occurrences in the image for colour i is incremented. As shown in Figure 2.2a, at the window's current location each of the 64 (8×8) pixels in the window are one of three colours, thus the counts for these three colours are incremented by one. The distance between two MPEG-7 CSTs can be calculated using the L_1 or L_2 distance metrics. Much like ACCs, searching image databases using MPEG-7 CSTs will return images containing blocks of the queried colours, and, since colour location is not recorded, they can only be used with the query-by-colour query style.



(a) The `sliding window' approach of the MPEG-7 Colour Structure Descriptor.

(b) An image partitioned into 64 (8×8) blocks represented in the MPEG-7 Colour Layout Descriptor.

Figure 2.2: A graphic showing the `sliding window' approach of the MPEG-7 Colour Structure Descriptor (left) and an image of a telephone box partitioned into 64 (8×8) blocks represented in the MPEG-7 Colour Layout Descriptor (right).

2.1.2.3.3 MPEG-7 Colour Layout Descriptor (MPEG-7 CLD) is generated by partitioning an image into 64 (8×8) blocks and calculating the average colour of each block in a YCbCr colour space, creating three 8×8 matrices - one for the luminance (Y) and the blue and red chrominance (Cb and Cr) colour components [59]. This partitioning can be seen in Figure 2.2b. Each of these matrices is then transformed using an 8×8 *discrete cosine transform*. A discrete cosine transform expresses a sequence of data points in terms of a sum of cosine functions set to varying frequencies, helping to separate each matrix into parts of differing importance. The result of this transform, coupled with quantisation, is a matrix for each of the colour components in which the most important frequencies occur in the top left-hand corner. Finally, each colour component is converted to a 64-dimensional vector using a zig-zag scan initiated from the top left-hand corner of its matrix. The distance between two MPEG-7 colour layout descriptors can be measured using the formula shown in Figure 2.5, where i is one of the 64 partitions and w_{Yi}, w_{Cbi}, w_{Cri} are optional weightings for the transformed Y, Cb, and Cr colour components in specific blocks.

$$D = \sqrt{\sum_{i}^{64} w_{Yi}(Y_i - Y_i')^2} + \sqrt{\sum_{i}^{64} w_{Cbi}(Cb_i - Cb_i')^2} + \sqrt{\sum_{i}^{64} w_{Cri}(Cr_i - Cr_i')^2} \tag{2.5}$$

Since the location of colours in images is account for, the MPEG-7 CLD *can* be used for performing colour-based image retrieval via the query-by-sketch query style.

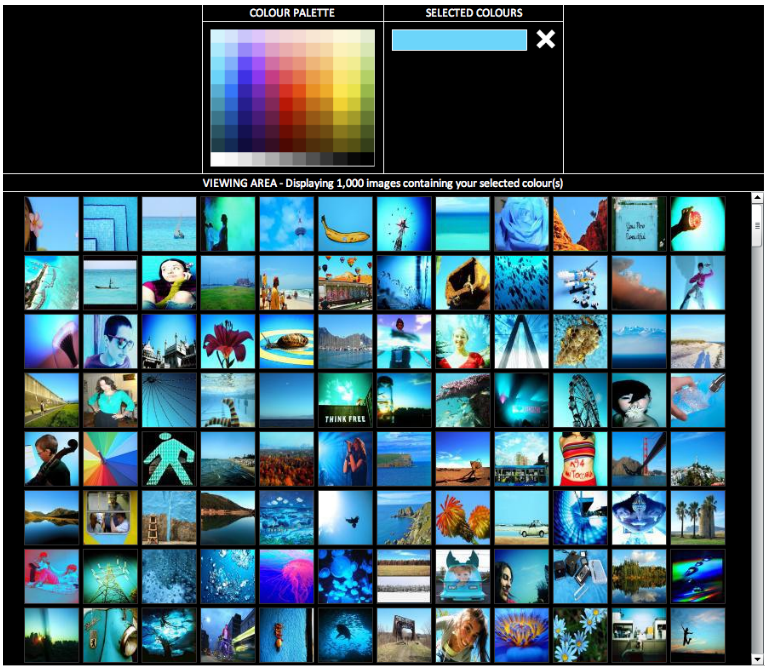


Figure 2.3: Screenshot of a colour-based image retrieval system which adopts the query-by-colour query style.

2.2 Query-by-colour Query Style

The *query-by-colour query style* is the most commonly found approach in colour-based image retrieval systems [23, 43, 22, 42, 44, 17, 3], the most widely used of which are the popular online image search engines Google Images [44] and Bing Images [17]. For query-by-colour, users select their requisite colour(s) from a graphical colour palette (and, in systems such as QBIC [23] and Chromatik [22], also specify a desired ratio for the selected colours). It is from these colour selections that query-by-colour systems formulate query feature vectors. An example of a graphical colour palette can be seen at the top of the system shown in Figure 2.3.

2.3 Alternative Query Styles to Query-by-colour

As described above, the query-by-colour query style is the most commonly adopted approach for colour-based image retrieval systems, but is not compatible with feature vectors accounting for the location of colour within images. In this Section, we describe two alternative query styles to query-by-colour which *can* be used with feature vectors accounting for colour location.

2.3.1 Query-by-example

There are a number of systems which adopt the *query-by-example* query style for image retrieval (e.g. [23, 70, 57, 9]). For query-by-example, users query an image database by uploading a target image representative of their requirements. An example of a query-by-example system interface is shown in Figure 2.4. A fundamental drawback with this query style, however, is that when users do not have a suitable target image to upload, it is very difficult for them to even begin to find database images suiting their requirements. This target image requirement issue has been referred to previously as the *page-zero problem* [102, 54]. It is for this reason that we discounted the query-by-example query style from our research.

2.3.2 Query-by-sketch

To combat the page-zero problem, systems such as QBIC [23] and Virage [30], as well those presented in [47] and [20], adopt the *query-by-sketch* query style along with, or instead of, query-by-example. For the query-by-sketch query style, users manually draw a representation of their required target. Query-by-sketch systems then compute a query feature vector from the synthesised image for comparison with images in the database. The screenshot of the query-by-sketch system developed by Jacobs *et al.* [47] in Figure 2.5 shows the database images returned (right) as a result of the user's sketched query (left). The query-by-sketch query style can be used with the same feature vectors as per query-by-example. It is important to note that we found no reported comparisons of query-based approaches for colour-based image retrieval.

Aston University

Illustration removed for copyright restrictions

Figure 2.4: Screenshot of the IMGSEEK image retrieval system [8] which adopts the query-by-example query style.

Aston University

Illustration removed for copyright restrictions

Figure 2.5: Screenshot of the image retrieval system presented by Jacobs *et al.* [47], adopting the query-by-sketch query style.

2.4 Users of Colour-based Image Retrieval Systems

Surprisingly little research literature exists regarding the user demographic of colour-based image retrieval systems. As a starting point, we asked Tanguay (the Software Engineer at Google) for his opinion on the types of users most likely to benefit from Google Images' newly added colour filter. As well as stating that a ``very large number'' of users invoke the colour filter, he wrote;

``*The kinds of people who can benefit from these filters are looking for images of something particular. Graphic designers, for example, may be looking for an image that fits a particular colour scheme in a web site*''.

Graphic designers are responsible for projects such as web page design, leaflet and poster design, product packaging design and magazine or brochure productions, to name but a few. In consulting several professional graphic designers on the subject of this research, many recalled occasions in which images comprising colours complementing a client's specification or brand were required for a project. These graphic designers cited the usefulness of the query-by-colour facility provided by the Google Images [44] and iStockPhoto [46] systems in achieving such colour-based image searches and selections. Other creative design industries in which individuals must regularly execute colour-based image retrieval include interior and fashion design. The ImageKind [43] and Montage [53] systems have been designed to facilitate colour-based image retrieval for individuals in the interior and fashion design industries, respectively.

Colour-based image retrieval is also an activity in which *graphic artists* must engage for creating life-like computer-generated imagery (commonly referred to as CGI) for use in television and film. Graphic artists are required to apply suitably coloured and textured images to objects and characters existing in virtually-generated environments in order make them appear more realistic and life-like. We found evidence of this upon examining keyword queries logged by a popular online texture repository. We observed that users of CG Textures [99], a web site offering images of real world textures for use in CGI, commonly entered keyword combinations which included at least one colour, such as ``*tile blue*'' or ``*red rust*'', when retrieving images.

User studies undertaken into the image retrieval practises of *image journalists* suggest that, they too, are another user group that regularly performs colour-based image retrieval. Markkula and Sormunen [60] analysed the indexing practices of, and journalist requests to, a digital newspaper archive. One of their observations was that expressive and aesthetic criteria such as colour played an important role during the final selection phase of a typical retrieval session. Similarly, McDonald and Tait [61] also observed that colour was an important consideration factor for image journalists selecting images. Finally, when image journalists participating in the user study of Westman *et al.* [101] were given the option to query an image database according to text, colour, sketch, quality or semantic category, it was found that colour was the second highest query style in the study overall, second only to text [101].

2.5 Similarity-based Visualisations

In contrast to querying, *similarity-based visualisations* [32] arrange and present database images according to similarity, with visually similar images appearing closer together in the visualisation than dissimilar images. Whilst the computation required to create such visualisations is typically complex, the calculations can be performed off-line so that users may interactively explore image databases without the need for a target image or image concept (i.e., users can browse the database for images of interest). In the following Sections, we describe the three most commonly adopted similarity-based visualisation styles.

2.5.1 Mapping-based Visualisations

Feature vectors used to represent (and define similarity between) database images can be high-dimensional data. For the *mapping-based visualisation* style, various *dimension reduction techniques* have been developed which describe the relationships between images in the high-dimensional feature space using a low-dimensional representation that humans can more readily understand - namely the x and y co-ordinates of a digital display. Image thumbnails are placed at the co-ordinates arrived at through dimension reduction to create a mapping-based visualisation (i.e., a layout of database images). These visualisations can be explored interactively using zoom and pan tools, often included so that users may inspect areas of interest within the visualisation in greater detail before selecting the images they wish to retrieve [71, 6, 16]. In the next Section, we describe the dimension reduction techniques which have previously been adopted for generating similarity-based visualisations.

2.5.1.1 Dimension Reduction Techniques

Whilst much research has been undertaken regarding methods for dimension reduction, in this Section we focus our attention on techniques which have been explicitly adopted for generating mapping-based visualisations of image databases; namely *Principal Components Analysis* and *multi-dimensional scaling*.

2.5.1.1.1 Principal Component Analysis Principal Component Analysis is the simplest dimensionality reduction approach previously adopted for generating mapping-based visualisations (operating with an $O(n)$ - or linear - complexity). The starting point for Principal Components Analysis is the *symmetric covariance matrix* of the feature data. Covariance, expressed formally in Equation 2.6, is a measure of how much two dimensions vary from the mean with respect to each other. A symmetric covariance matrix is, therefore, a matrix whose i, j value represents the covariance between the i^{th} and j^{th} feature vectors from the feature data.

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \tag{2.6}$$

The next step in Principal Component Analysis is to calculate the *eigenvectors* and their respective *eigenvalues* for the covariance matrix. The eigenvectors are non-zero vectors which, when multiplied by the covariance matrix, produce multiples of themselves. The respective eigenvalue is the multiple. The computed eigenvectors are ranked in descending order of eigenvalues. The eigenvector with the highest eigenvalue is the principal component of the dataset as it represents the most significant relationship between data dimensions. The remaining principal components are selected from the top eigenvectors according to the number of dimensions required (i.e. for 2D the top two eigenvectors are selected). These principal components are then combined to form a transformation matrix which is then applied to the original feature data. Principal Components Analysis has the advantage that it is relatively simple. Since it maximises the variance of the captured data, however, it does not necessarily best preserve the mutual relations between the individual data items.

The Personal Digital Historian system [62], developed for sharing digital photographs amongst groups of users, adopted Principal Component Analysis to visualise subsets of database images which were projected from above on to a table below. Keller *et al.* [50] also adopted Principal Component Analysis to visualise images in a virtual 3D environment based on their texture content.

2.5.1.1.2 Multi-dimensional Scaling In contrast to Principal Component Analysis, multi-dimensional scaling [52] attempts to preserve the original relationships (i.e. distances) of the high-dimensional space as best as possible in the low-dimensional projection. Multi-dimensional scaling starts with a similarity matrix which describes all of the pair-wise distances between objects (or images) in the original, high-dimensional space. The aim of multi-dimensional scaling is to best maintain these distances within the low-dimensional projection by minimising a *stress* measure, often defined as per Equation 2.7 whereby δ_{ij} is the original distance between objects i and j, and $\hat{\delta}_{ij}$ is the distance in the low-dimensional space [52].

$$\text{STRESS} = \frac{\sum_{i,j}(\hat{\delta}_{ij} - \delta_{ij})^2}{\sum_{i,j}\delta_{ij}^2} \tag{2.7}$$

Starting from either a random initial configuration or co-ordinates derived from Principal Component Analysis, the multi-dimensional scaling algorithm iteratively repositions images in the low-dimensional space in order to reduce the overall stress. The algorithm repositions points in the low-dimensional space that correspond to points in the original space until a termination condition has been reached, typically a maximum number of iterations or a predefined threshold stress value. Whilst this means that multi-dimensional scaling is able to provide a more accurate representation of the relationships between images in the high-dimensional feature space compared with Principal Component Analysis, the computational cost is much higher - with multi-dimensional scaling incurring an $O(n^2)$ quadratic complexity.

2.5.1.1.2.1 Use of Multi-dimensional Scaling The use of multi-dimensional scaling for image database visualisation was first proposed by Rubner *et al.* [77, 78]. Rubner *et al.* suggested that multi-dimensional scaling could be used to visualise either query results or indeed entire databases to provide users with an overview of an image database. Whilst we agree that multi-dimensional scaling is suited to computing a global visualisation (as it can be computed off-line and therefore the time required for computation is inconsequential), we would argue that multi-dimensional scaling is too complex a technique to generate visualisations in response to a user query, particularly since systems such as Google Images [44] return 1,000 images or more in response to users' queries. It would appear that Pecenovic *et al.* [69] also reached the same conclusion when designing their image retrieval system. Pecenovic *et al.'s* system uses multi-dimensional scaling to generate a global database visualisation, but presents query results as a ranked list.

Rodden *et al.* have investigated several aspects of mapping-based visualisations generated using multi-dimensional scaling [73, 74, 75]. In [74], Rodden *et al.* compare the accuracy of different feature vectors (and associated distance measures) for producing mapping-based visualisations via multi-dimensional scaling. Their findings suggest that using lower-dimensional feature vectors as input for multi-dimensional scaling - such as a 3-dimensional average colour feature vector - can produce visualisations which appear remarkably similar to those produced using much more complex feature vectors.

As well as their theory-based analysis of mapping-based visualisations outlined above, Rodden *et al.* have also conducted user-based studies investigating mapping-based visualisations generated using multi-dimensional scaling. For their first user study, they compared the times required by users to retrieve target images from a mapping-based visualisation and a randomly assorted arrangement [73]. Results of the study showed that users could locate their target images significantly faster in the mapping-based visualisation than the randomly assorted arrangement. In a later user-based study, however, Rodden *et al.* observed the opposite result [75]. Surprisingly, users participating in the second study took significantly longer to retrieve images (to accompany a fictional travel article) from the mapping-based visualisation than when retrieving from the randomly assorted grid. Rodden *et al.* note, however, that users in the study were generally more satisfied with the appropriateness (for use in the fictional travel article) of the images retrieved from the mapping-based visualisation than their respective selections from a randomly assorted grid.

Another key finding from the user studies conducted by Rodden *et al.* was that users disliked the overlapping of images that can often occur in mapping-based visualisations. As can be seen in Figure 2.6, co-ordinates that are close together in the high-dimensional feature space inevitably become even closer when reduced to just two dimensions, causing images to overlap, or be completely occluded, by other images in the visualisation. In the following Section, we describe several proposed solutions to this image overlap issue.

Aston University

Illustration removed for copyright restrictions

Figure 2.6: Visualisation of the UCID image database [83] generated using multi-dimensional scaling. Many of the images have been overlapped, or completely occluded, by other images in the visualisation.

2.5.1.2 Overcoming Image Overlap

Nguyen and Worring [67] specify two requirements for managing overlap in mapping-based visualisations; a *structure preservation requirement* and an *image visibility requirement*. The first requirement states that the structure of the relationships between images in the feature space should be retained, while the second demands that images should be visible enough so that the content of the image is distinguishable. It is clear that these two requirements are intrinsically conflicting. Moving an image in order to make it more visible will detract from the original visualisation structure, whilst maintaining the structure could cause a loss of visibility for certain images. As a proposed solution, Nguyen and Worring define a cost function which considers both image overlap and structure preservation. In order to detect overlap in their solution, a circle is placed about the centre of the image - as it is assumed that an object of focus will be about the centre of an image. If the circles of two images overlap, the position of the images is altered according to values derived from the cost function. A similar cost function is also adopted in the Personal Digital Historian system [62].

2.5.1.2.1 Fitting Visualisations to a Regular Grid As a solution to their finding regarding image overlap, Rodden *et al.* [73] proposed an algorithm for fitting mapping-based visualisations to a regular grid. Figure 2.7 demonstrates the outcome of their proposed solution. For their algorithm, a *minimum spanning tree* is first calculated from the similarity matrix (used as input for multi-dimensional scaling) and used to order images. This minimum spanning tree is the shortest possible path that interlinks all

Aston University

Illustration removed for copyright restrictions

Figure 2.7: Visualisation of 100 images generated using multi-dimensional scaling (left), fitted to a regular grid (right) [73].

of the images in the visualisation. Images are mapped to cells of the grid in this order according to their visualisation co-ordinates. If an image's intended cell is already occupied, a spiral search (emanating from this original cell) is performed in order to locate the closest free cell (see Figure 2.8a). In addition to this *basic* strategy, where the image is simply mapped to the next closest free cell, a further *swap* strategy was also proposed. As shown in Figure 2.8b, for the swap strategy an image is moved from the initial cell to the next closest cell, and the new image is placed in the original cell. Finally, in their third *bump* strategy, the images in the line of cells between the original cell and the next closest cell are all moved outwards (from the original cell at the centre) by one cell, with the new image placed in the original cell. This strategy is shown in Figure 2.8c. In their experiments, Rodden *et al.* [73] found that the *bump* strategy produced the lowest average error (i.e., the lowest average distance an image was from its original cell). This *bump* strategy of Rodden *et al.* has a computation complexity of $O(m^2) + O(n^2)$ (where m is the size of the grid and n is the number of images to be located).

Aston University

Illustration removed for copyright restrictions

(a) Basic Strategy (b) Swap Strategy (c) Bump Strategy

Figure 2.8: Spreading strategies proposed by Rodden *et al.* [73]. Dashed arrows represent the spiral scan, whilst solid arrows depict the path of an image.

Liu *et al.* [55] proposed a more efficient strategy for fitting mapping-based visualisations to regular grids. Liu *et al's.* algorithm first creates an ordered set of the visualisation images based on their y co-ordinates. Images are then placed in grid cells sequentially; the image with the lowest y co-ordinate is placed in the first cell of the top row of the grid, whilst the image with the highest y co-ordinate is assigned to the final cell of the bottom grid row. Then, for each row of the grid, images are sorted according to their x co-ordinate; the image assigned to the first cell of a row has the lowest x co-ordinate of all the images within that particular row, whilst the image located to the last cell will have the highest x co-ordinate of all images in the row. The resultant visualisation is thus optimised in one dimension whilst sub-optimised in the other, and has a complexity of $O(2n\log(n) - n\log(m))$ (where m is the number of rows in the grid and n is the number of images).

2.5.1.3 Mapping-based Visualisation Interactive Tools

When large image databases are visualised on a single 2D plane, the resolution of images is often reduced so much that their content becomes almost impossible for users to interpret. As we have noted, systems adopting mapping-based visualisations typically include zoom and pan tools, amongst others, to enable users to inspect areas of a visualisation (and thus individual images) in greater detail. The systems described in [13, 81] provide a navigational toolbar which enables users to zoom and pan database visualisations. As well as including such a toolbar, Barthel's ImageSorter system [5] also permits zooming and panning via mouse interactions; namely scroll wheel manipulation for zooming, and a left-click and drag operation for panning.

Combs and Bederson [16] recruited 30 participants for a study into the effectiveness of zooming for image retrieval. As part of their study, Combs and Bederson compared their prototype ZIB (Zoomable Image Browser) system with the commercially available ThumbsPlus system [10]. In their ZIB system, users could zoom in and out of an image set, whilst in ThumbsPlus, users were required to scroll. Combs and Bederson report that there was no significant difference in the times required by users to retrieve specific images using the two systems. Furthermore, Combs and Bederson report that, despite up to 225 images being displayed simultaneously in their ZIB system and all users being trained on the functionality of the system, only half of the participants actually performed a single zoom operation. Given that these users would have been viewing such small images, it is perhaps unsurprising to learn that the error count for the ZIB system - that is, the number of images retrieved by users that were not identical to the image they were asked to find - was found to be higher than for Thumbsplus.

It is our belief that the effectiveness of zooming for image retrieval will correlate directly with the number of images being visualised - that is, the more images that are displayed, the more effective zooming will become. As such, because the study of zooming conducted by Combs and Bederson only adopted datasets containing a maximum of 225 images, no notable difference was observed between the zoomable and non-zoomable systems. If the study were to be repeated with much larger datasets (e.g., 10,000 images or more), we would expect a much higher percentage of users to perform zoom operations in the ZIB system and, therefore, we would expect to observe a reduced error rate, as users would have to zoom in order to determine the content of images. Furthermore, we would anticipate a faster retrieval rate in the zooming-based system than the non-zooming alternative, as users will be able to navigate to images of interest more quickly

Whilst zooming increases the resolution of all images in the visualisation, some image retrieval systems facilitate previewing individual images within a visualisation by hovering the mouse cursor over an image of interest. This maintains the overall structure of the database visualisation by first rendering only small thumbnails image first, with higher resolution images loaded upon request. Google Images [44] and PhotoMesa [6] include this previewing facility. Mouse over previewing can also be applied to a small group of images via a fisheye lens [79]. Using this mechanism, images located at the centre of the lens are rendered at a higher resolution (i.e., the images of interest), while those around the focussed area are distorted [71], causing the images of interest to appear more prominently on the display.

In the following Section, we describe clustering-based visualisations which reduce the number of images that are displayed to users at any one time, thus removing the need for the interactive tools above.

2.5.2 Clustering-based Visualisations

The mapping-based visualisation style described above attempts to display *all* of the images to be visualised which, as we saw in Figure 2.6, can lead to `cluttered' visualisations. Whilst we have discussed approaches for fitting such visualisations to a regular grid, research has been undertaken which reduces the total number of images to be displayed to users by clustering - or grouping - similar images together. Whilst not always the case [37, 14], clustering is typically performed in a hierarchical manner. Clustering-based visualisations create groups of images with similar feature vectors which are represented visually using a single image (or, in some cases, a small set of images) representative of the images contained within the group. Image groups can contain sub-groups of images, which themselves may also harbour further sub-groups of images, and so on. Users can explore this hierarchical structure by navigating through groups of interest using an interface similar to that of file browsers found in common operating systems (i.e., the representative images of groups act much like folders and individual images become analogous to files). In Figure 2.9, for example, a user has traversed the hierarchy to a cluster of video stills captured from the same news broadcast.

2.5.2.1 Uses of Hierarchical Clustering

Krischnamachari and Abdel-Mottaleb [1, 51] have investigated using agglomerative and divisive clustering techniques to produce image database visualisations. Agglomerative, or bottom-up, clustering begins with treating each individual sample as an individual cluster. Using some predetermined defi-

Figure 2.9: Screenshot of Borth *et al's.* Navidgator system [7]. A user has traversed the hierarchy to cluster of video stills captured from the same news broadcast.

nition of similarity, clusters are merged with their most similar neighbours and this process is repeated until a pre-defined number of clusters remain. These clusters then form the top layer of the generated tree. In contrast, divisive, or top-down, clustering begins with all samples starting as a single large cluster which is then iteratively split into smaller clusters until a termination criterion is met (such as all clusters corresponding to individual samples or, in our case, images) [48].

Schaefer and Ruszala [80, 82] adopted an alternative technique for hierarchically clustering image databases. At the root layer, all database images are assigned to cells based on co-ordinates derived through multi-dimensional scaling. If a root layer cell contains only a single image, this is the image that is displayed in the cell. If more than one image is mapped to a cell, however, the image whose co-ordinates are closest to the cell centre is displayed as a representative image for that image cluster. Then, at the next level of the hierarchy, this cluster is expanded by subdividing the cell into a set of smaller cells and performing the above procedure again (that is, mapping each image in the cluster to a sub-cell, and, should multiple images be assigned to the same sub-cell, descend into the next level of the hierarchy).

Schaefer and Ruszala employed two approaches in an attempt to reduce the number of empty cells and layers within a hierarchy, the first of which is shown in Figure 2.10. If an empty cell has three or more neighbouring cells that contain images, a fraction of these images are moved to the empty cell. The images moved from each of the occupied cells are those with derived co-ordinates closest to those of the centre of the empty cell (i.e., those images closest to the borders between the two cells). The overall effect of this approach is that more cells are filled, hence making better use of the visualisation space.

Figure 2.10: The evening strategy of Schaefer and Ruszala [82]. Images (represented as black dots) have been moved from the cells neighbouring the vacant cell to produce a more even distribution of images.

For their second approach, Schaefer and Ruszala combined the basic, swap and bump spreading strategies proposed by Rodden et al. [73] (discussed previously in Section 2.5.1.2.1) in order to prevent the creation of an undue number of layers. When a given parent cell consisted of fewer than 25% occupied sub-cells, sub-cells containing more than a single image would share their images with the nearest vacant sub-cells. If an empty cell was located in the first ring of neighbours, then the `basic' strategy was applied. If the closest vacant cell was in the second ring of neighbours, however, then the `swap' strategy was applied. Finally, should the closest cell be located within the third ring of neighbours, Schaefer and Ruszala applied the `bump` strategy.

2.5.2.2 Selection of Representative Images

For visualisation purposes, each clustered group of images can be represented using a single image or a small group of images [55, 21]. In many approaches, such Schaefer and Ruszala's [82], the centroid image of the cluster is selected. Formally, this is the image with the minimal cumulative distance from all other images in the cluster. Alternatively, other systems such as CAT [28] select the image closest to the centroid of the cluster in the feature space. These representative images serve as links which users can follow to traverse through the hierarchy, usually by clicking the representative images of interest. From this, users are presented with all of the images from the cluster that the selected representative image represents.

The representative images of clusters, particularly at higher levels of the hierarchy, are important for effective image retrieval in this visualisation style. If none of the representative images at the root level bear relevance to the target image or image concept required by a user, it will be unclear which cluster contains images of interest. Similarly, whilst clustering images according to content reduces the number of images required to be visualised at any one time, this visualisation style is heavily reliant upon the underlying features and metric used in defining similarity. Should the definition of similarity not match human perception well, an image could be placed within clusters that users would not intuitively think to inspect in order to retrieve it.

2.5.3 Graph-based Visualisations

For completeness, this Section details *Graph-based visualisations*; the least researched of the three visualisation styles. As can be seen in Figure 2.11, graph-based visualisations utilise links between images to construct a graph where the nodes of the graph are the images, and the edges form the links between similar images. On inspection of the graph-based visualisation shown in Figure 2.11, we believe that a visualisation of 1,000 images or more would become too `cluttered', making it difficult for users to retrieve images effectively. It is for this reason that we discounted graph-based visualisations from our research.

2.5.3.1 Pathfinder Networks

Chen *et al.* [13] adopted pathfinder networks to formulate graph-based visualisations for their InfoViz system. The pathfinder algorithm, originally used to analyse proximity data in psychology research, removes all but the shortest edges (links) in the graph [12]. Users of the InfoViz system could interactively explore the graph-based visualisation using zooming and panning tools to inspect areas of the visualisation in more detail. The major disadvantage with this global graph-based visualisation style would appear to be the number of images it can handle. Chen *et al.* test their visualisations with just 200 images, which is far short of the 1,000 image results returned by image search engines like Google Images [44].

Figure 2.11: A graph-based visualisation of database images [49].

2.5.3.2 NN^k Networks

Heesch and Rüeger [33, 34] devised NN^k networks in their research, where NN stands for *nearest neighbour* and k describes a set of different features. In an NN^k network, a directed graph is formed between every image and its nearest neighbours if there exists at least one possible combination of features for which the neighbour image is ranked top. As an example, a link may be made between images A and B if B is the most similar image in the database to A according to colour content. A link may then also be made between A and another image, C, as C is the most similar image in the database to A according to texture. In contrast to Chen *et al's.* InfoViz system, users of the systems designed by Heesch and Rüeger are required to navigate through the database one image at a time, starting from an initial set of images designed to represent the overall content of the database. Selecting a representative image places it as a query image about the centre of the screen, surrounded by its linked neighbours. The user is then able to select a neighbour, which then becomes the query image, surrounded by its respective neighbours. The inherent disadvantage with this visualisation is the difficulty also encountered with the clustering-based visualisation style; if none of the representative images match the target image or image concept users are aiming to retrieve, users could be left confused as to where their images of interest may be located.

2.5.4 Comparison of Similarity-based Visualisation Styles

Liu *et al.* [55] undertook a user study comparing three image retrieval systems which presented query results either in the form of a ranked list, a mapping-based visualisation, or a clustering-based visualisation. Nine users were given 17 query terms and asked to find the image in the test database (containing 3,400 images) which best matched each of the pre-determined query terms. It was found that users could find images 26% faster when using the systems presenting query results in the form of mapping-based and clustering-based visualisations as opposed to the ranked list presentation. Whilst the search times recorded for the systems adopting mapping-based and cluster-based visualisations for presenting query results were not found to be significantly different, users did assign higher preference ratings to the mapping-based visualisation system - claiming it to be more intuitive, interesting, and convenient than the ranked list and clustering-based visualisation systems for searching and comparing images. Seven out of nine preferred the mapping-based system, whilst the other two preferred the clustering-based alternative.

2.6 Colour-based Image Retrieval Evaluations

As we have stated previously, there are several examples of colour-based image retrieval systems in existence today [23, 43, 22, 42, 44, 17, 3]. Surprisingly, however, there have been very few reported evaluations of colour-based image retrieval systems. In this Section, we describe and discuss the few existing examples of colour-based image retrieval system evaluations.

2.6.1 Estimating Colour Proportions

Some colour-based image retrieval systems adopting the query-by-colour query style, such as QBIC [23] and Chromatik [22], permit users to explicitly set the required ratio of their selected colours. Chan and Wang [11] conducted a user study investigating the usefulness of this proportion tool, measuring the accuracy of users' estimates of the quantity of colour present within an image. In their study, Chan and Wang presented participants with images of a solid red ellipse against a white background. Participants were asked to estimate the area of the image (as a percentage) that was occupied by the red ellipse. Each subject was shown a series of images across which the size of the red ellipses varied. Their investigation revealed that, on average, users over-estimated the size of these ellipses. For the second part of the study, participants were shown images with dispersed red shapes on a white background, and asked to estimate the total area of each image occupied by the red shapes. The red shapes varied in size and number across the images. The results showed that users yet again over-estimated the proportion of the images occupied with coloured shapes. Chan and Wang [11] concluded that further research was required to develop a statistical model for use in querying which considers the inaccuracy of colour quantity estimation apparent in users.

2.6.2 Graphical Colour Palette Design

A more fundamental feature of the query-by-colour query style is the graphical colour palette from which users select their colours. Broek *et al.* [98, 97] investigated which of several graphical colour palettes was most suited for colour-based image retrieval systems. Various palettes, including those adopted for the QBIC [23] and VisualSEEK [90] systems, were assessed using *Fitts' law*. Fitts' law is a calculation designed to model human movement, predicting the time required to point (either physically or virtually using a mouse cursor) to a target area. It is calculated using the distance to, and the size of, a given target. Broek *et al.* used Fitts' law to estimate the time that would be required by users to generate a sketched query based on parameters relating to the distances between, and sizes of, the colour cells of the palettes tested.

As part of their study, Broek *et al.* also tested a novel colour palette, based on the theory that humans divide colour into eleven discrete categories - red, green, blue, yellow, brown, purple, pink, orange, black, white, and grey. The novel palette of Broek *et al.* also included the colour cyan. Broek *et al.* conclude that, according to Fitts' law, their newly proposed graphical colour palette supported more efficient colour selections for query-by-sketch than the others tested. While Broek *et al.* [98] demonstrate that the design of graphical colour palettes can impact upon colour selection times, we believe it would have been more beneficial to have measured how real-world users responded to the tested graphical colour palettes whilst undertaking an assessed image retrieval task (i.e., how long did it take them to retrieve specific images from a database, which of the palettes tested did users prefer most for generating their sketches, etc.).

2.6.3 Evaluating the Query-by-colour Query Style

To evaluate the query-by-colour facility of the QBIC system, Faloutsos *et al.* [23] supplied users with target images and asked them to mark each image in a database (of approximately 1,000 images) as either relevant or irrelevant with respect to the supplied target image. The participating users were then asked to submit a single colour query to the QBIC system that they believed would be sufficient to retrieve each target image from the database (i.e., they submitted one query for each target image). Using the top 20 images returned by QBIC from the user's query, Faloutsos *et al.* calculated the average rank ($AVRR$) of all the relevant images occurring in the result set and the ideal average rank ($IAVRR$). The ideal average rank can be defined as $IAVRR = (0 + 1 + \cdots + (T - 1))/T$, where T is the total number of images in a database relevant to a target image. According to these definitions, an effective colour-based image retrieval system will achieve an $AVRR$ close to the $IAVRR$ value. In the study by Faloutsos *et al.*, these values were averaged over ten different target images.

In the following Section, we consider the suitability of the approach above - as well as several others that are available - for assessing the effectiveness of a colour-based image retrieval system.

2.7 Image Retrieval Evaluation Approaches

In the previous Section, we have highlighted the only recorded approaches for measuring the effectiveness of colour-based image retrieval systems. We believe an effective colour-based image retrieval is one with which users can find the database images closest to their colour requirements, quickly and easily. Whilst the other approaches are novel, the evaluation method adopted by Faloutsos *et al.* is based on the commonly adopted *precision and recall* measures (typically reported together as they are intrinsically linked) used to evaluate image retrieval systems. *Precision* is calculated by dividing the number of relevant images returned by a system by the total number of images returned [63], whilst *recall* is defined as the number of relevant images returned by a system, divided by the total number of relevant images in the test database. The measures of precision and recall are presented formally in Equations 2.8 and 2.9 respectively. To measure the relevance of returned images, the n images retrieved by a system are compared with a list of `correct' images for the given target image. This list is commonly referred to as a *ground-truth*.

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{2.8}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \tag{2.9}$$

In the evaluation method of Faloutsos *et al.*, users were required to manually label all images in a database as either relevant or irrelevant to a given target image. Not only is asking human judges to manually label all relevant images to a target image a very arduous, expensive, and time consuming task, this list of `correct' images is open to the high degree of varying subjectivity that is possible between different human assessors. The research field of image retrieval currently lacks a standardised large-scale image database with an associated ground-truth [64, 56] for benchmarking. The major difficulties involved in creating such a standardised image database relate not only to copyright issues associated with the reproduction and distribution of images, but also the inherent difficulty in producing a reliable ground-truth. In the early days of image retrieval research, each new system or approach was evaluated using a different database and ground-truth. As well as making it impossible to replicate studies, Müller *et al.* [63] have also shown how using different subsets of image databases for testing can be manipulated to enhance the perceived performance of a system.

To overcome this lack of a standardised image database with an associated ground-truth for benchmarking, evaluation initiatives - such as TRECVID [4] (a track of the Text REtrieval Conference dedicated to devoted to research in automatic segmentation, indexing, and content-based retrieval of digital video) and ImageCLEF [15] (an evaluation forum for the cross-language annotation and retrieval of images) - have been set up to promote collaboration between members of the image retrieval community and provide frameworks and resources for systematic and standardised evaluation of image and video retrieval systems (note that, since video retrieval analyses individual frames, it is considered analogous to image retrieval). In particular, they provide reusable large-scale image databases with associated ground-truths and pre-determined tasks (e.g., search and annotation) to participants so that systems and approaches can be benchmarked and meaningfully compared. One way in which such evaluation initiatives formulate their ground-truths is through *pooling*. In this approach, ground-truths are formulated automatically using the relevance assessments of systems participating in the initiative. As an example, if many annotation systems participating in the initiative assign a common keyword to a particular image, the ground-truth used by all participants of the initiative will be updated to include the commonly assigned keyword.

Given that image retrieval is a highly interactive process, a fundamental problem with the precision and recall measures is they do not consider the usability of tested systems [15]. Attempts have been made to run interactive tasks as part of ImageCLEF - e.g., the *target search* [16, 73, 72, 28] and *category search* [75, 29, 66] tasks - but participation in these tasks is typically low due, perhaps due to the inevitable difficulties involved with recruiting test-users required to complete them. For the category search task, test-users are shown a target image or keyword and instructed to find as many relevant images as possible in a database (using the system being tested) within an allocated time limit. The number of images found in the time available is used as a measure of the system's effectiveness. For target search, users are instead asked to retrieve a specific image from a database using the tested system, with the time taken used as a measure of the tested system's performance. As we describe in the next Section, however, we believe that even the user-based category and target search tasks are unsuitable for evaluating the effectiveness of colour-based image retrieval systems.

2.7.1 Drawbacks for Colour-based Image Retrieval

Despite the efforts of the research initiatives listed in the previous Section, no large-scale image database and ground-truth currently exists for assessing colour-based image retrieval systems and approaches. As we have stated above, there are a variety of difficulties in producing such a database without a concerted effort from a large group of researchers. Furthermore, as we have also stated, the precision and recall do not account for the usability of an image retrieval system - a particular problem given the highly interactive nature of colour-based image retrieval. We believe, therefore, that the precision and recall measures are not suited for evaluating the effectiveness of colour-based image retrieval systems.

As we also report in the previous section, the category and target search tasks - in contrast to the precision and recall measures - are interactive evaluation approaches which do account for the usability of tested systems. The problem with these approaches, however, is that each fails to accurately reflect the manner in which highly creative users, such as those identified in Section 2.4, assess the suitability of images for use in their projects. These users typically determine the compatibility of images within the context of the project for which the retrieved image is intended. For example, graphic designers add images to web page designs before deciding on their appropriateness. Graphic designers will then decide whether to keep their chosen image, or, discard the image and refine their search criteria according to its shortcomings. This iterative image selection process is known formally as *reflection-in-action* [84]. This is in contrast to the category and search tasks since test-users have no overall project in which the suitability of selected images can be assessed.

Given the drawbacks reported in this Section, we believe that, prior to the research documented in this thesis, no suitable evaluation approach existed for assessing the effectiveness of colour-based image retrieval systems.

2.8 Summary and Research Aims

In this Chapter, we have reviewed the current state of colour-based image retrieval research. It surprised us to learn in Sections 2.1, 2.2 and 2.3 that the underlying query-based technology used for colour-based image retrieval in Google Images [44] and Bing Images [17] was not as innovative as we expected and, is in fact, based on a retrieval model devised almost two decades ago [23]. We found this to be somewhat of a revelation given that image databases have grown exponentially over the last twenty years and can now contain millions of images [68]. We were also surprised at the lack of literature available regarding the user demographic of colour-based image retrieval systems. From the little previous research available, coupled with our consultations with graphic designers, we deduced in Section 2.4 that individuals operating in creative industries such as graphic, fashion and interior design, search image databases by colour. Tanguay, the Software Engineer at Google, also revealed that ``a very large number'' of users invoke the colour filter in Google Images each day.

Discovering this ageing retrieval approach and large user demographic made us eager to research more effective approaches for colour-based image retrieval.

In Section 2.5, we detailed the substantial body of research regarding *similarity-based visualisations* for image retrieval. In contrast to query-based approaches, similarity-based visualisations display and arrange database images so that images with similar content are located closer together than images with dissimilar content [32]. This removes the need for queries, because users can instead visually explore the database using interactive navigation tools to retrieve images of interest from the database. Since we were unable to find any evidence of similarity-based visualisations being evaluated previously for colour-based image retrieval, coupled with our finding that no research existed directly comparing the existing query-by-colour and query-by-sketch query styles for colour-based image retrieval, we formulated the research aim; *to evaluate and compare the effectiveness of query-based approaches (i.e., the query-by-colour and query-by-sketch query styles) and similarity-based visualisations for colour-based image retrieval*.

Further to the above finding, we found two other aspects of similarity-based visualisations that we believed warranted further investigation. First, many systems adopting mapping-based visualisations for image retrieval include a zooming facility for exploring databases. As described in Section 2.5.1.3, however, there has been just one (seemingly flawed) study into the importance of zooming in these systems - that is, the study of Combs and Bederson [16]. As we believed the importance of zooming for image retrieval would correlate directly with the number of images being visualised (i.e., the more images that are displayed, the more effective zooming will become), we decided to make an additional aim for this research; *to assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases*.

The second aspect of similarity-based visualisations that we believed warranted further investigation was their potential for use in presenting the results of colour-based image retrieval queries. As outlined in Section 2.5.4, Liu *et al.* [55] undertook a user study comparing three image retrieval systems which presented keyword query results either in the form of a ranked list (where images ranked by the system as most similar to a query appear at the head of the list), a mapping-based visualisation, or a clustering-based visualisation. Since Google Images and Bing Images currently adopt a ranked list visualisation for presenting both keyword and colour query results, we wanted to evaluate what effect changing the visualisation of results would have on the large number of users searching by colour (as opposed to keywords as per Liu *et al.*'s investigation). Consequently, we added the research aim; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*.

In order for us to fulfil the research aims stated above, our first task was to overcome the lack of a reliable method for systematically and meaningfully evaluating colour-based image retrieval systems. In Section 2.6, we reviewed the few reported system evaluations that have been conducted previously. Furthermore, in Section 2.7.1 we described the fundamental drawbacks of existing evaluation approaches; namely the field of colour-based image retrieval lacking a standardised database with associated ground-truth, and their failure to support *reflection-in-action* [84] - that is, the iterative image selection process conducted by creative individuals. As a solution, we devised the *Mosaic Test*. The Mosaic Test is a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system under evaluation. The time and users' perception of the workload required to complete this creative task, as well as the visual accuracy of their image mosaics (in comparison with the initial target images), are used to assess the effectiveness of the system being tested. In the following Chapter, we describe how we fulfil another aim of our research; *to show that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems*.

2.8.1 Research Aims

For clarity, the aims of this research can be summarised as follows:

- Show that that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems,
- Evaluate and compare the effectiveness of query-based approaches (i.e., the query-by-colour and query-by-sketch query styles) and similarity-based visualisations for colour-based image retrieval,
- Assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases,
- Compare the effectiveness of colour-based image retrieval query result visualisations - namely, clustering and mapping-based visualisations against ranked lists.

Chapter 3

The Mosaic Test

3.1 Introduction

In order for us to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval, as well as accomplish the other aims of this research, we first needed to overcome the field lacking a reliable method for evaluating colour-based image retrieval systems. As stated in the previous Chapter, the field of colour-based image retrieval does not have a standardised strategy for evaluation which has resulted in very few previous studies assessing the effectiveness of systems and/or approaches for retrieving images from a database by colour. The previous Chapter also outlined the limitations of alternative evaluation strategies typically adopted for assessing image retrieval systems; the precision and recall, and category and target search tasks, are unreliable due to the field of colour-based image retrieval lacking a standardised database with associated ground-truth, and their failure to support *reflection-in-action* [84] - that is, the iterative image selection process conducted by creative individuals.

In this Chapter, we describe how we fulfilled our research aim; *to show that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems*. The Mosaic Test is a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system under evaluation. The time and users' perception of the workload required to complete this creative task, as well as the visual accuracy of their image mosaics (in comparison with the initial target images), are used to assess the effectiveness of the system being tested. Since our Mosaic Test adopts a standardised image database, automatically assesses the relevance of selected images by measuring the accuracy of generated mosaics, and enables users to reflect on the suitability of images for use in their mosaics, we believed that the Mosaic Test would provide a reliable mechanism by which to meaningfully and systematically evaluate and compare systems for colour-based image retrieval.

To support the above claim, this Chapter reports on findings from our first user study with 24 participants which relate directly to aspects of our novel Mosaic Test. As well our observations of participants undertaking reflection-in-action during the study, we provide empirical evidence to support our use of MPEG-7 colour structure descriptors and the L_1 distance metric for measuring image mosaic accuracy (to automatically assess the relevance of images selected by test-users during a system evaluation). In further support of our newly proposed Mosaic Test, we discuss our finding that recruiting either expert and non-expert users of systems for colour-based image retrieval has no significant effect on the effectiveness data recorded from a Mosaic Test, simplifying participant recruitment.

3.1.1 Chapter Outline

This Chapter is organised as follows. In Section 3.2 we describe the Mosaic Test and associated *Mosaic Test Tool*. We also describe here how effectiveness data is recorded during a Mosaic Test session. In Section 3.3 we describe the design of our first user study, before presenting and discussing our findings in Sections 3.4 and 3.5 which support our claim that the Mosaic Test allows for *meaningful* comparisons of colour-based image retrieval systems. Finally, Section 3.6 concludes the Chapter.

3.2 The Mosaic Test

As described above, the Mosaic Test is a user-based evaluation approach in which participants complete an image mosaic of a predetermined target image. An image mosaic is an art form in which a target image is divided into cells, each of which is then replaced by an image with similar colour content to the corresponding cell in the target image. Viewed from a distance, the smaller images collectively appear to form the target image, whilst viewing an image mosaic close up reveals the detail contained within each of the smaller images [88]. Using the *Mosaic Test Tool*, participants of a Mosaic Test manually produce an image mosaic by retrieving images from the test image database using the system that is being evaluated. In the following Section, we describe the Mosaic Test Tool.

3.2.1 The Mosaic Test Tool

As stated above, Mosaic Test participants are asked to manually create an image mosaic of a predetermined target image. Unfortunately, no software currently exists specifically for *manually* creating image mosaics. Whilst McKean [25] has demonstrated that it is possible to create mosaics manually using image editing software, not everyone has extensive experience (or training) using such applications, and licensing costs can be expensive and thus prohibitive. Therefore, to support the manual creation of image mosaics during a Mosaic Test, we have developed a software tool in which mosaics can be created via simple drag-and-drop functions. We refer to our novel software as the *Mosaic Test Tool*.

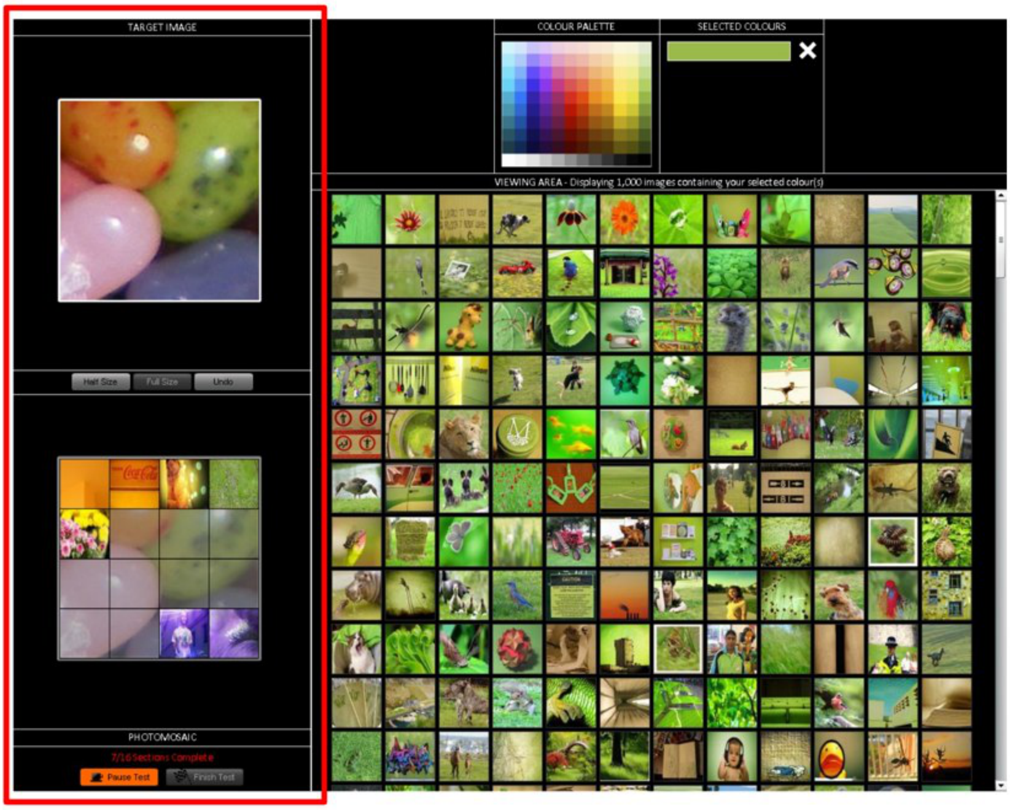


Figure 3.1: Screenshot of the Mosaic Test Tool (outlined in red) and a colour-based image retrieval system (right) under evaluation during a Mosaic Test session. The five cells across the top, and the two cells in the bottom right hand corner, of the image mosaic have been filled using images from the database.

3.2.1.1 Displaying the Mosaic Test Tool

The Mosaic Test Tool has been designed to be displayed simultaneously with the colour-based image retrieval system under evaluation. As can be seen in Figure 3.1, the Mosaic Test Tool is displayed on the left of the screen, using 40% of the available screen space. The colour-based image retrieval system being tested is then displayed on the right in the remaining 60% of the screen. This removes the need for users to constantly switch between application windows, and permits users to easily drag images from the tested colour-based image retrieval system to their image mosaic in the Mosaic Test Tool. It is important to note that the facility to export images through drag-and-drop operations is the only requirement of a colour-based image retrieval system for it to be compatible with the Mosaic Test Tool and thus the Mosaic Test.

3.2.1.2 Creating an Image Mosaic

As can be seen in Figure 3.1, the target image (i.e., the image users are trying to replicate as a mosaic) is displayed in the top half of the Mosaic Test Tool. This is so the target image can act, much like the picture on a jigsaw-puzzle box, as a guide to completing the task. The image mosaic under construction is in the lower half of the Mosaic Test Tool. It comprises a reduced opacity version of the target image overlaid with a grid which partitions the image into cells to be filled with selected images from the database. Inclusion of the reduced opacity target image beneath the grid is designed to act as a guide to help users identify the layout and proportion of colours required in the image selected to fill the corresponding cell of the mosaic.

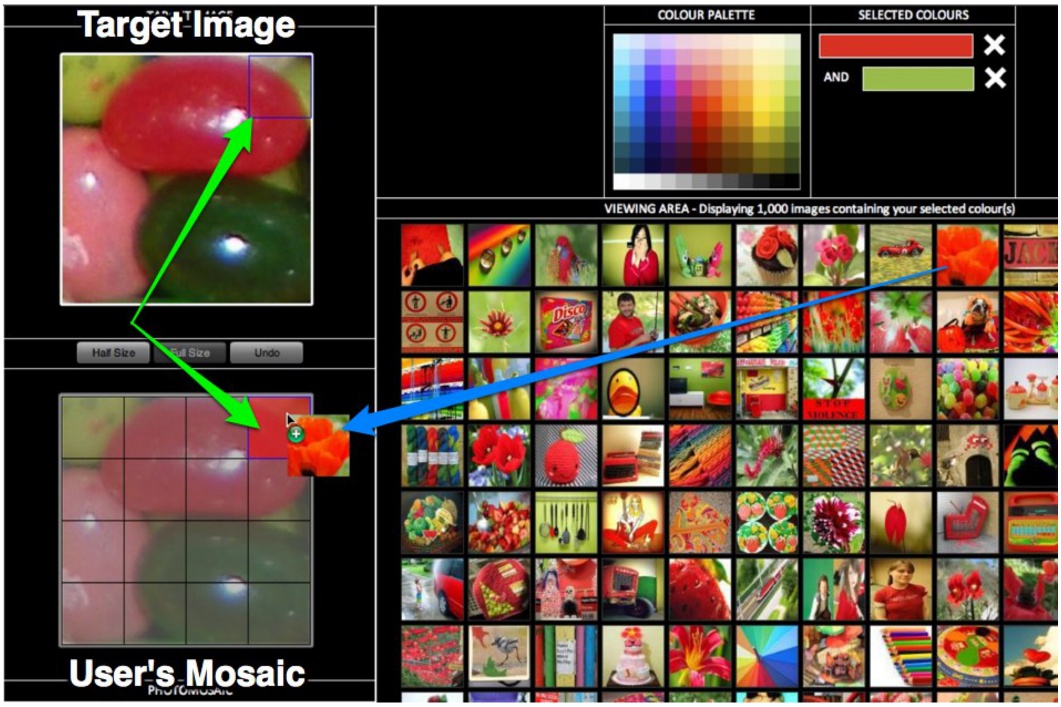
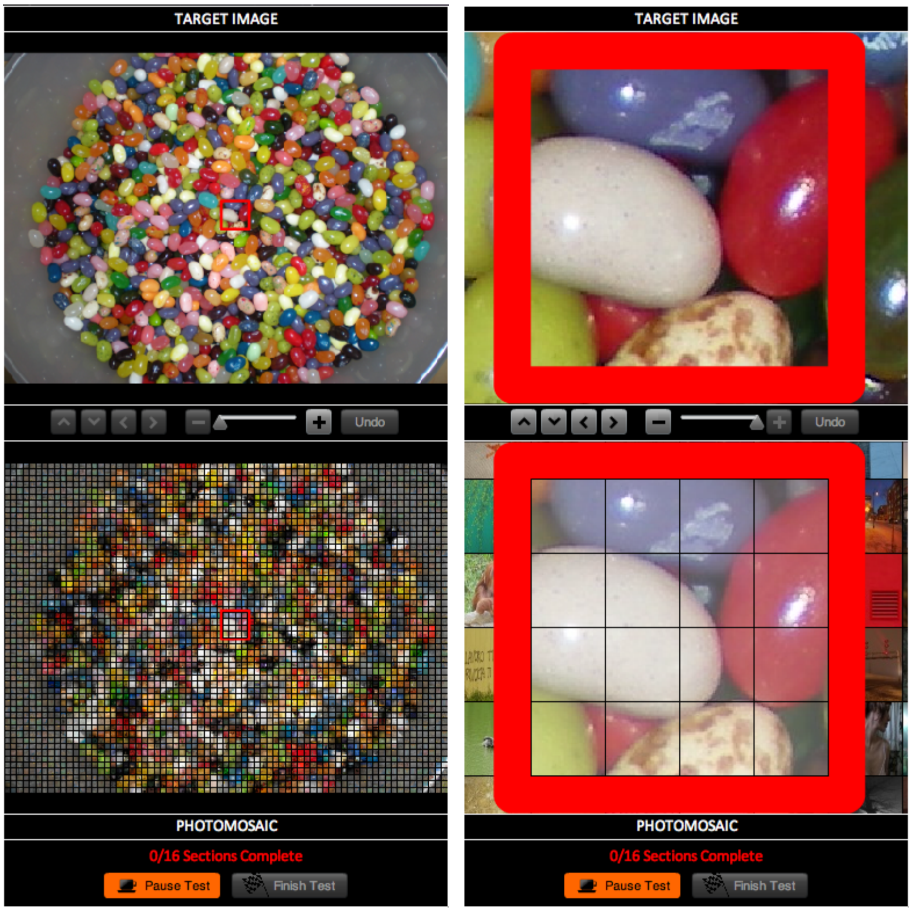


Figure 3.2: Screenshot of a Mosaic Test participant attempting to fill an image mosaic cell.

To inspect the colour content that is required of a database image, users can place the mouse cursor over a cell in the mosaic. As shown by the green arrows in Figure 3.2, this highlights the corresponding target image section in the full colour target image. Once users of the Mosaic Test Tool have found an image in the database that they believe to be suitable to fill a mosaic cell, users can drag the identified image from the colour-based image retrieval system to the desired mosaic cell in the Mosaic Test Tool (as shown by the blue arrow in Figure 3.2).

3.2.1.3 Enabling Reflection-in-action

If, upon reflection, users decide subsequently that an image dragged to a mosaic cell is not suitable, they can simply drag the image out of the mosaic cell, or revert to an earlier image for that cell using the *Undo* button located on the toolbar. There is no imposed limit regarding the number of times users can add or remove images to or from a mosaic cell. It is in this way that we believe the Mosaic Test overcomes a major drawback of existing evaluation methods - that is, it enables users to perform the creative practise of reflection-in-action [84].



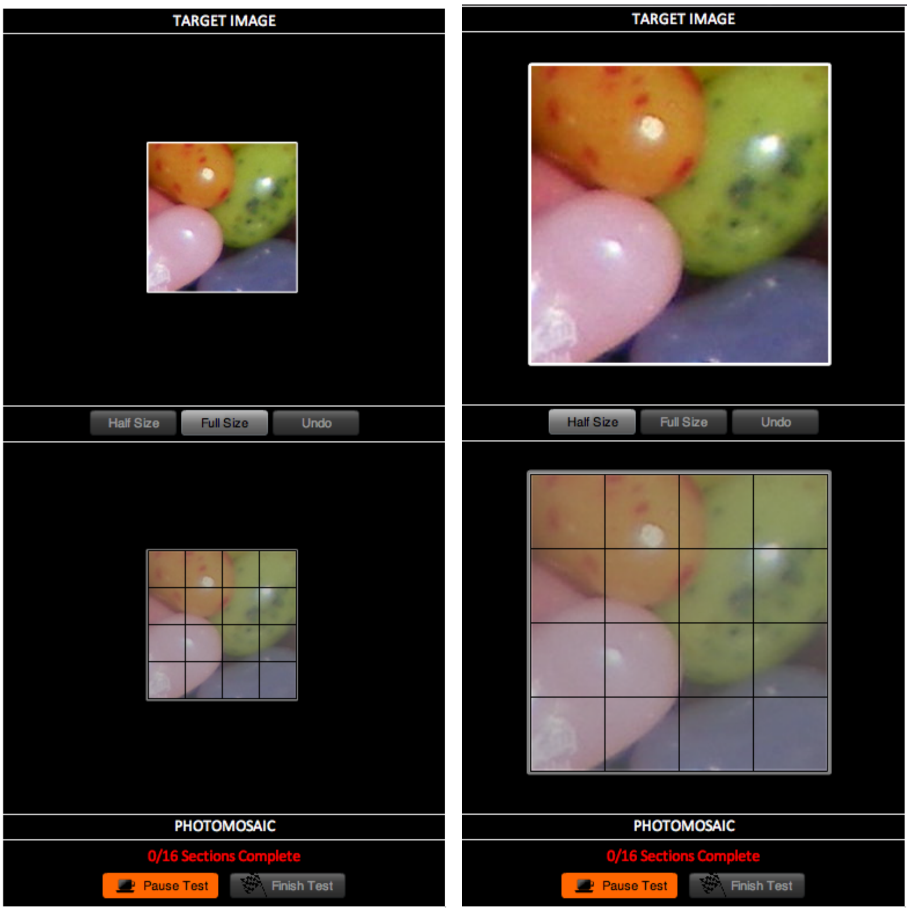
(a) Zoomed out

(b) Zoomed in

Figure 3.3: Screenshots of the initial Mosaic Test Tool with which users could zoom and pan around the entire target and mosaic images.

To aid users in assessing the accuracy of images used within their mosaics, the Mosaic Test Tool includes a facility for altering the size of the mosaic and target images. In the original Mosaic Test Tool (shown in Figure 3.3), an entire image mosaic (i.e., not just a small subsection) and corresponding target image were displayed. Users could zoom and pan around the target and mosaic images to reflect on and assess their image selections by artificially viewing the image mosaic from afar (i.e., zooming out). Users could zoom and pan using either the buttons located on the toolbar between the target image and image mosaic, or scrolling the mouse wheel for zooming and clicking (and dragging) the right mouse button to pan.

We decided to pilot our initial Mosaic Test Tool by asking five participants (3 female and 2 male, aged between 23 and 51) to complete three image mosaics each. We observed that users experienced difficulty when navigating to the subsection of the image mosaic (highlighted with a red square in Figure 3.3) that they were required to complete. Consequently, we simplified the Mosaic Test Tool to show only the mosaic cells to be completed and support just two zoom levels; *half-size* and *full-size*. Mosaics can be viewed at these zoom levels using the buttons located on the toolbar between the target and mosaic images (shown in Figure 3.4).



(a) Zoomed out (b) Zoomed in

Figure 3.4: Screenshots of the revised Mosaic Test Tool showing only the mosaic cell to be completed and supporting just two zoom levels; *half-size* (Figure 3.4a) and *full-size* (Figure 3.4b).

3.2.1.4 Completing a Mosaic Test

Located at the bottom of the Mosaic Test Tool interface are the *Pause Test* and *Finish Test* buttons. Since the time taken to complete mosaics is an important measure of the effectiveness of the colour-based image retrieval system under evaluation, the Mosaic Test can easily be paused should participants require a break for any reason - thus preventing inappropriately extended task completion times. To prevent users submitting an incomplete mosaic, the *Finish Test* button is only enabled once all mosaic cells have been filled. When participants submit their mosaic, the Mosaic Test Tool automatically records the total time taken (in seconds), as well as a bitmap of the user-generated mosaic for subsequent analysis. In the following Section, we describe how this data can be used to compare colour-based image retrieval systems.

3.2.2 Evaluating Colour-based Image Retrieval Systems

Our definition of an effective colour-based image retrieval system is one with which users can quickly and easily retrieve images from a database most relevant to their colour requirements. According to this definition, an effective colour-based image retrieval system will enable users to quickly and easily generate accurate image mosaics (when compared with the corresponding target images). Supported by the Mosaic Test Tool, the Mosaic Test measures each component of our effectiveness definition.

The Mosaic Test uses the number of seconds taken by participants to complete a mosaic as its measure of time. An effective colour-based image retrieval system will enable users to create image mosaics in fewer seconds than a less effective system. To measure the ease of use of a system for colour-based image retrieval, after completing their image mosaic participants are asked to subjectively rate the workload they experienced whilst using a tested system via the NASA-TLX workload assessment tool [31]. Users are asked to break their assessment of workload down according to six dimensions - mental demand, temporal demand, physical demand, frustration, effort and performance, and to rate each on a 20-point scale. These NASA-TLX scales are included in Appendix A. We would expect users creating a mosaic with an effective colour-based image retrieval system to experience less workload than when completing a mosaic with a less effective system.

We adopt image mosaic accuracy as a measure of relevance for the Mosaic Test. Recall from Section 2.4 that, in consulting several professional graphic designers on the subject of this research, many recalled occasions in which images comprising colours complementing a client's specification or brand were required for a project. The most relevant images in the database for these users, therefore, will be those which contain colours which match and/or compliment those used within the project for which the image is intended. This concept of colour relevance also applies when creating image mosaics. Mosaics require images comprising colours which match those contained within the target image. Given that an accurate mosaic is one which appears visually similar to its target image, the most relevant images in the database for each mosaic cell will be those which consist of colours best matching the colours within corresponding sections of the target image.

To clarify our definition of image mosaic accuracy further, let us consider Figure 3.5. The images used to create the mosaic in Figure 3.5a have colour content which closely match the colours in the corresponding areas of the target image (Figure 3.5b); more so than the images contained within the mosaic in Figure 3.5c. The images used within the mosaic depicted in Figure 3.5a are therefore *more* relevant for the project than those used in the mosaic shown in Figure 3.5c. Since the images included are more relevant, the mosaic in FIgure 3.5a appears more visually similar to the target image than the mosaic in FIgure 3.5c. It is for this reason that we believe image mosaic accuracy represents a suitable measure of relevance for the Mosaic Test. We would expect image mosaics created using an effective colour-based image retrieval system to be more accurate (or relevant) than mosaics generated using a less effective system.

(a) More accurate image mosaic. (b) The original target image. (c) Less accurate image mosaic.

Figure 3.5: Two image mosaics of varying accuracy in comparison with the original target image.

3.2.2.1 Automatically Assessing Image Mosaic Accuracy

As we have discussed previously, since no standardised image database and associated ground-truth currently exist for evaluating colour-based image retrieval, it is difficult to replicate studies and reliably compare systems. To overcome this drawback in the Mosaic Test we adopted the MIRFLICKR-25000 collection [40], which consists of 25,000 images downloaded from the photo sharing community website Flickr [45]. Furthermore, to remove the need for a ground-truth, we aimed to automate the assessment of image relevance by comparing the accuracy of user-generated mosaics against their corresponding target images.

Image mosaics are typically created automatically by a computer program that analyses and compares the colour content of a cell in the target image and images in a database [88]. Currently, there is no standardised approach to benchmarking the accuracy of image mosaics. In the following sections, we describe several measures that have been proposed previously for benchmarking the accuracy of image mosaics.

3.2.2.1.1 Tran placed the image mosaic and target image side-by-side and measured the closest physical distance (in feet and inches) at which an adjudicator could stand so that the mosaic and target images appeared identical [96]. According to this measure, the closer the distance at which the images appeared identical, the more accurate the image mosaic.

3.2.2.1.2 Nakade and Karule proposed the *Average Pixel-to-Pixel* (APP) distance for *automatically* measuring mosaic accuracy [65]. This measure computes the average L_2 distance between the colours of corresponding mosaic and target image pixels in an RGB colour space. The APP distance is expressed formally in Equation (3.1), where i is one of a total n corresponding pixels in the mosaic image M and target image T, and r, g and b are the red, green and blue colour values of a pixel.

$$APP = \frac{\sum_{i=0}^{n} \sqrt{(r_M^i - r_T^i)^2 + (g_M^i - g_T^i)^2 + (b_M^i - b_T^i)^2}}{n} \tag{3.1}$$

3.2.2.1.3 Zhang et al. defined an image mosaic accuracy measure in [103] referred to as the *multi-level colour histogram* distance. The multi-level colour histogram distance is designed to account for the phenomenon occurring in image mosaics whereby up close, the individual images appear clearly, whilst from a distance, the images collectively form the overall target image. The multi-level colour histogram distance, expressed formally in Equation 3.4, can be summarised as the absolute difference between a global and a local histogram distance calculated between an image mosaic (M) and target image (T).

The global histogram distance is calculated by measuring the L_2 distance between 64-bin colour histograms (with four values for each of the red, green and blue colour channels) extracted from the target image and image mosaic. The global histogram distance is shown in Equation 3.2, where H_i^M and H_i^T represent the i^{th} of 64 bins in the global colour histograms of the image mosaic (M) and target image (T) respectively. The local histogram distance is computed by dividing the target image and image mosaic in to 25 (5×5) blocks, computing a 64-bin colour histogram for all blocks, then measuring the average L_2 distance between the colour histograms of corresponding blocks in the target image and image mosaic. The local histogram distance is shown in Equation 3.3, where $H_i^M(j)$ and $H_i^T(j)$ represent the i^{th} bin (of 64) in the j^{th} local colour histogram (of 25) from the image mosaic (M) and target image (T) respectively.

$$d_{GCH}(M,T) = d_{L2}(M,T) = \sqrt{\sum_{i=1}^{64} (H_i^M - H_i^T)^2} \tag{3.2}$$

$$d_{LCH}(M,T) = \frac{\sum_{j=1}^{25} \sqrt{\sum_{i=1}^{64} (H_i^M(j) - H_i^T(j))^2}}{25} \tag{3.3}$$

$$d_{MLCH}(M,T) = |d_{LCH}(M,T) - d_{GCH}(M,T)| \tag{3.4}$$

3.2.2.1.4 Drawbacks of Existing Mosaic Accuracy Measures We believed that each of the mosaic accuracy measures identified above incurs at least one drawback casting doubt on its suitability for use with the Mosaic Test. Tran's [96] *manual* analysis of image mosaic accuracy is highly subjective and thus unreliable. Further to Tran's proposed accuracy measure being highly subjective, the approach also relies heavily on the vision of the human assessor which can, of course, vary greatly between people.

Interestingly, Zhang *et al.* [103] dismiss the APP measure used by Nakade and Karule [65], stating that a pixel-by-pixel comparison is inappropriate as the accuracy of an image mosaic should be derived from its overall `look-and-feel' rather than microscopic details. Another criticism of Nakade and Karule's APP distance measure is that it measures pixel distances in an RGB colour space; a colour space in which differences between colours do not reflect the differences perceived by humans [91]. Finally, the fundamental flaw with the multi-level colour histogram distance proposed by Zhang et al. [103] is that, by making the global histogram measure *worse*, you can actually improve the multi-level colour histogram distance achieved.

Given the drawbacks outlined above, we needed to find a reliable measure of image mosaic accuracy for use with the Mosaic Test. In the following Section, we describe our first user study from which we identified a mosaic accuracy measure suitable for use with the Mosaic Test.

3.3 User Study

In this Section, we describe the first user study conducted as part of our research in which 24 participants were each asked to complete three Mosaic Tests using three different colour-based image retrieval systems. There were two primary motivations for this user study; as well as comparing the Mosaic Test measures achieved by the tested colour-based image retrieval systems (the focus of Chapter 4 in this Thesis), we wanted to test whether the Mosaic Test was able to overcome the major drawbacks which we believe render existing strategies unfit for evaluating colour-based image retrieval systems and approaches - that is, the field of colour-based image retrieval lacking a standardised database with associated ground-truth, and their failure to support *reflection-in-action* [84]. Furthermore, as many image retrieval systems have been evaluated by recruiting either `expert` [75] or `non-expert' [76] users, we also wanted to measure what effect user expertise had on the effectiveness measures derived from a Mosaic Test. Participants were asked to complete a Mosaic Test using three colour-based image retrieval systems. In the following Sections, we provide describe the demographics of participants and our study procedure.

3.3.1 Participants

We recruited 24 users to participate in our study, 12 of which had previous experience working within the graphic design industry and were therefore experienced in retrieving images on the basis of colour for use in creative design projects. We recruited these expert users by contacting graphic design agencies in the local area. In this *expert* group, 7 participants were male, 5 female, and all were aged between 21 and 45. The remainder of our participants were classed as non-experts; these participants were recruited as a consequence of an advert emailed to undergraduate students at Aston University. Respondents were selected on the basis that they had little or no experience of using colour-based image retrieval systems. In the *non-expert* group, 10 participants were male, 2 female, and all were aged between 21 and 56. Every participant in the study was required to sign a consent form, included in Appendix F, confirming they had no known colour-blindness or colour-vision deficiency. Participants were also required to complete a questionnaire so that we could accurately capture data regarding their age, gender and level of expertise. This is included in Appendix A.

3.3.2 Procedure

In this Section, we describe the underlying procedure of our first study. Since the focus of this Chapter is the suitability of the Mosaic Test for meaningfully evaluating and comparing colour-based image retrieval systems and approaches (as opposed to comparing the systems tested by users), we provide only a summary of the design and implementation of colour-based image retrieval systems tested in our first study.

3.3.2.1 Image Database

The colour-based image retrieval systems tested in our first user study indexed the same image database, namely the 25,000 thumbnail images (64×64 pixels in resolution) of the MIRFLICKR-25000 collection [40]. We extracted the 120 bin colour histograms using a similar approach to that of Ibrahim *et al.* [41]. Specifically, image pixels were mapped to their closest corresponding colour in the adopted colour space - namely the 120 colours contained within the graphical colour palette of the MultiColr system [42] - by measuring the L_2 distance between the red, green and blue components of pixel and palette colours.

3.3.2.2 Colour-based Image Retrieval Systems

The first of the three systems tested adopted the query-by-colour query style. We based this system on the popular colour-based image retrieval system MultiColr [42] to represent the *state of the art* of query-by-colour systems in our study. The second system featured a mapping-based visualisation - created by reducing colour histograms extracted from database images to 2-dimensional co-ordinates for plotting thumbnails on screen - which could be interactively explored using zoom and pan tools. The third system used by study participants was essentially the mapping-based visualisation systems with the zooming facility disabled (i.e., users could only pan the visualisation displayed at full resolution). In the following Sections, we summarise the design and implementations of the three colour-based image retrieval systems tested in the study. We describe the technical aspects of these systems in greater detail in Chapter 4.

3.3.2.2.1 Query-by-colour System We adopted the design of the MultiColr system [42] for our query-by-colour system interface, mainly due to the many positive reviews it had received online (e.g., [26]). To formulate a query-by-colour, users could select multiple colours from the graphical colour palette displayed at the top of the system interface (shown in Figure 3.6). The query-by-colour system generated a 120-bin query colour histogram from the user's selected colours which was updated each time the user's query was modified (i.e., when a colour was added or removed from a query). This histogram was then compared to the colour histograms extracted from database images using the L_2 distance metric, with the 1,000 database images with a histogram most similar to the query returned by the system in the form of a scrollable list; the most relevant (i.e., closest in colour content) of which were displayed at the head of the list. Users could scroll through this list, using a scroll bar or the mouse scroll wheel, to find the image most relevant to their colour requirements. Upon finding a suitable database image, users could click and hold the left mouse button to drag a copy of the image to cells in their image mosaics.

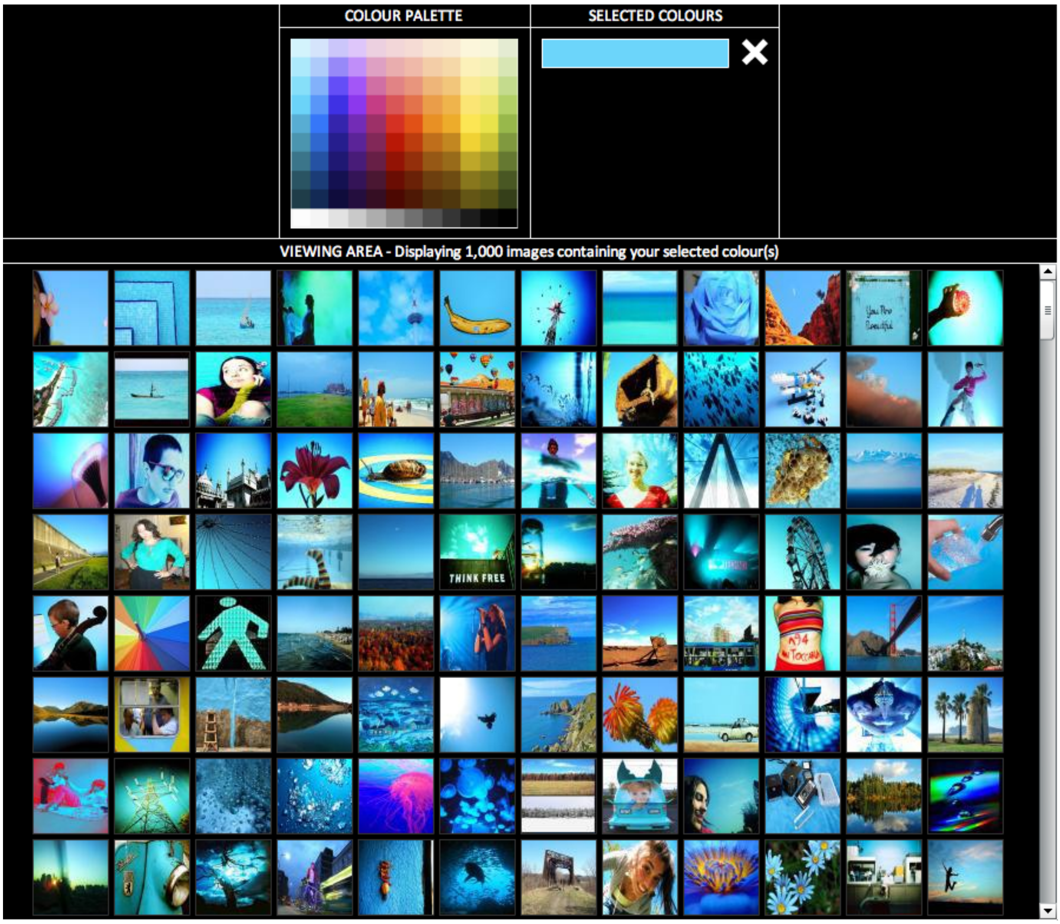
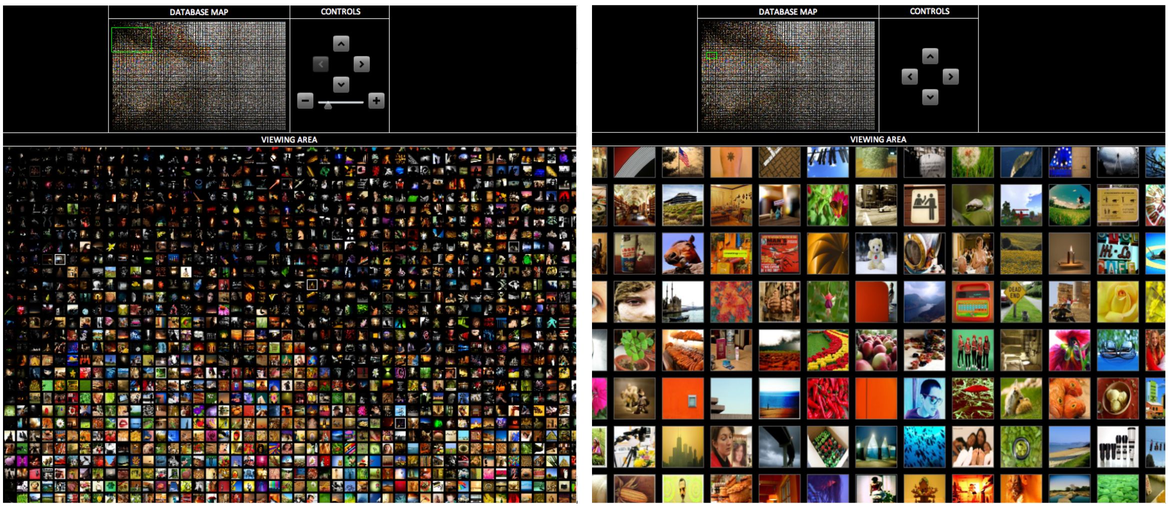


Figure 3.6: Screenshot of the query-by-colour system tested as part of our first user study.

3.3.2.2.2 Zoomable Mapping-based Visualisation System We adopted the approach of Rodden *et al.* [73] to produce the mapping-based visualisation for the two systems in our study. First, we reduced the 120-bin colour histograms of database images to just 2-dimensions via Principal Component Analysis (described previously in Section 2.5.1.1.1).

As can be seen in Figure 3.7a, users of the zoomable mapping-based visualisation system could zoom and pan using the controls located within the ``*Controls*'' pane at the top of the system interface. Users could also zoom and pan the visualisation with mouse operations, using the scroll wheel to zoom and the right mouse button (by clicking and dragging the visualisation) to pan. Panning of the visualisation was also supported via the ``*Database Map*'' component displayed at the top of the system interface; users could drag the green rectangle (representing the area of the visualisation currently being viewed in the ``*Viewing Area*'' panel) or click the left mouse button on a point of the map to pan to an alternative area of the visualisation. Finally, as with the query-by-colour system, users were required to click and hold the left mouse button whilst dragging to copy an image from the system to cells in their image mosaics.



(a) Zoomable mapping-based visualisation system (b) Pannable-only mapping-based visualisation system

Figure 3.7: Screenshots of the zoomable and pannable-only mapping-based visualisation systems tested in the first user study

3.3.2.2.3 Pannable-only Mapping-based Visualisation System For the pannable-only mapping-based visualisation system, the visualisation was the same as that used for its zoomable counterpart. On account of the fact that the zooming facility was removed for this system, the visualisation was displayed at full resolution (i.e., fully zoomed in, as can be seen in Figure 3.7b). All other user controls, (e.g., those for panning and copying images), however, were implemented exactly as per the zoomable mapping-based visualisation system.

(a) Target image 1 (b) Target image 2 (c) Target image 3

Figure 3.8: Target images recreated as image mosaics by participants of the study. Each target image is predominantly comprised of three jelly beans of varying colour and orientation.

3.3.2.3 Target Images

We chose to use the images of jelly beans shown in Figure 3.8 as target images for our Mosaic Tests. Not only do the images of jelly beans create a bright, interesting target image for participants to create in image mosaic form, but in addition it is possible for users to generate an mosaic appearing visually similar to the target image. During our pilot study of the Mosaic Test Tool (introduced previously in Section 3.2.1.3), subsections of famous artworks and major world landmarks were trialled as potential target images. We found, however, that participants had great difficulty in recreating such target images as image mosaics. As well as containing areas of intricate detail, the target images also had areas in which there are only subtle changes in colour (e.g. skin tones in paintings of faces). These slight differences in colour were mostly disregarded by the participants of the trial, resulting in the production of inaccurate and unconvincing image mosaics. Photographs of jelly beans, however, provide large areas of distinct colours, thus overcoming the difficulties experienced by participants in the pilot study. We opted to use three different target images to prevent users learning a set of suitable database images to use. The three target images were selected to have the same number of jelly beans (and thus colours), with only the colour and orientation of the jelly beans varying between the target images.

3.3.2.4 Prior to the Mosaic Tests

Prior to completing their first Mosaic Test, participants were first given written instructions explaining the concept of an image mosaic and the functionality of the Mosaic Test Tool. Each participant undertook a practice session following these written instructions, in which the participant completed an image mosaic using a small selection of relevant images. A screenshot taken from the beginning of a Mosaic Test practice session is shown in Figure 3.9.

Once participants were familiar with the functionality of the Mosaic Test Tool, and the evaluator had observed each participant completing a set of training tasks listed on a check sheet (such as dragging and removing images from the colour-based image retrieval system to the Mosaic Test Tool image mosaic), participants proceeded to the Mosaic Test for the first colour-based image retrieval system they were due to evaluate. The written instructions and `training tasks' check sheet used for practise sessions in the study can be found in Appendix A.

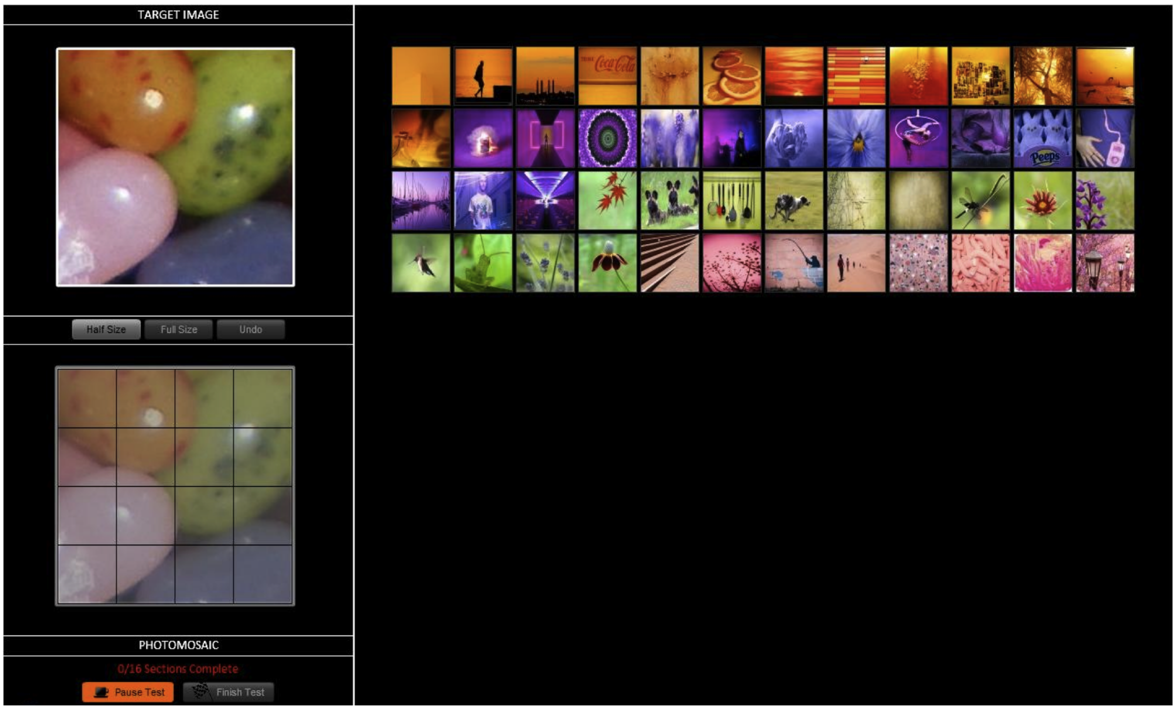


Figure 3.9: Screenshot a Mosaic Test practice session. The Mosaic Test Tool (left) is displayed simultaneously with a selection of small selection of suitable images (right) so that participants can practise creating an image mosaic.

3.3.2.5 The Mosaic Tests

Before participants created an image mosaic using each colour-based image retrieval system (i.e., they completed three image mosaics, one per a system), they were first trained, and given an opportunity to practise with, the functionality of the system they were about to use. The training material for each of the tested systems is included in Appendix B. When participants had indicated to the evaluator that they were satisfied with the controls of the system, they proceeded to the assessed Mosaic Test. After completing an image mosaic with a colour-based image retrieval system, participants were asked to complete a NASA-TLX [31] subjective workload assessment for the system they had just used. Before completing their first NASA-TLX assessment, participants were provided with instructions on how the assessment should be completed. This background information, along with the NASA-TLX assessment used, is included in Appendix A.

To ensure that results were not affected by any one of the image mosaics being more difficult to complete than the others, the order in which the target images were presented remained constant whilst the colour-based image retrieval system order was counterbalanced across participants. This also guarded against learning effects and user fatigue. Furthermore, the Mosaic Test Tool and colour-based image retrieval systems used were, for each participant, run on a Sony VAIO laptop, running Windows Vista, with a 17-inch (1600 x 900 resolution) display. This was to ensure that the colours displayed to users remained constant (as rendered colours can vary between graphic card and monitor manufacturers [58]).

3.3.2.6 After the Mosaic Tests

Once test-users had completed an image mosaic with all three colour-based image retrieval systems, participants were asked to rank their mosaics in order of perceived closeness to the corresponding target images. We could then compare the rankings made by the 24 participants to the rankings automatically generated using a variety of measures of image mosaic accuracy, to discover which of the tested measures correlated most strongly with the image mosaic accuracy assessments formed by participants - as an aim of the study was to select a measure for assessing image mosaic accuracy automatically. Furthermore, so that we could accurately assess which of the systems within a group was preferred by users, after creating their three image mosaics, participants were asked to rank their preference for each of the tested systems on a single 20-point scale. The preference assessment completed by participants is included in Appendix A.

3.4 Results

In this Section, we present our main findings related to the Mosaic Test overcoming the drawbacks of existing approaches available for evaluating colour-based image retrieval systems.

3.4.1 Automatic Mosaic Accuracy Measures

We measured the accuracy of user-generated image mosaics (in comparison with their corresponding target images) using the following feature vectors and distance metrics;

- Average Pixel-to-Pixel (APP) [65]
- Multi-level colour histogram distance (MLCH) [103]
- The MPEG-7 Colour Structure (MPEG-7 CST), Scalable Colour (MPEG-7 SCD) and Colour Layout (MPEG-7 CLD) Descriptors [59]
- Auto Colour Correlogram (ACC) [38]

For measuring the distance between the MPEG-7 CLDs of target images and image mosaics, we used the default metric outlined in [59] (and described previously in Section 2.1). For the MPEG-7

SCDs, MPEG-7 CSTs and ACCs, we generated rankings using both the L_1 and L_2 distance metrics between target image and mosaic descriptors, and also generated rankings using several different levels of quantisation (e.g., with 16, 32, 64, 128 or 256 colours). Finally, we also tested a local MPEG-7 SCD (LOCAL MPEG-7 SCD). For this, we divided the target image and image mosaic in to 16 (4×4) blocks, computed a MPEG-7 SCD for each block, then measured the average distance between the MPEG-7 SCDs of corresponding blocks in the target image and mosaic. As with the MPEG-7 SCD, we tested the LOCAL MPEG-7 SCD using the L_1 and L_2 distance metrics, and at several levels of quantisation (16, 32, 64, 128 and 256 colours). To measure the correlation between the human-assigned rankings and those generated automatically by the descriptors above, we calculated the Spearman's rank correlation coefficient (r_s) which measures the strength of correspondence between two sets of data. We proposed to adopt the mosaic accuracy measure with the highest r_s value (i.e., the image mosaic accuracy measure with rankings most strongly linked to the rankings assigned by participants of the study) for measuring image relevance in the Mosaic Test.

Rank	Descriptor	Quantisation	Metric	r_s
1	MPEG-7 CST	256	L_1	0.528
2	MPEG-7 CST	128	L_1	0.418
3	MPEG-7 SCD	256	L_2	0.407
4	MPEG-7 CST	64	L_2	0.363
5	MPEG-7 CST	64	L_1	0.355
6	MPEG-7 CST	256	L_2	0.341
7	MPEG-7 CST	128	L_2	0.33
8	APP	N/A	L_2	0.275
= 9	ACC	64	L_2	0.253
= 9	MPEG-7 SCD	16	L_2	0.253

Table 3.1: Spearman's rank correlation coefficients (r_s) between the image mosaic distance rankings made by humans and the 10 most strongly correlated rankings generated by the tested image mosaic accuracy measures.

Table 3.1 shows the Spearman's rank correlation coefficients (r_s) calculated between the human-assigned rankings and the top 10 most strongly correlated rankings generated by the tested image mosaic accuracy measures. A table showing the Spearman's rank correlation coefficients of all the tested mosaic accuracy measures is included in Appendix C. As shown in Table 3.1, rankings formed by extracting the MPEG-7 colour structure descriptor (MPEG-7 CST) from image mosaics and corresponding target images (both quantised to 256 colours), and computing the L_1 distance between the descriptors, correlated the most strongly ($r_s = 0.528$) with the participant's perceptions of image mosaic accuracy in our study.

3.4.2 Reflection-in-action Observations

As part of our user study, we observed the actions performed by participants when creating an image mosaic. It was clear that the majority of users, in both the expert and non-expert groups, *did* perform reflection-in-action [84] when assessing the relevance of images retrieved from the database. The manner in which reflection-in-action physically manifested, however, varied between users. Some users relied on the `undo' button: to assess the potentially greater suitability of an alternative image from the database relative to the previously selected image in a mosaic cell, some users would over-write the pre-existing image with the newly retrieved image; thereafter, if the newly retrieved image was considered less suitable than the pre-existing image, users would click `undo' to revert back to the pre-existing image. This observed behaviour corresponds with similar `undo'-based reflection-in-action as witnessed amongst creative individuals by Terry and Mynatt [95]. Another popular reflection-in-action strategy across users in both groups was to drag and `hover' a retrieved image over the intended image mosaic cell to inspect its suitability.

3.4.3 User Expertise Comparison

To compare the performance of the `expert' and `non-expert' participants of our study, we analysed each of the effectiveness elements measured by the Mosaic Test, namely task completion time (measured in seconds), user workload (measured using the NASA-TLX workload assessment) and mosaic accuracy (assessed by calculating the L_1 distance between the MPEG-7 colour structure descriptors of user-generated mosaics and corresponding target images) for both groups.

For the statistical comparisons reported in this Section, it is important to note that we adopt a 5% significance level. If our test data could be considered normally distributed, we adopted one-way ANOVA tests for our statistical analysis. When normality could not be assured, we used the Kruskal-Wallis test instead, which is a non-parametric alternative to ANOVA (i.e., unlike ANOVA, the Kruskal-Wallis test does not require the data to be normally distributed). Shaprio-Wilk tests, which test the null hypothesis that a sample $x_1, ..., x_n$ came from a normally distributed population, were used to verify the normality of test data.

The box-plots included throughout this Section should be interpreted as follows: the lower box line represents the first quantile (i.e., the point at which 25% of the data occurs), the bold line within the box represents the median (i.e., 50% of the data), and finally the upper box line is the third quantile (i.e., 75% of the data). The region between the first and third quantiles (i.e., the height of the box) is commonly referred to as the *interquartile range*. The maximum box-whisker length (shown as dashed lines in our plots) is 1.5 times the interquartile range. We consider any data points beyond these whiskers to be outliers (represented as circles in our box-plots).

3.4.3.1 Time Comparison

We had hypothesised previously that `expert' users (i.e., those with graphic designer experience) would require less time to complete their image mosaics than `non-experts' (i.e., participants without graphic design experience). We were surprised to observe that, as can be seen in Figure 3.10, a `non-expert' user completed an image mosaic in the quickest time (117 seconds) whilst an `expert' user took the longest to complete the task (2,546 seconds). The mean time taken by `expert' participants ($\mu$ = 959.97 seconds) was also slightly greater than the `non-expert' (μ = 864.75 seconds) participants. Kruskal-Wallis tests on the time data revealed no significant differences between the times taken by participants across the groups, both when the outliers - displayed as circles in Figure 3.10 - were included ($\chi^2_{(1)} = 0.4341, p = 0.51$) and omitted ($\chi^2_{(1)} = 0.3255, p = 0.5683$).

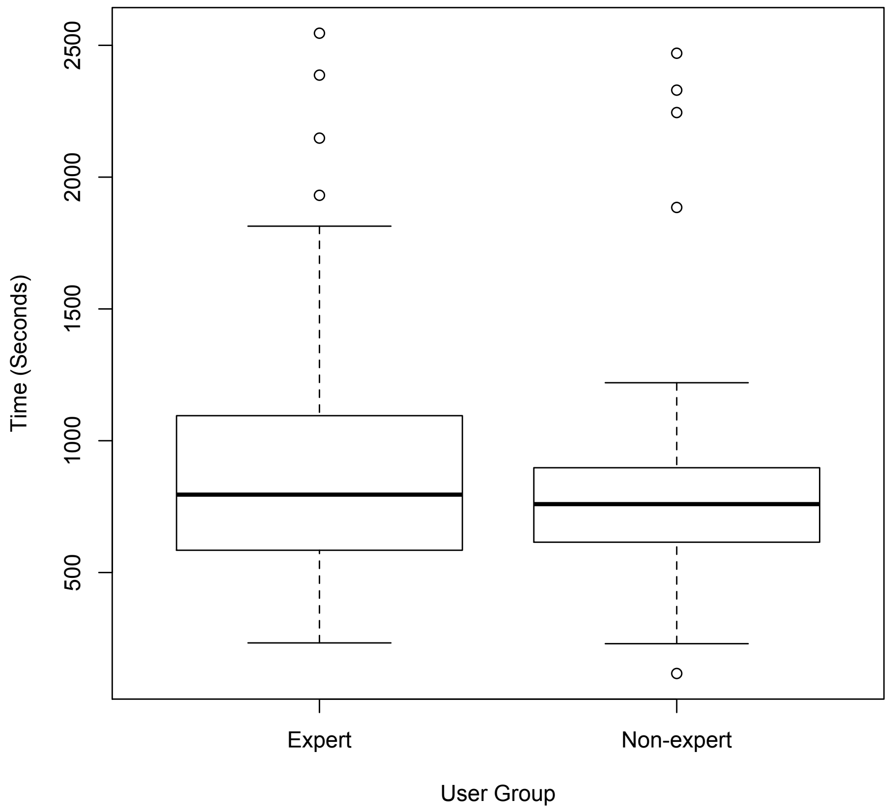


Figure 3.10: Box-plots of the times required by study participants in the `expert' and `non-expert' user groups to complete their image mosaics.

3.4.3.2 Mosaic Accuracy Comparison

As described previously in Section 3.4.1, we adopted the L_1 distance between the MPEG-7 colour structure descriptors of user-generated image mosaics and its corresponding target image for automatically assessing the relevance of images retrieved by users during a Mosaic Test. As illustrated in Figure 3.11, the accuracy of image mosaics generated by participants in our `expert' group varied greatly, with 'expert' participants creating mosaics closest ($L_1 =$ 1,898) and furthest away ($L_1 =$ 3,947) from the corresponding target images. Recall our hypothesis that `expert' participants of the study would create more relevant image mosaics (i.e., image mosaics that were visually closer to the initial target image) than `non-expert' participants. Results from the study actually show that the mean $L_1$ distance between the MPEG-7 colour structure descriptors of target images and image mosaics generated by `non-expert' participants of the study ($\mu =$ 2,652.3) was lower than the mean L_1 distance between target images and image mosaics created by `expert' participants ($\mu =$ 2,772.8). A one-way ANOVA test to compare the accuracy measures for the two groups found no significant difference between the two groups ($F_{1,70} = 1.31, p = 0.26$).

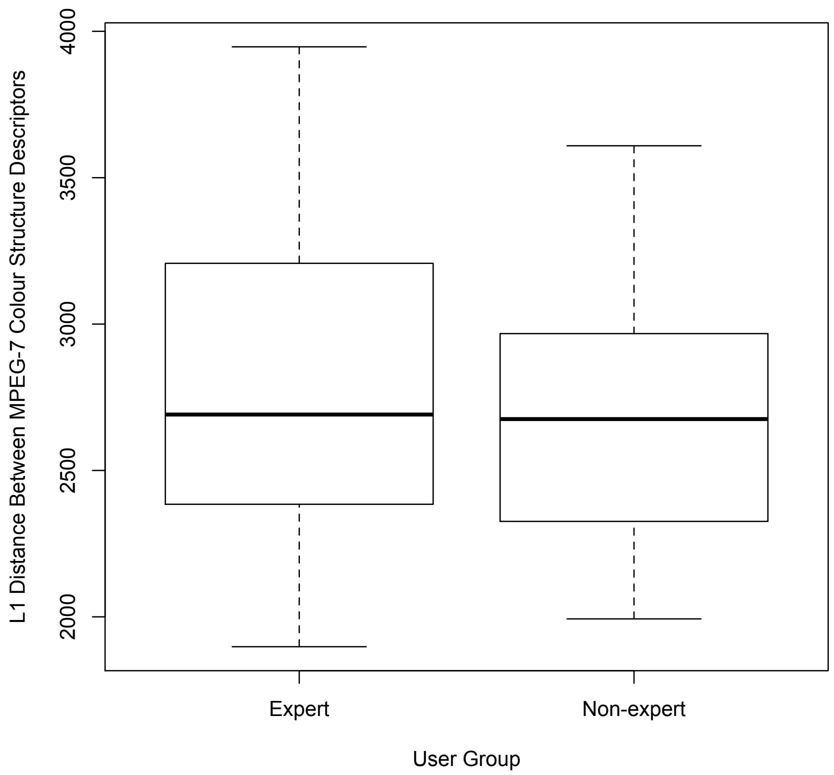


Figure 3.11: Box-plots of the accuracy of image mosaics (i.e., the L_1 distance between the MPEG-7 Colour Structure descriptors of generated mosaics and their corresponding target images) created by participants in the `expert` and `non-expert` user groups using the three colour-based image retrieval systems.

3.4.3.3 Workload Comparison

The next aspect of colour-based image retrieval effectiveness measured by the Mosaic Test is user-perceived workload. Figure 3.12 shows that study participants in our `expert' group experienced more varied levels of workload than the study participants in our `non-expert' group, with `expert' users registering both the lowest (1.83) and highest (16.17) NASA-TLX overall workload ratings. Contrary to our initial hypothesis, the average workload experienced by the `expert' users in our study ($\mu = 10.498$) was marginally higher than the overall workload experienced by our `non-expert' participants ($\mu = 10$). A one-way ANOVA test did not find the difference in overall workload experienced by the `expert' and `non-expert' user groups to be significant ($F_{1,70} = 0.49, p = 0.486$). It is also important to note that no significant differences were found between the `expert' and `non-expert' user groups for any of the individual workload elements (i.e., mental, physical, and temporal demands, frustration, effort, and performance).

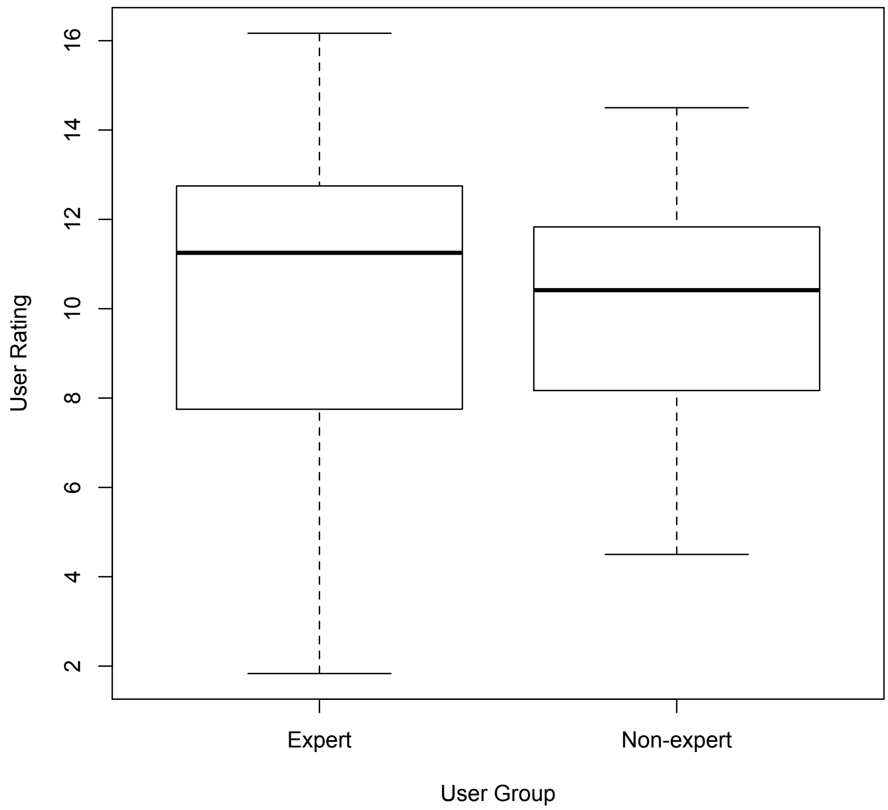


Figure 3.12: Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants in the `expert` and `non-expert` user groups whilst using the three colour-based image retrieval systems to create their image mosaics.

3.5 Discussion

In this Section, we discuss the findings of our first user study that support our claims that the Mosaic Test overcomes the drawbacks of existing evaluation techniques for assessing colour-based image retrieval systems - namely, their requirement of a standardised image database with associated ground-truth and/or failure to support reflection-in-action.

3.5.1 Automatically Measuring Mosaic Accuracy

The results of our study presented in Section 3.4.1 show that by measuring the L_1 distance between the MPEG-7 colour structure descriptors (MPEG-7 CSTs) of target images and user-generated image mosaics (reduced to 256 possible colours), the Mosaic Test can automatically calculate the relevance of retrieved images, according to the low-level feature of colour, in a manner that appears to correlate with human perception. Thus, assuming MPEG-7 CST as a validated measure for the accuracy of selected images, the Mosaic Test overcomes a drawback with the precision and recall and category search tasks measures; their requirement of a standardised image database with associated ground-truth.

3.5.2 Reflection-in-action

As reported in our results, we observed the actions performed by participants when creating their image mosaics. It was clear that the majority of users, in both the expert and non-expert groups, *did* perform reflection-in-action when assessing the relevance of images retrieved from the database. Irrespective of their chosen approach, the fact that participants were observed iteratively refining their image selections in this way demonstrates that the Mosaic Test, unlike precision and recall and the category and target search evaluation strategies, *does* support the creative image selection process of reflection-in-action.

3.5.3 User Expertise Comparison

The results of our study indicate that there is no significant differences between the time, workload and mosaic accuracy measures achieved by `expert' and `non-expert' participants when creating image mosaics with three colour-based image retrieval systems. As a result, when using the Mosaic Test to evaluate the effectiveness colour-based image retrieval systems, it is not necessary to recruit `expert' users for testing - often an expensive, difficult to plan, and most of all time-consuming task. Instead, a sample population of computer-literate participants can be tested, thus overcoming the difficulties of recruiting `expert' users. This is another clear advantage of the Mosaic Test.

3.5.4 Limitations of the Mosaic Test

The findings of our study suggest that the Mosaic Test overcomes the drawbacks associated with existing approaches for evaluating colour-based image retrieval systems. For balance, however, we should also report the limitations of our findings and thus the Mosaic Test. As per many user-based studies, the number of participants recruited to the study was relatively small and, to substantiate our Mosaic Test reliability claims further, we would need to perform the study again with many more users. Given that some participants of the study required approximately 40 minutes to complete their image mosaics, however, increasing participation significantly will also dramatically increase the time needed to undertake a study. A possible solution to this problem would be to reduce the number of mosaic cells participants are asked to complete and increase the participation rate via crowd-sourcing or distribution of a free mobile phone application - possibilities we discuss further in Section 6.2.1.

Another limitation of our Mosaic Test evaluation is the manner in which we selected the image mosaic accuracy measure. Since completing three separate Mosaic Tests proved to be a prolonged exercise, we felt it would be unfair on participants - as well as detrimental to the reliability of the human rankings obtained - to also ask users to rank a large set of image mosaics in order of accuracy (that is, in comparison with their corresponding target images) upon completing their three image mosaics. We therefore only asked participants to rank their own image mosaics in order of accuracy (i.e., each user ranked the three image mosaics he or she had just created), giving us just one set of human rankings. Since we were only able to compare the rankings generated with each of the tested mosaic accuracy measures against a single set of human rankings, we can not rule out that the correlations achieved were, in fact, false positives or, indeed, false negatives (i.e., the rankings assigned by the mosaic accuracy measures correlated, or did not correlate, with the human rankings by coincidence). It could also be that the accuracy ratings assigned by users could have been subject to bias based on which system he or she preferred most for completing the task. In order to prove beyond doubt that the L_1 distance between the MPEG-7 colour structure descriptors (MPEG-7 CSTs) of target images and user-generated image mosaics (reduced to 256 possible colours) is the most accurate mosaic accuracy measure available, a further study is needed. To overcome the limitations of our study, a much greater number of participants should be recruited to rank more sets of pre-generated mosaics (i.e., image mosaics not created by participants of the study).

Unfortunately, given the limited time and resources at our disposal, it was not feasible to conduct a further study to fully vindicate our selected mosaic accuracy measure. As our comparison of image mosaic accuracy measures reported in this Chapter is the only such evaluation to have been conducted, however, we opted to select the accuracy measure which correlated most strongly with the human assigned rankings in our study for the Mosaic Test.

3.6 Conclusion

Prior to our research, the field of colour-based image retrieval did not have a standardised strategy for evaluation. Since no standardised image database with associated ground-truth existed for use in colour-based image retrieval evaluations, the precision and recall measures could not be adopted. Furthermore, the interactive target and category search task, alongside precision and recall, do not support *reflection-in-action*, the iterative image selection process often adopted by creative individuals (e.g., graphic, fashion and interior designers). Consequently, very few studies have been conducted previously assessing the effectiveness of systems and/or approaches for retrieving images from a database by colour. In this Chapter, we have described our solution to the lack of a reliable method for evaluating colour-based image retrieval systems - the *Mosaic Test*. The Mosaic Test is a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system(s) under evaluation. The time and users' perception of the workload required to complete this creative task, as well as the visual accuracy of their image mosaics (in comparison with the initial target images), are used to assess the effectiveness of the system being tested.

The findings from our first user study, in which we evaluated the Mosaic Test using twelve `expert' and twelve `non-expert' participants (users with and without graphic design experience respectively), have confirmed that the Mosaic Test overcomes the two major drawbacks associated with previous evaluation methods: in addition to providing valuable effectiveness data relating to efficiency and user workload (by recording the task time and asking users to complete the NASA-TLX workload assessment tool), the Mosaic Test enables participants to reflect on the relevance of retrieved images within the context of their image mosaic (i.e., to perform reflection-in-action [84]), and automatically measures the relevance of retrieved images, by computing MPEG-7 colour structure descriptors (from the user-generated image mosaics and target images, quantised to 256 possible colours) and calculating the L_1 distance between them. The results of our study also show that participants in a Mosaic Test need not be `expert' users (i.e., users who retrieve images based on colour content on a regular basis) in order to use the test to reliably evaluate the effectiveness of colour-based image retrieval systems. This is important as it removes difficulties, such as time and finance, often associated with recruiting `expert' users for software testing.

Based on the findings presented in this Chapter, we believe we have fulfilled an aim of this research; *to show that the Mosaic Test provides a reliable mechanism by which to meaningfully and systematically evaluate and compare colour-based image retrieval systems.* Furthermore, we believe that our Mosaic Test is a significant contribution to the field of colour-based image retrieval, allowing us and others in the future to *meaningfully* evaluate and compare the effectiveness of colour-based image retrieval systems to further advance the field. In the following Chapter, we detail the system comparison aspect of our first user study.

Chapter 4

Comparing Query-by-colour and a Mapping-based Visualisation for Colour-based Image Retrieval

4.1 Introduction

In this Chapter, we turn our attention to the system comparison aspect of our first user study - comparing the effectiveness of the query-by-colour query style and a mapping-based visualisation for colour-based image retrieval, and assessing the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases, using the Mosaic Test. Recall from the previous Chapter how 24 participants were recruited for evaluating three colour-based image retrieval systems using our Mosaic Test. The first of these systems was based on the MultiColr system [42], adopting the query-by-colour query style to facilitate colour-based image retrieval. The second system tested adopted a mapping-based visualisation for colour-based image retrieval (recall that *mapping-based* is a style of similarity-based visualisation) which users could interactively explore using zooming and panning tools. These two systems were included in the study so that we could fulfil part of our research aim *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*; by directly comparing the query-by-colour query style with a mapping-based visualisation. The third colour-based image retrieval system tested in our first user study featured the same mapping-based visualisation as the second system. For this third system, however, we removed the zoom facility so that users could only pan the visualisation to search for images. Removing this functionality from the third system enabled us to meet another of our research aims; *to assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases*.

4.1.1 Chapter Outline

As the design and procedure of our first user study has already been comprehensively described in Section 3.3 of the previous Chapter, we include here only details of how the test image database used was indexed (Section 4.2) and the manner in which the tested colour-based image retrieval systems were implemented (Sections 4.3, 4.4 and 4.5). In Section 4.6, we outline our hypotheses for our comparison of the tested colour-based image retrieval systems. In Section 4.7, we compare the Mosaic Test measures and user preference ratings recorded for each of the tested systems, before discussing our findings in Section 4.8. Finally, Section 4.9 concludes this Chapter.

4.2 Image Database

To ensure parity across our system comparison, each of the tested systems indexed the 25,000 64×64 pixel thumbnails of the MIR-FLICKR 25000 image collection [40] in the same way - that is, using 120 bin colour histograms extracted from database images. We extracted the 120 bin colour histograms using a similar approach to that of Ibrahim *et al.* [41]. Specifically, image pixels were mapped to their closest corresponding colour in the adopted colour space - namely the 120 colours contained within the graphical colour palette of the MultiColr system [42] (shown in Figure 4.1) - by measuring the L_2 distance between the red, green and blue components of pixel and palette colours. The 120 bin colour histogram was formulated by counting the total number of image pixels mapped to each of the 120 palette colours, so that the n^{th} histogram bin contained a count of the pixels that had been mapped to the n^{th} palette colour. We chose to use the 120 colours of MultiColr for our colour space because, as described in the following Section, our query-by-colour system interface was based on that of MultiColr.

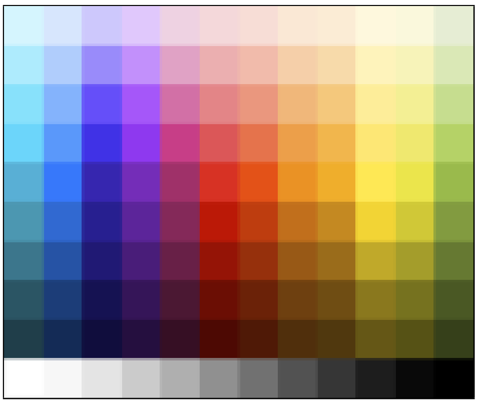


Figure 4.1: Graphical colour palette of the MultiColr system [42], comprising the 120 colours used to index the MIR-FLICKR 25000 image collection [40] with 120 bin colour histograms.

4.3 Query-by-colour System

We adopted the design of the MultiColr system [42] for our query-by-colour system interface, mainly due to the many positive reviews it had received online (e.g., [26]). To formulate a query-by-colour, users could select multiple colours from the graphical colour palette displayed at the top of the system interface. As can be seen in Figure 4.2, the colours comprising a user's current query were displayed to the right of the graphical colour palette (under the label ``*Selected Colours*''). Colours could be removed from the query by clicking the large white cross located adjacent to the required colour swatch. As per MultiColr, users could select a maximum of five colours from the graphical colour palette to query the database and could select the same colour multiple times to increase the relative prominence of that colour.

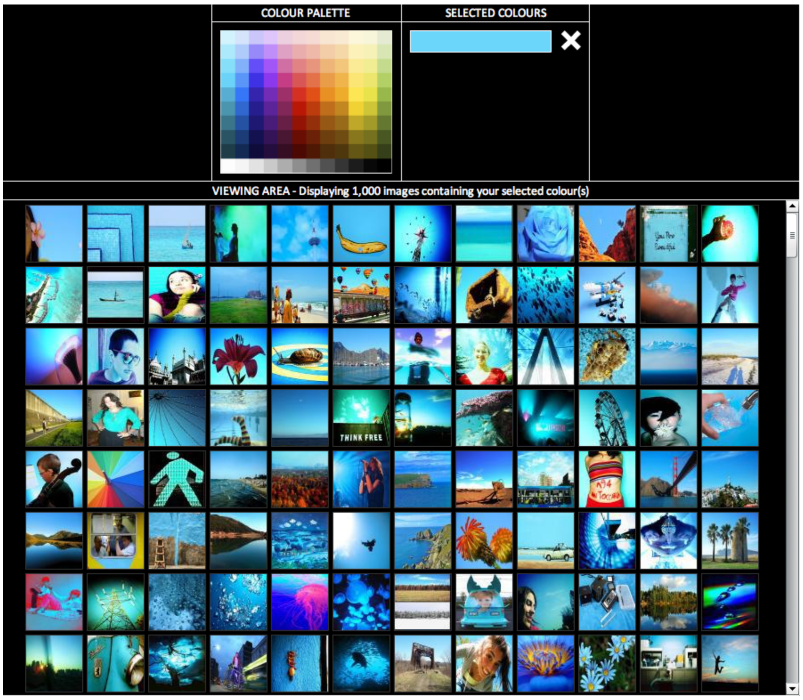


Figure 4.2: Screenshot of the query-by-colour system. The system has returned the database images which contain the most pixels with the user's selected cyan colour.

The query-by-colour system generated a 120-bin query colour histogram from the user's selected colours which was updated each time the user's query was modified (i.e., when a colour was added or removed from a query). This histogram was then compared to the colour histograms extracted from database images. The weight of each query histogram bin was calculated using the formula shown in Equation 4.1, where H_c is the histogram bin for colour c, t is the total number of colours selected and n is the number of times colour c occurs within the selection (as users had the ability to select the same colour more than once). It is important to note that the order in which users selected colours was not considered by the system during the formulation of the query histogram (i.e., it was not a ranked calculation).

$$H_c = \frac{1}{t} \times n \tag{4.1}$$

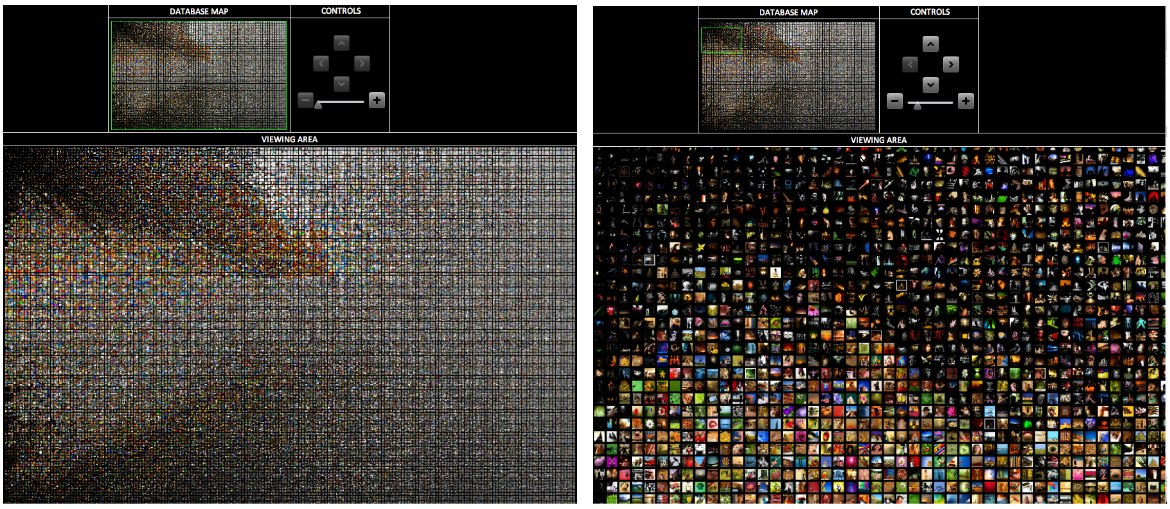
For each query-by-colour, the system compared the extracted query colour histogram to the colour histograms of the 25,000 database images using the L_2 distance metric. The 1,000 database images with a histogram most similar to the query were returned by the system in the form of a scrollable list, with the most relevant (i.e., closest in colour content) images displayed at the head of the list. Users could scroll through this list, using a scroll bar or the mouse scroll wheel, to find the image most relevant to their colour requirements. Upon finding a suitable database image, users could click and hold the left mouse button to drag a copy of the image to cells in their image mosaics.

4.4 Zoomable Mapping-based Visualisation System

We adopted the mapping-based visualisation style for our user study on account of its popularity (i.e., it is more often included for use in image retrieval systems) in comparison with the clustering-based and graph-based visualisation styles, and so that we could assess the importance of zooming. Recall that mapping-based visualisations use dimensionality-reduction techniques to map the often high-dimensional image feature vectors to just 2-dimensions, used as co-ordinates for plotting database images on screen. Users can then explore the resultant database visualisations via zooming and panning tools.

We implemented the approach of Rodden *et al.* [73] to produce the mapping-based visualisation for the two systems in our study. First, we reduced the 120 bin colour histograms of database images to just 2-dimensions via Principal Component Analysis (described previously in Section 2.5.1.1.1). We opted to use this dimensionality-reduction technique because it is much less computationally expensive than iterative techniques such as multi-dimensional scaling. For convenience, we translated the derived co-ordinates so that all had positive x and y values. Since Rodden *et al.* [73] has previously demonstrated that users dislike image overlap that can occur within mapping-based visualisations, we fitted the databases images to a grid of 100×250 cells using a variation of Rodden *et al.'s bump* strategy (described previously in Section 2.5.1.2.1). We first mapped images to cells on the grid according to the 2-dimensional co-ordinates derived through Principal Component Analysis. Our spreading algorithm then located the grid cell with the highest number of assigned images and initiated a spiral search (emanating from this original cell) in order to locate the closest free cell. The images in the line of cells between the original cell and the next closest cell are all moved outwards (from the original cell at the centre) by one cell, with the image to be spread from the original cell moved to the newly vacated adjacent cell. Images are spread from grid cells until just one remains. The order in which images are spread from the grid cells is based on the distance between their derived co-ordinates and the grid cell centre (i.e., the image with derived co-ordinated closest to the grid cell centre is the one that remains). The result of the spreading algorithm is each grid cell contains exactly one image.

We used OpenZoom [27] so that our mapping-based visualisation could be interactively zoomed and panned by study participants. OpenZoom is a software development kit for the Adobe® Flash® platform which allows developers to create applications in which high resolution images can be interactively explored via zoom and pan operations. As can be seen in Figures 4.3a and 4.3b, users of the zoomable mapping-based visualisation system could zoom and pan using the controls located within the ``*Controls*'' pane at the top of the system interface. Users could also zoom and pan the visualisation with mouse operations, using the scroll wheel to zoom and the right mouse button (by clicking and dragging the visualisation) to pan. Panning of the visualisation was also supported via the ``*Database Map*'' component displayed at the top of the system interface; users could drag the green rectangle (representing the area of the visualisation currently being viewed in the ``*Viewing Area*'' panel) or click the left mouse button on a point of the map to pan to an alternative area of the visualisation. Finally, as with the query-by-colour system, users were required to click and hold the left mouse button whilst dragging to copy an image from the system to cells in their image mosaics.



(a) Before a ``zoom in'' operation

(b) After a ``zoom in'' operation

Figure 4.3: Screenshots of the zoomable mapping-based visualisation system before and after a ``zoom in'' operation has been applied.

4.5 Pannable-only Mapping-based Visualisation System

For the pannable-only mapping-based visualisation system, the visualisation was the same as that used for its zoomable counterpart. On account of the fact that the zooming facility was removed for this system, the visualisation was displayed at full resolution (i.e., fully zoomed in, as can be seen in Figure 4.4). All other user controls, (e.g., those for panning and copying images), however, were implemented exactly as per the zoomable mapping-based visualisation system.

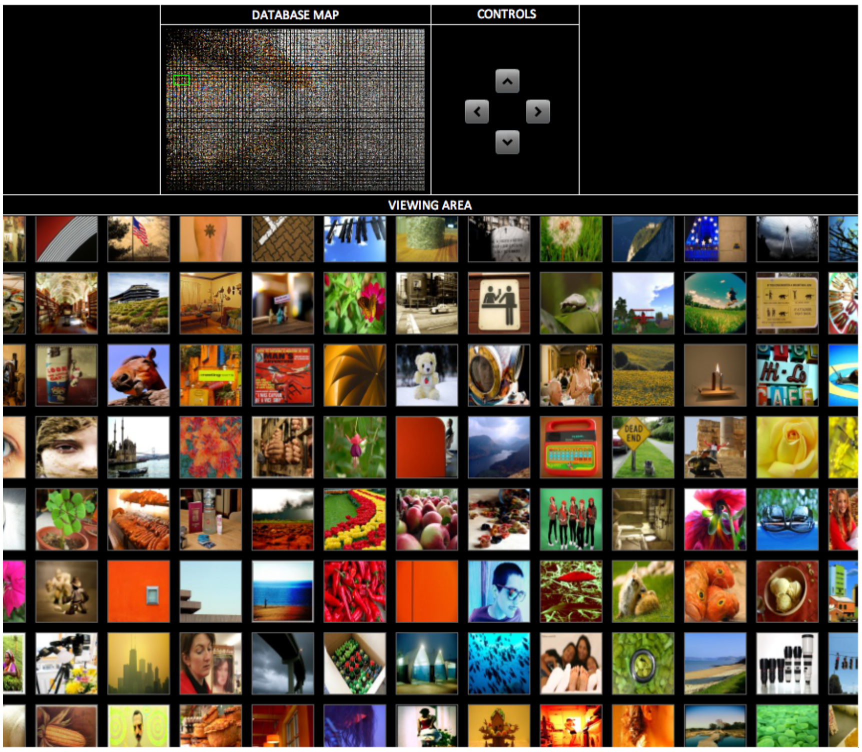


Figure 4.4: Screenshot of the pannable-only mapping-based visualisation system.

4.6 Hypothesis

Prior to this research, there had been no reported comparison of systems adopting either the query-by-colour query style or a mapping-based visualisation for colour-based image retrieval. Consequently, we were unsure as to which would prove to be the most effective for retrieving images on the basis of colour. For our comparison of the zoomable and pannable-only mapping-based visualisation systems, we hypothesised that the zoomable system would be more effective than its pannable-only counterpart for colour-based image retrieval (i.e., users of the zoomable mapping-based visualisation system would be able to generate more accurate image mosaics, more quickly, and with less self-perceived workload). This was hypothesised because we believed that the zoomable system would not only enable study participants to navigate to interesting areas of the visualisation more quickly and with fewer interactions, but also because the zooming facility would enable users to review more images more quickly - as the number of images that are displayed on screen at any one time could be manually configured by users. In the following Section, we present the findings from our systems comparison, and discuss the degree to which the hypothesis defined in this Section matched the observed results.

4.7 Results

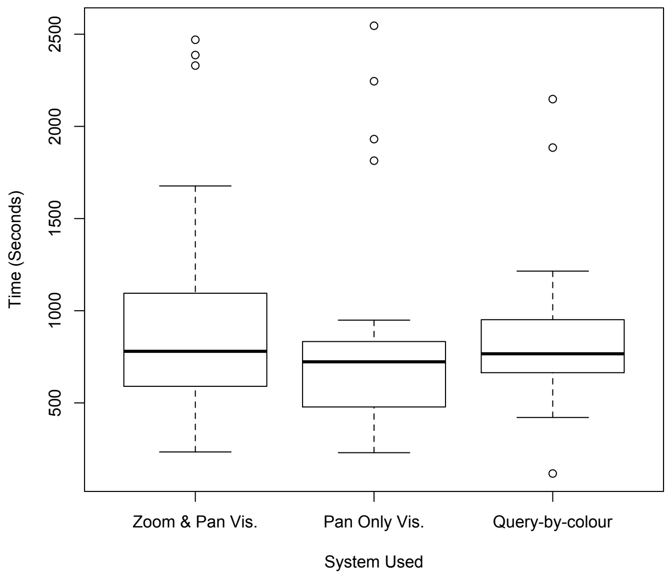
In this Section we present the results from our first user study as they relate to the comparison of the three tested colour-based image retrieval systems. We compare the Mosaic Test measures recorded for the tested systems, namely the time required by users to complete their image mosaics, the accuracy of users' generated mosaics, and the subjective workload experienced by users whilst creating them. Furthermore, we also compare the preference ratings that users assigned to each of the tested systems upon completing their three image mosaics.

For the statistical comparisons reported in this Section, it is important to note that we adopt a 5% significance level. If our test data could be considered normally distributed, we adopted one-way ANOVA tests for our statistical analysis. When normality could not be assured, we used the Kruskal-Wallis test instead, which is a non-parametric alternative to ANOVA (i.e., unlike ANOVA, the Kruskal-Wallis test does not require the data to be normally distributed). Shaprio-Wilk tests, which test the null hypothesis that a sample $x_1, ..., x_n$ came from a normally distributed population, were used to verify the normality of test data.

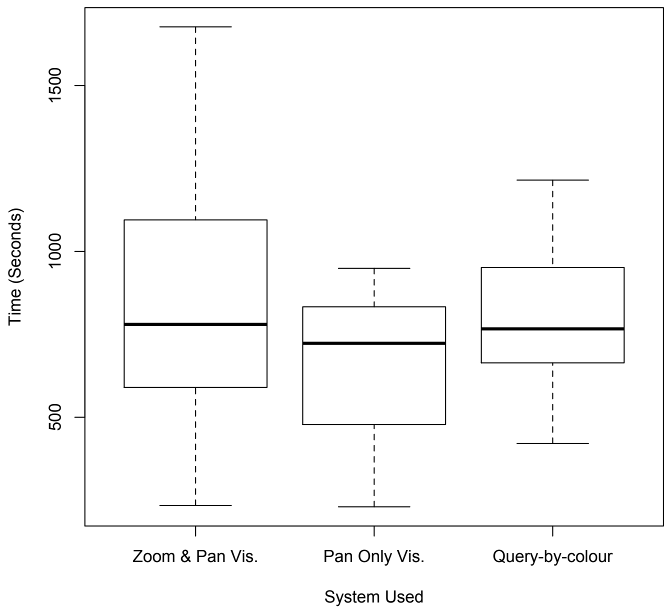
The box-plots included throughout this Section should be interpreted as follows: the lower box line represents the first quantile (i.e., the point at which 25% of the data occurs), the bold line within the box represents the median (i.e., 50% of the data), and finally the upper box line is the third quantile (i.e., 75% of the data). The region between the first and third quantiles (i.e., the height of the box) is commonly referred to as the *interquartile range*. The maximum box-whisker length (shown as dashed lines in our plots) is 1.5 times the interquartile range. We consider any data points beyond these whiskers to be outliers (represented as circles in our box-plots).

4.7.1 Time

As can be seen in Figure 4.5a, the times taken by participants to generate their image mosaics ranged from a mere 117 seconds using the query-by-colour system to 2,546 seconds (almost 43 minutes) using the pannable-only mapping-based visualisation system. Upon statistically analysing the time data, however, we identified these times as outliers (circles in our box-plots). With the outliers removed, both the quickest and slowest times were both recorded by participants using the zoomable mapping-based visualisation system (234 and 1,677 seconds respectively, shown in Figure 4.5b). A Kruskal-Wallis test on the full time data revealed no significant difference between the times taken by participants to create their image mosaics across the three systems ($\chi^2_{(2)} = 1.8717, p = 0.3922$). When we omitted the outliers from the data, however, a one-way ANOVA test revealed a significant difference between the times recorded for the three systems ($F_{(2,62)} = 4.127, p = 0.021$), with a post-hoc Tukey's HSD (Honestly Significant Difference) test - used in conjunction with an ANOVA to find which means are significantly different from one another - revealing the pannable-only mapping-based visualisation system to be significantly faster than its zoomable counterpart ($p = 0.043$) *and* the query-by-colour system ($p = 0.034$) for creating image mosaics.



(a) Box-plots *with* outliers (shown as circles).



(b) Box-plots *without* outliers.

Figure 4.5: Box-plots of the time required by participants to complete an image mosaic using each of the three systems tested in our first user study with and without outliers.

4.7.2 Image Mosaic Accuracy

With both the outlier (shown in Figure 4.6 as a circle) included ($F_{(2,69)} = 2.1, p = 0.131$) and excluded ($F_{(2,68)} = 2.87, p = 0.0636$), one-way ANOVA tests revealed no significant difference between the accuracy of mosaics across the three systems. Interestingly, as Figure 4.6 shows, the most accurate (or relevant) image mosaic was generated using the zoomable mapping-based visualisation system ($L_1 = 1,898$), whilst the least accurate image mosaic was created using the query-by-colour system ($L_1 = 3,947$).

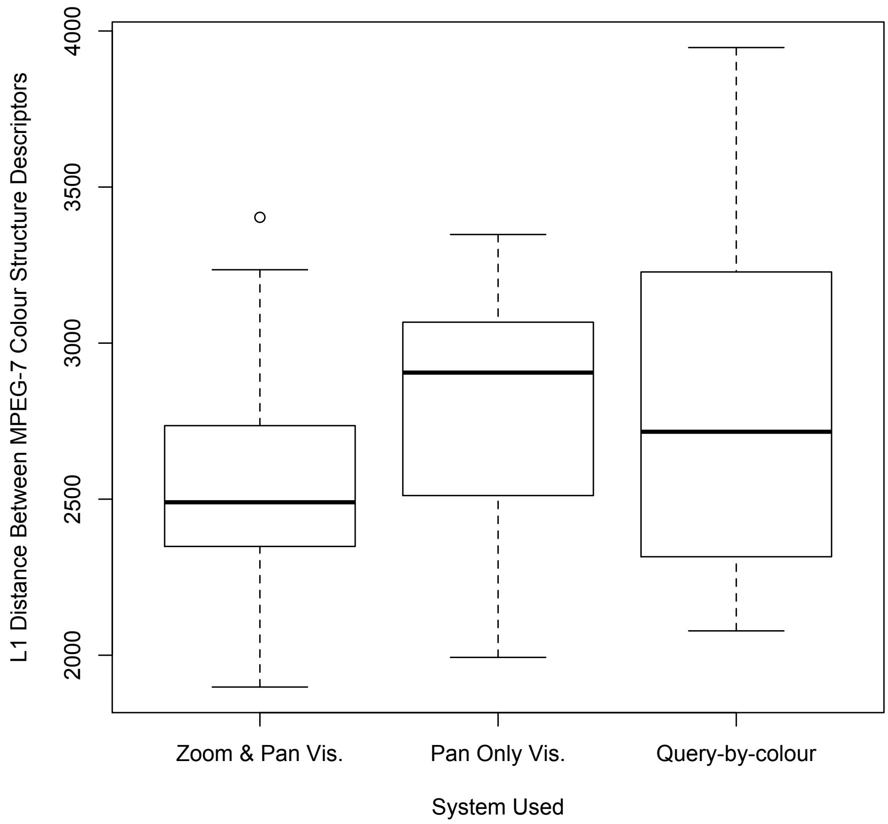


Figure 4.6: Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants of our first user study with each of the tested systems, and their corresponding target images.

4.7.3 Workload

Upon removing the outlier (circled in Figure 4.7) from the pannable-only mapping-based visualisation system's ratings, we did encounter a significant difference when comparing the overall workload (i.e., the mean of the six workload dimensions - mental, physical and temporal demand, performance, effort and frustration) experienced by participants whilst using the three systems ($F_{(2,67)} = 10.31, p = 0.0001$). Post-hoc Tukey's HSD tests showed that the overall workload ratings reported by participants when using the query-by-colour system were significantly lower than those reported when using either of the systems featuring a mapping-based visualisation. Contrary to the finding above, however, Figure 4.7 shows that the worst (highest) overall workload rating was actually assigned to the query-by-colour system (16.17). It is important to note that the difference between the overall workload ratings reported by participants using the zoomable and pannable-only mapping-based visualisation systems were not found to be significant. In the following Sections, we report on the workload dimensions for which we observed a significant difference in the results recorded across the tested systems.

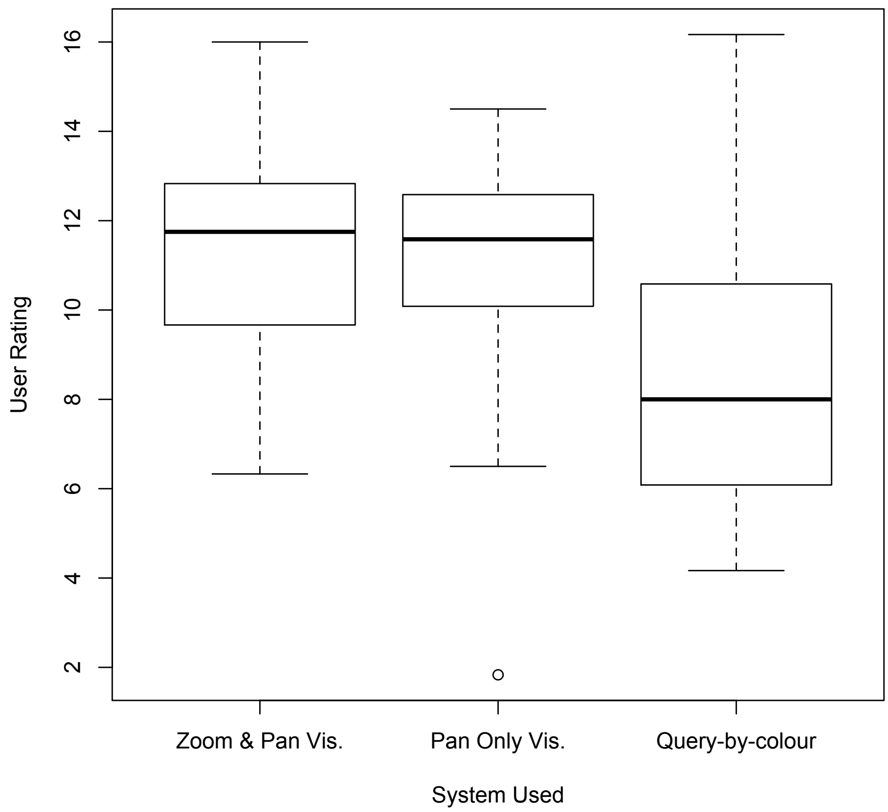


Figure 4.7: Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants of our first study when using the three colour-based image retrieval systems to create their image mosaics.

4.7.3.1 Mental Demand

As can be seen in Figure 4.8, the mental demand ratings assigned by participants to the pannable-only mapping-based visualisation system contained one outlier (a rating of 0.5). Whilst a Kruskal-Wallis test comparing the systems' mental demand ratings *with* the outlier identified as a circle in Figure 4.8 revealed no significant difference ($\chi^2_{(2)} = 5.83, p = 0.054$), we did find a significant difference across the systems when performing the same the test with the outlier excluded ($\chi^2_{(2)} = 6.69, p = 0.035$). A post-hoc test Wilcoxon Signed-rank test - used to assess whether the mean ranks of populations differ - showed there to be significant differences between the query-by-colour and zoomable mapping-based visualisation systems ($p = 0.028$) *and* between the query-by-colour and pannable-only mapping-based visualisation systems ($p = 0.026$).

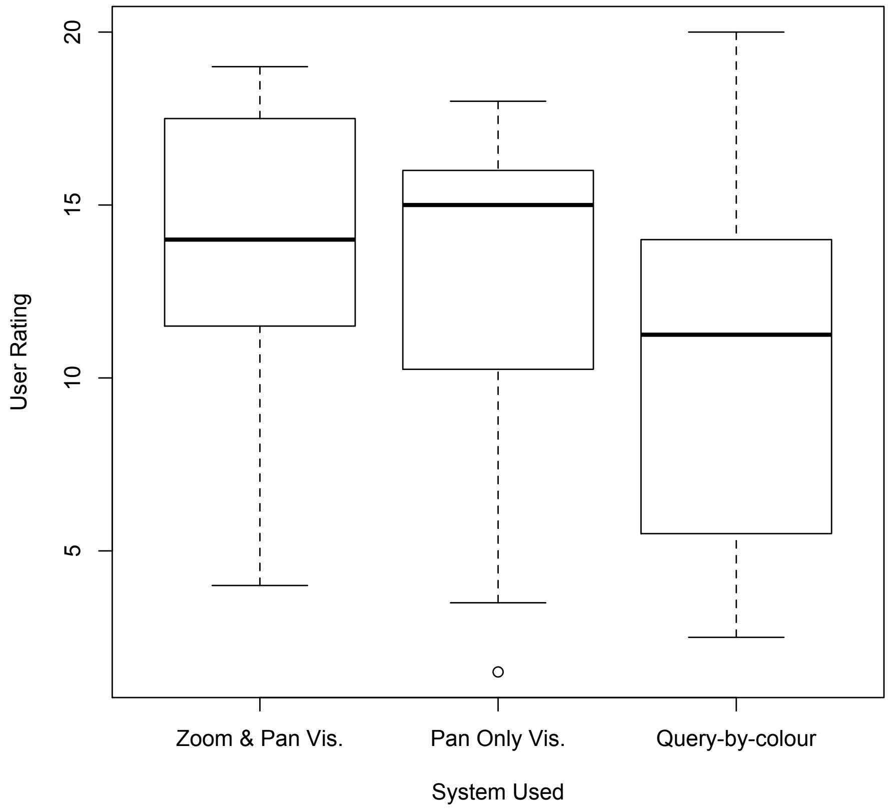


Figure 4.8: Box-plots of the mental demand ratings assigned by participants of our first study using the three colour-based image retrieval systems to create their image mosaics.

4.7.3.2 Effort

With the outlier identified in Figure 4.9 as a circle removed, we performed a one-way ANOVA test to compare the effort ratings of participants across our three systems. This revealed a significant difference between the times recorded for the three systems ($F(2, 68) = 3.25, p = 0.045$), with a post-hoc Tukey's HSD test unveiling participants of our first study felt that they had expended significantly more effort creating image mosaics using the pannable-only mapping-based visualisation system than when using the query-by-colour system ($p = 0.354$).

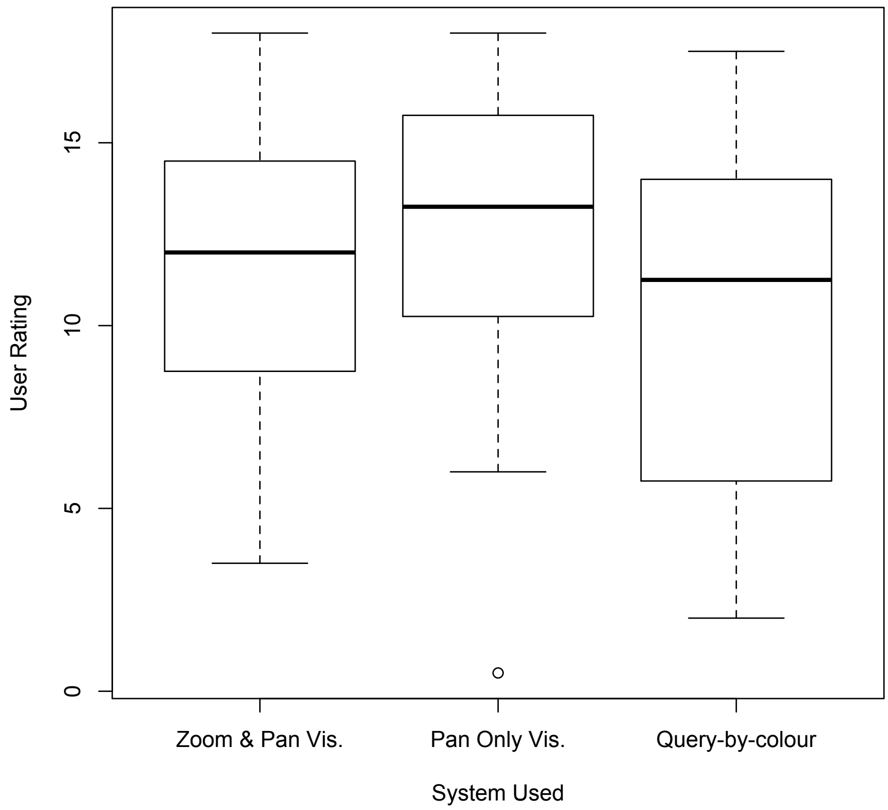


Figure 4.9: Box-plots of the effort ratings assigned by participants of our first user study using the three colour-based image retrieval systems to create their image mosaics.

4.7.3.3 Frustration

As previously discussed in Section 4.7.1, a major source of participant frustration when using the systems featuring a mapping-based visualisation was that the organisation of images appeared random, making it difficult for users to recall which areas of the visualisation they had already reviewed. It was therefore unsurprising to find that the frustration ratings of users significantly differed across the three systems both with ($F_{(2,69)} = 6.06, p < 0.01$) and without ($F_{(2,68)} = 7.38, p < 0.01$) the inclusion of the outlier in the data, shown in Figure 4.10 as a circle. Post-hoc Tukey's HSD tests showed that there was a significant difference between the frustration ratings for the query-by-colour and the zoomable mapping-based visualisation systems both with and without the outlier in the data included, as well as a significant difference between the frustration ratings supplied for the query-by-colour and pannable-only mapping-based visualisation systems (once again, both with and without the outlier included).

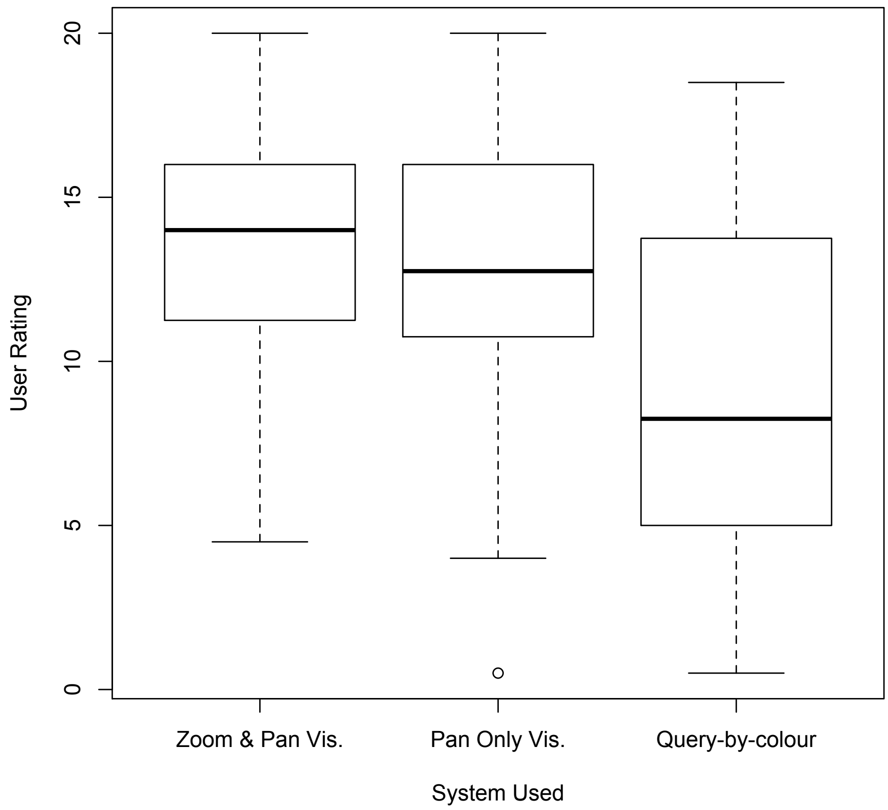


Figure 4.10: Box-plots of the frustration experienced by participants of our first study when using the three colour-based image retrieval systems to create their image mosaics.

4.7.4 User Preference Ratings

The workload results in the previous Section would suggest that the query-by-colour system was perceived by users as being easier and, in part, more effective than the the mapping-based visualisation systems for colour-based image retrieval. This finding is also further substantiated when we compare the user preference ratings assigned to the three systems by study participants. Recall that, upon completing their three image mosaics, participants were asked to rate their preferences for the systems on a single 20-point scale.

As can be seen in Figure 4.11, the highest rating was awarded to the query-by-colour system (20), whilst the system awarded the lowest rating was the pannable-only mapping-based visualisation system (0). A one-way ANOVA test on the data with the outliers removed revealed that the user's preference ratings were significantly different ($F_{(2,67)} = 33.53, p < 0.01$). Post-hoc Tukey's HSD tests showed that there was not only a significant difference between the preference ratings for the query-by-colour system and each of the mapping-based visualisation systems, but also a significant difference between the preference ratings assigned to the zoomable and pannable-only mapping-based visualisation systems.

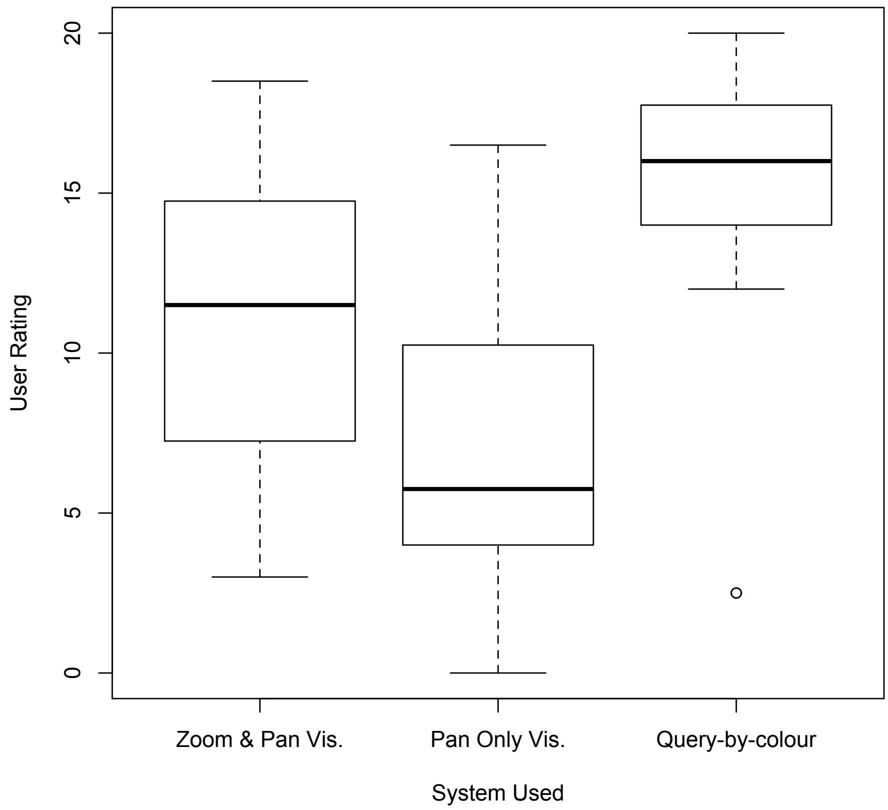


Figure 4.11: Box-plots of the preference ratings assigned to each of the three colour-based image retrieval systems by participants for creating their image mosaics.

4.8 Discussion

In this Section we discuss the results presented above. In addition, we discuss observations of test sessions and the aspects of our post-test discussions with study participants which we believe contributed towards the results of our first user study.

4.8.1 Time

Contrary to our initial hypothesis, users created image mosaics faster using the pannable-only mapping-based visualisation as opposed to the zoomable system. This could be due to the fact that fewer interactions (i.e., mouse operations) were required in order to navigate to an area of interest within the visualisation in the pannable-only system.

Many study participants stated that, because they had more confidence in retrieving the most suitable images in the database using the query-by-colour system (as opposed to when using the mapping-based visualisation systems), their image standards were higher (i.e., they wanted to find more accurate images for use in their mosaics) and, as such, were willing to interact with the query-by-colour system for longer than the other systems to find these more accurate images. This explains participants creating image mosaics faster using the pannable-only mapping-based visualisation system than when using the query-by-colour system.

4.8.2 Image Mosaic Accuracy

Post-test discussions revealed that several participants believed they had `got lucky' when finding some of the more visually accurate images using the two systems featuring a mapping-based visualisation; whilst attempting to retrieve an image suitable for a mosaic cell, participants had accidentally located another image in the visualisation suitable for use in another mosaic cell. Hilliges [36] claims that such serendipitous discovery of images is an advantage of similarity-based visualisation systems over query-based approaches.

4.8.3 Workload

As the results of the study show, participants experienced significantly higher levels of workload when using the two mapping-based visualisation systems as opposed to he query-by-colour system. On examining the individual workload dimensions, the results reveal that participants of the study marked significantly higher ratings on the frustration, effort and mental demand scales for the mapping-based visualisation systems as opposed to the query-by-colour system.

Post test discussions with study participants revealed an increased level of frustration was caused by what appeared to many as a `seemingly random' organisation of images. A majority of users reported unintentionally revisiting and reviewing the same areas of the visualisation multiple times because, as they were unsure where in the visualisation the images with colours they required resided, they resorted to randomly exploring the database. Performing this retrieval tactic in the pannable-only mapping-based visualisation required a great deal of interaction from users and this is reflected in the participants' effort ratings for the said system. Many of the participants also outlined a source of frustration applicable to all three of the systems tested; whilst they could find several images in the database with colour content suitable for occupying a particular image mosaic cell, they could not find images with the required colours in the correct positions. During post-test discussions, several study participants alluded to a query-by-sketch type interface as a potential solution to the issue. These users suggested that the query-by-colour system could be significantly improved if a facility for specifying the desired location of query colours was added (e.g., a facility to specify that the colour red was required in the bottom left-hand corner of the desired image).

As well as being left frustrated whilst using the mapping-based visualisation systems, many participants felt that the visualisation required them to process an excessive amount of visual information, leaving them `overwhelmed'. One participant likened the visualisation to `a sea of images'. These users preferred instead the query-by-colour system, as they could `filter' the visual information by selecting requisite colours. This will have undoubtedly contributed negatively towards the mental demand ratings of the mapping-based visualisation systems. We should note for balance, however, that there were several participants who, despite the extent of visual information, particularly enjoyed the high degree of interaction afforded by the mapping-based visualisation systems.

4.8.4 Preferred Approach for Colour-based Image Retrieval

For the majority of study participants, the query-by-colour system was the preferred of the three systems tested. Many participants felt confident that they had retrieved the most suitable images in the database for their mosaics when using the query-by-colour system, but were not so convinced that they had selected the most suitable database images when using either of the mapping-based visualisation systems. Whilst this participant sentiment is not reflected in the mosaic accuracy measures recorded across the three systems, it does appear to have effected user's preference ratings. The high level of workload experienced by participants when creating their mosaics using the mapping-based visualisation systems also appears to have impacted the preference ratings of users.

4.8.5 Zooming Preference

As stated previously, participants assigned significantly higher preference ratings to the zoomable mapping-based visualisation system than its pannable-only counterpart. Many participants revealed during post-test discussions that they had felt restricted with the pannable-only mapping-based visualisation system, as they did not have the ability to adjust the number of images displayed on screen at one time. When using the zoomable mapping-based visualisation system, most participants opted to display hundreds of images on screen at a time (as per Figure 4.3b). Several participants of the study remarked that, because their image requirements were solely based on colour rather than semantic content, they could assess the suitability of images at such a low resolution (i.e., more zoomed out). Another user behavioural observation made with respect to zooming during the study was that most participants opted to use the mouse scroll wheel for zooming in and out of areas of interest within the visualisation, as opposed to the available zooming buttons and slider.

We should note, however, that not all the participants of the study preferred the zoomable mapping-based visualisation system to its pannable-only counterpart. A minority of users stated that they preferred the pannable-only mapping-based visualisation system because, due to the fact that fewer interactions (i.e., mouse operations) were required in order to navigate to an area of interest within the visualisation, they perceived the system as being easier to use.

4.9 Conclusion

Prior to our first user study, no research existed evaluating the effectiveness of any similarity-based visualisation for colour-based image retrieval. So that we could compare a mapping-based visualisation with the typically adopted query-by-colour style, we recruited 24 users to evaluate three colour-based image retrieval systems using the Mosaic Test. Participants were asked to create image mosaics using both a query-by-colour system and a system featuring a zoomable mapping-based visualisation, fulfilling part of our research aim *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*. The third colour-based image retrieval system tested in our first user study featured the same mapping-based visualisation as the second system. For this third system, however, we removed the zoom facility so that users could only pan the visualisation to search for images. Removing this functionality from the third system enabled us to meet another of our research aims; *to assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases*.

We hypothesised that the system featuring a zoomable mapping-based visualisation would be more effective than its pannable-only counterpart for colour-based image retrieval, on account of users being able to navigate to visualisation areas and thus images of interest more quickly (as we expected zooming to require fewer interactions than panning). It was a surprise to find, therefore, that participants created their image mosaics significantly faster using the pannable-only mapping-based visualisation system than when using the zoomable system. We believe this could be due to the fact that fewer interactions (i.e., mouse operations) were required in order to navigate to an area of interest within the visualisation in the pannable-only system. We found there to be no significant difference between the other Mosaic Test measures recorded across the two systems. We did, however, also observe a significant difference in the users' preference ratings, with participants assigning higher ratings to the zoomable system. Many of the study participants suggested that they had felt limited whilst using the pannable-only system, preferring instead the ability to customise the number of visualisation images displayed on screen at one time via zooming. On the basis of this finding, we conclude that systems facilitating colour-based image retrieval via a mapping-based visualisation should include a zoom facility to improve the overall user experience.

As a system featuring a mapping-based visualisation had not been compared to a query-by-colour system prior to this research, it was difficult for us to hypothesise which of the two would be most effective for colour-based image retrieval. Whilst we found there to be no significant difference between the Mosaic Test time and accuracy measures recorded across the two systems, we did observe significant differences in the users' preference and subjective workload ratings. Participants of the study assigned higher preference ratings to the query-by-colour system than for the systems featuring a mapping-based visualisation, as well as registering higher overall workload ratings for the latter. Analysis of the participant's subjective workload ratings revealed that users experienced significantly more effort, mental demand and frustration when creating image mosaics using the mapping-based visualisation systems than when using the query-by-colour system. Post-test discussions revealed that the frustration experienced and extra effort required by many participants when creating image mosaics using the mapping-based visualisation systems - due to the seemingly random organisation of images causing users to accidentally revisit and review areas of the visualisation multiple times - as well as the mental demand required to process the large amount of visual information, impacted negatively on their preference ratings and workload ratings.

The findings of our first user study suggest that the query-by-colour query-style is more effective than a similarity-based visualisation for colour-based image retrieval. Of course, we have tested just one feature vector (the colour histogram), one dimensionality reduction technique (Principal Component Analysis), and one visualisation style (the mapping-based visualisation style). It may be that using a different combination of feature vectors, spreading strategies, dimension reduction techniques (for mapping-based visualisations), and even visualisation style, could produce a more effective similarity-based visualisation for colour-based image retrieval which may have resulted in a different outcome for our research. We have, at the very least, contributed a starting point for others in the field to undertake further comparisons between similarity-based visualisations and the query-by-colour query style for colour-based image retrieval. Given that this study is the only recorded investigation of its kind, we must conclude that mapping-based visualisations are less effective than the query-by-colour query style for colour-based image retrieval. We therefore recommend that mapping-based visualisations should not be adopted instead of the query-by-colour query style in colour-based image retrieval systems.

In the following Chapter, we describe how we satisfied the final aim of our research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*.

Chapter 5

Comparing Colour-based Image Retrieval Query Result Visualisations

5.1 Introduction

In this Chapter, we describe our second user study which was undertaken to fulfil the final aim of this research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*. Recall from Section 2.5.4 that Liu *et al.* [55] conducted a user study comparing three image retrieval systems; presenting keyword query results either in the form of a ranked list (where images ranked by the system as most similar to the query appear at the head of the list), a mapping-based visualisation, or a clustering-based visualisation. Test users were given a set of query terms and asked to search for images in the test database which they believed best matched each of these pre-determined query terms. It was observed that users could find images fastest using the systems presenting query results in the form of mapping and clustering-based visualisations (as opposed to the ranked list presentation). Whilst the search times recorded for the systems adopting mapping-based and cluster-based visualisations were not found to be significantly different, users did assign higher preference ratings to the mapping-based visualisation system - claiming it to be more intuitive, interesting, and convenient than the ranked list and clustering-based visualisation systems for searching and comparing images.

In Liu *et al.*'s study, participants were asked to perform high-level image retrieval (i.e., search for images based on the objects or subject matter depicted within them). We believed the outcome of Liu *et al.*'s study could have been different if participants had instead been searching by colour (i.e., performing much lower-level image retrieval); especially given that participants of our first user study preferred a system presenting query results as a ranked list over a mapping-based visualisation for colour-based image retrieval. Since Google Images [44] and Bing Images [17] currently adopt a ranked list for presenting both keyword and colour query results, we wanted to evaluate what effect changing the visualisation of results would have on the large number of users searching by colour.

Another motivation for conducting our second user study was to fulfil one more aspect of the research aim; *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*. As reported in Section 2.8, we found no evaluation had been conducted previously comparing the effectiveness of the query-by-colour and query-by-sketch query styles for colour-based image retrieval. We found this particularly intriguing given that many colour-based image retrieval systems typically support either the query-by-colour *or* query-by-sketch query styles [47, 42, 22, 44], but rarely both [92].

Let us consider the predicament a software designer would face if asked to create a new colour-based image retrieval system with this gap in existing knowledge - should the new system implement the query-by-colour query style, the query-by-sketch query style, or both? If the designer were to opt for either one of the two query styles, it would be based on nothing more than intuition - leaving him prone to selecting the least effective of the two approaches. This would lead to less users being attracted to purchasing or using the system (as they would instead use an alternative system implementing the more effective query style), resulting in lost revenue for the company developing the system. On the other hand, the software designer could decide to `hedge their bets' and include both the query-by-colour and query-by-sketch query styles to ensure the that most effective approach was definitely implemented in the new system. Taking this design route would inevitably require more time and manpower to create the system, thus increasing development costs. The development company would deem this to be a waste of their resources, however, if end-users were to only value and hence use the more effective query style in the newly developed system.

To better support designers of colour-based image retrieval systems making the critical decision outlined above in the future, we wanted to compare the effectiveness of the query-by-sketch and query-by-colour query styles for colour-based image retrieval as part of our second user study.

5.1.1 Chapter Outline

This Chapter is organised as follows. In Section 5.2, we describe our second user study. In Section 5.2.1, we outline the hypotheses, based on the findings from both our first user study and Liu *et al.*'s. Next, in Sections 5.2.2 and 5.2.3, we respectively describe the participants recruited to, and the image database used for, the study. We also detail how the adopted image database was indexed in Section 5.2.3 for use with the six tested colour-based image retrieval systems. These systems are described in Section 5.2.4. We present the Mosaic Test procedure undertaken for evaluating the tested colour-based image retrieval systems in Section 5.2.5, before presenting and discussing the results of the study in Sections 5.3 and 5.4 respectively. Finally, Section 5.5 concludes this Chapter.

5.2 User Study

Participants recruited to our second user study were divided into two equally sized groups, the first of which created image mosaics using three systems adopting the query-by-colour query style, whilst the second group generated their image mosaics using three systems with the query-by-sketch query style. We grouped users in this way so that we could fulfil another part of our research aim; *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*, by directly comparing the query-by-colour and query-by-sketch query styles. Within each user group, the colour-based image retrieval systems visualised query results as either a ranked list (i.e., in descending order of closeness to the submitted query), a mapping-based visualisation, or as a clustering-based visualisation. This allowed us to satisfy the final aim of our research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*.

5.2.1 Hypotheses

We formulated two hypotheses for our second user study. The first concerned the comparison of the query-by-colour and query-by-sketch query styles for colour-based image retrieval. We hypothesised that the query-by-sketch systems tested in the study would prove to be more effective for colour-based image retrieval than those adopting the query-by-colour query style, on account of the number of users in our first user study bemoaning their inability to specify the location of colours in queries. In regards to our query result visualisation comparison, we hypothesised that - contrary to the findings of Liu *el al.* which were based on users performing higher-level image retrieval - the systems presenting query results in the form of a ranked list would be the most effective for colour-based image retrieval. This was based on the finding from our first user study that participants preferred (and experienced less workload when) using the query-by-colour system (which presented query results as a ranked list), as opposed to the mapping-based visualisation systems, for colour-based image retrieval.

5.2.2 Participants

We recruited 36 users to participate in our study and divided them into two equally sized groups. The first group of users tested the three query-by-colour systems, whilst the second group tested the three query-by-sketch systems. The query-by-colour user group were made up of 9 males and 9 females. The ages of the males in this group ranged between 18 and 35, whilst the females were aged between 18 and 20 years old. In the query-by-sketch user group, there were 6 females and 12 males, all aged between 18 and 35. As we have shown previously in Chapter 4 that it was not necessary to recruit expert users (e.g., graphic designers) to evaluate colour-based image retrieval systems with the Mosaic Test, we chose to recruit participants via an email and poster campaign at Aston University. The only prerequisites for participation were that test users had a formal qualification in Information

Technology (or a similar discipline) and had no known colour-blindness or colour-vision deficiency. As per our first user study, every participant was required to sign a consent form confirming they had no known issues with colour-blindness or colour-vision deficiency. This consent form is included in Appendix F. Furthermore, test-users recruited to our second user study were also required to complete a short colour blindness test prior to participation. Contrary to our first study, test-users recruited to our second user study were offered £10 as an incentive for participation.

5.2.3 Image Database

We chose to increase the size of the image database used in our second user study to make it more comparable with the magnitude of commercial collections available these days [68]. Specifically, we adopted the one million 64×64 pixel thumbnails of the MIR-FLICKR 1M image collection [40]. We extracted three feature vectors from each database image. As per our first user study, we extracted 120 bin colour histograms from database images by mapping pixels to their closest corresponding colour in the adopted colour space - namely the 120 colours contained within the graphical colour palette of the MultiColr system [42]. The second feature vector extracted from database images was the MPEG-7 colour layout descriptor [87], described previously in Section 2.1. The final feature vector extracted from database images was the average pixel colour in an RGB colour space. In the following Section, we describe how the tested colour-based image retrieval systems made use of the extracted feature vectors.

5.2.4 Tested Colour-based Image Retrieval Systems

Each of the six colour-based image retrieval systems evaluated in our second user study adopted the same fundamental interface design. As can be seen in Figure 5.1, at the top of the system interface (highlighted in green) is a *query panel* for users to formulate a colour query (either a query-by-colour or a query-by-sketch, depending on the user group). Below this is the *query result visualisation area* (highlighted in blue) where users browse query results. In the following Section, we describe the functionality of the query-by-colour panel, applied exclusively to the systems tested by our first user group, and the alternative query-by-sketch panel, implemented only in the systems tested by our second user group.

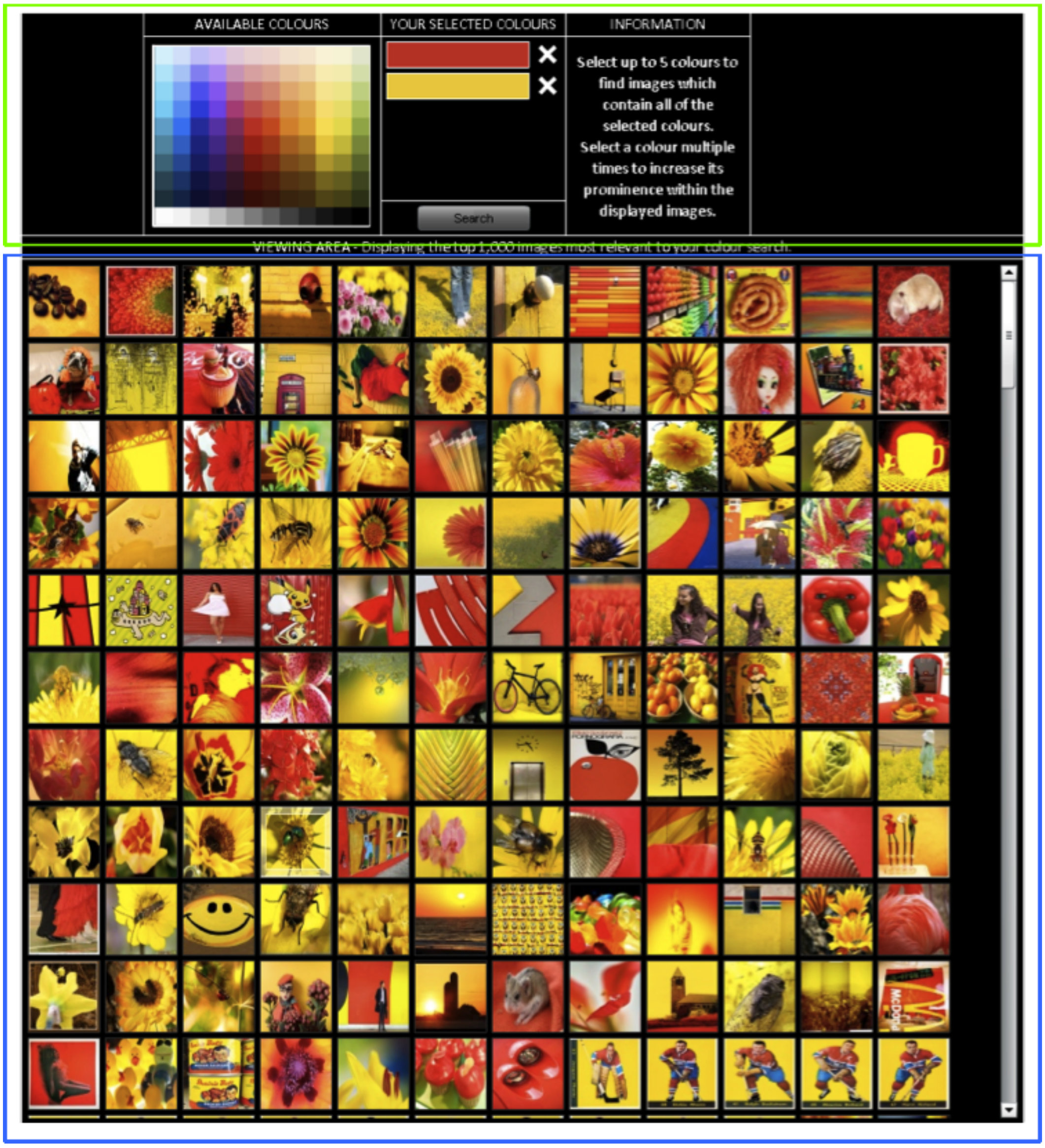


Figure 5.1: Fundamental interface design of the colour-based image retrieval systems tested in the second user study, comprising a *query panel* (highlighted in green) and a *query result visualisation area* (highlighted in blue).

5.2.4.1 Query-by-colour Panel

To formulate a query-by-colour, users could select multiple colours from the graphical colour palette in the query panel. As can be seen in Figure 5.2, the colours comprising a user's current query were displayed to the right of the graphical colour palette (under the label ``*Your Selected Colours*''). Colours could be removed from the query by clicking the large white cross located adjacent to the required colour swatch. As per the MultiColr system [42], users could select a maximum of five colours from the graphical colour palette to query the database and could select the same colour multiple times to increase the relative prominence of that colour. Based on feedback from participants of our first study, users of the query-by-colour systems in the second study could click a colour swatch to add the same colour again to their selected colours (as opposed to having to re-find the colour in the graphical palette).

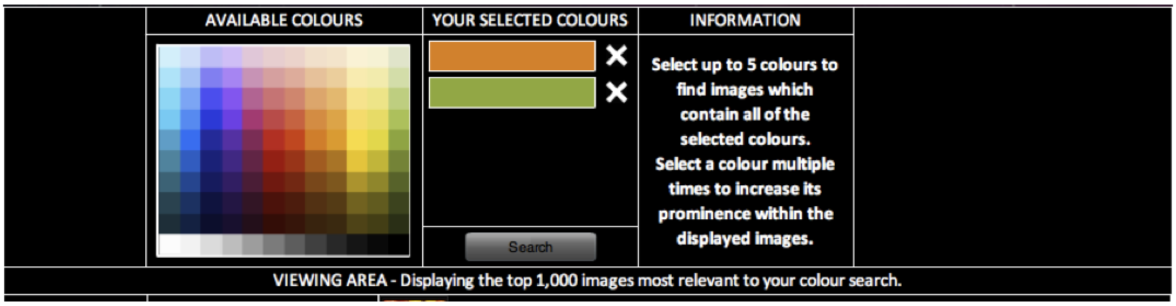


Figure 5.2: Screenshot of the query panel of query-by-colour systems tested in the second user study.

For the query-by-colour system used in our first study, a search was initiated each time the user's query was modified (i.e., when a colour was added or removed from a query). Since the query-by-colour systems in this study indexed a larger database, and the computation of query result visualisations was much more computationally expensive (in the case of the systems presenting query results as mapping and clustering-based visualisations), query results were only updated once users had chosen the colours for their query and clicked the ``*Search*'' button (located below the swatches of their selected colours).

As per the previous query-by-colour system, the query-by-colour systems tested in this study generated a 120 bin query colour histogram from the user's selected colours. The weight of each query histogram bin was calculated using the formula shown in Equation 5.1, where H_c is the histogram bin for colour c, t is the total number of colours selected and n is the number of times colour c occurs within the selection (as users had the ability to select the same colour more than once). It is important to note that the order in which users selected colours was not considered by the system during the formulation of the query histogram (i.e., it was not a ranked calculation).

$$H_c = \frac{1}{t} \times n \tag{5.1}$$

In addition to a colour histogram, the query-by-colour systems in our second user study also computed an average RGB colour from the user's query. Inspired by the pre-filtering performed by QBIC [23], this average RGB colour was compared to the average RGB colour of database images using the L_2 distance metric to find the 10,000 most similar database images. The system's query histogram was then compared (using the L_2 distance metric) to the colour histograms extracted from the 10,000 database images returned through pre-filtering by average colour. The 1,000 database images with colour histograms most similar to the user's query were then presented in one of the three tested query result visualisations (dependent on the system being tested). We describe the three query result presentations in Section 5.2.4.3.

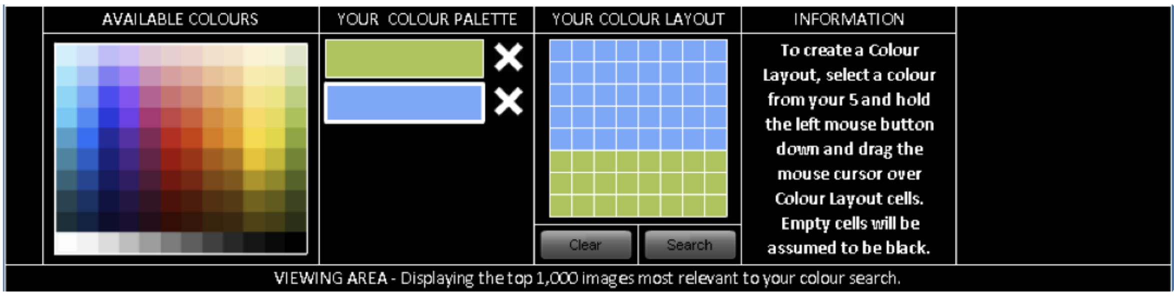


Figure 5.3: Screenshot of the query panel of query-by-sketch systems. The user has filled the bottom 24 cells of their query sketch in green and the upper 40 cells using blue.

5.2.4.2 Query-by-sketch Panel

To formulate a query-by-sketch, users could select multiple colours from the graphical colour palette (displayed under the label ``*Available Colours*''). As can be seen in Figure 5.3, the colours selected by users were displayed to the right of the graphical colour palette (under the label ``*Your Colour Palette*''). Users could select a maximum of five colours to use in their sketch. For the query-by-sketch panel, however, only one instance of a colour could be added to the user's colour palette. Users could select a colour from *their* colour palette by clicking the appropriate swatch. The currently selected colour was highlighted with a bold white line (e.g., the blue swatch in Figure 5.3). A user's query sketch (displayed under the label``*Your Colour Layout*'') comprised 64 cells. Users could fill cells in their query sketch with the currently selected colour by left clicking the sketch cell they wanted to fill. As shown in Figure 5.3, the user has filled the bottom 24 cells of their sketch with green and the upper 40 cells in blue. To support more fluid sketching, users could left click and drag the mouse cursor over the sketch cells they wished to colour (i.e., they did not need to click each individual cell). The colour of sketch cells could be overwritten with other colours, or removed via a right click (on the cell requiring clearing). As per colouring, users could remove the colour from multiple sketch cells in a fluid motion by right click and dragging the mouse cursor over the cells they wished to clear. Sketch cells with no colour were treated as black by the system. To remove all sketch cells of a given colour, users could click the large white cross located adjacent to the swatch (located in their colour palette) of the colour to be removed. This also removed the colour from their colour palette. Users could clear the colour from all sketch cells by clicking the ``*Clear*'' button, located directly below (and to the left of) their query sketch.

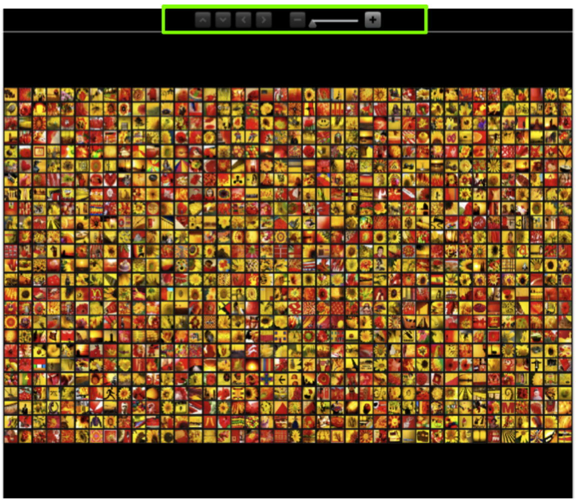
To initiate a query-by-sketch, users were required to click the ``*Search*'' button, located directly below (and to the right of) their query sketch. The query-by-sketch systems generated an MPEG-7 colour layout descriptor [87] from users' sketches. Furthermore, as was the case in the query-by-colour systems, an average RGB colour was also calculated from users' queries. This average RGB colour was compared to the average RGB colour of database images using the L_2 distance metric to find the 10,000 most similar database images. The system's query MPEG-7 colour layout descriptor was then compared (using the metric defined in [87]) to the colour layout descriptors extracted from the 10,000 database images returned through pre-filtering by average colour. The 1,000 database images with MPEG-7 colour layout descriptors most similar to the user's query were then presented in one of the three tested query result visualisations. These three query result presentations are described in the following Section.

Figure 5.4: Screenshot of query results in a ranked list visualisation. Users browse query results either by scrolling the scroll bar (highlighted in green) or via the mouse wheel.

5.2.4.3 Query Result Visualisation Area

As per the study of Liu *et al.*, all participants in our second user study tested three query result presentations - namely, a ranked list, a mapping-based visualisation, and a clustering-based visualisation. It is important to note that in all of the query result visualisations, users were required to click and hold the left mouse button whilst dragging to copy an image from the test system to cells in their image mosaics. The number of images displayed in the visualisation (i.e., the number of images returned from a users' query), was displayed at the top of the query result visualisation area in each of the tested systems. In the remainder of this Section, we describe how these presentations were produced in response to user queries. We also describe the manner in which test-users could browse query results across the three visualisations.

5.2.4.3.1 Ranked List As per query-based colour-based image retrieval systems available today, for our ranked list visualisation the database images ranked as most similar to a user query were displayed at the head of the list. Users could browse query results either by scrolling via the scroll bar (highlighted in green) or the mouse scroll wheel. This visualisation style is shown in Figure 5.4.



(a) Zoomed out

(b) Zoomed in

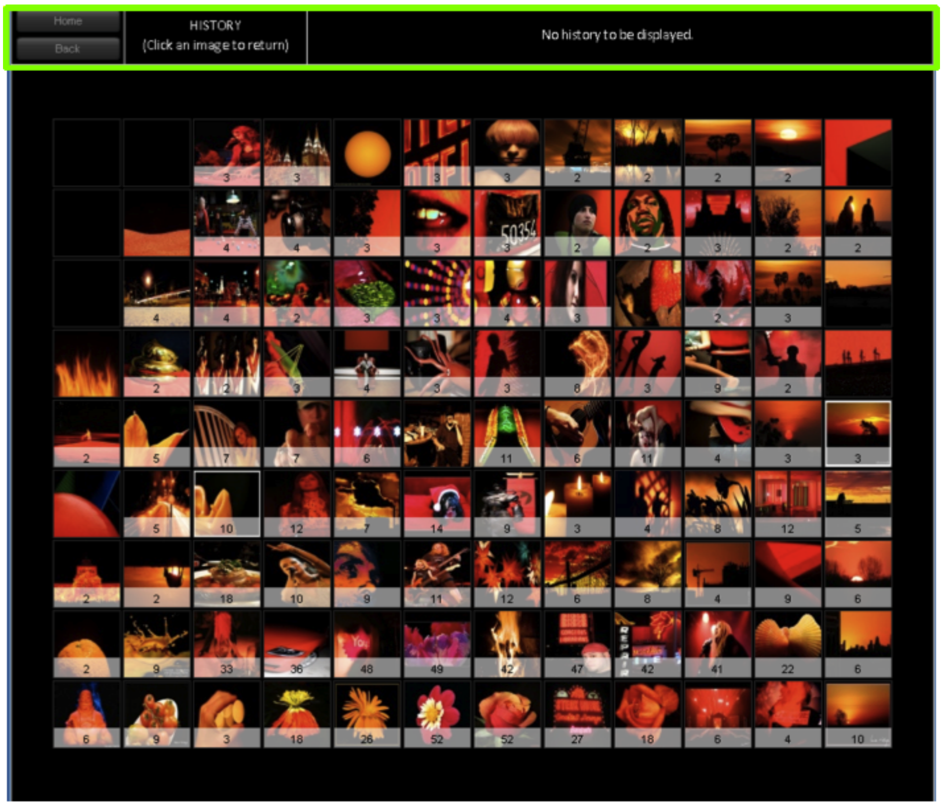
Figure 5.5: Screenshots of query results in a mapping-based visualisation. The toolbar with zoom and pan controls is highlighted in green.

5.2.4.3.2 Mapping-based Visualisation As per our previous study, we adopted the approach of Rodden *et al.* [73] to produce the mapping-based visualisation of query results for the two systems in our second study. First, the distance between the colour layout descriptors of all the images to be presented (i.e., the 1,000 images returned either through query-by-sketch or query-by-colour) was computed using the metric defined in [87] to formulate a distance matrix. This matrix was reduced down to just 2-dimensions via Principal Component Analysis (described previously in Section 2.5.1.1).

We opted to use this dimensionality-reduction technique because it is much less computationally expensive than iterative techniques such as multi-dimensional scaling. For convenience, we translated the derived co-ordinates so that all had positive x and y values. Since Rodden *et al.* [73] has previously demonstrated that users dislike image overlap that can occur within mapping-based visualisations, we fitted the databases images to a grid of 25×40 cells using a variation of Rodden *et al.'s bump* strategy (described previously in Section 2.5.1.2.1). We first mapped images to cells on the grid according to the 2-dimensional co-ordinates derived through Principal Component Analysis. Our spreading algorithm then located the grid cell with the highest number of assigned images and initiated a spiral search (emanating from this original cell) in order to locate the closest free cell. The images in the line of cells between the original cell and the next closest cell are all moved outwards (from the original cell at the centre) by one cell, with the image to be spread from the original cell moved to the newly vacated adjacent cell. Images are spread from grid cells until just one remains. The order in which images are spread from the grid cells is based on the distance between their derived co-ordinates and the grid cell centre (i.e., the image with derived co-ordinated closest to the grid cell centre is the one that remains). The result of the spreading algorithm is each grid cell contains exactly one image.

5.2.4.3.2.1 Controls As per the mapping-based visualisation systems in our first user study, we used OpenZoom [27] so that the query results could be interactively zoomed and panned by study participants. Users could zoom and pan using the controls located at the top of the query result visualisation area (highlighted in green in Figure 5.5) or via mouse operations; using the scroll wheel to zoom and the right mouse button (by clicking and dragging the visualisation) to pan.

5.2.4.3.3 Clustering-based Visualisation We adopted the approach of Schaefer and Ruszala [80, 82], described previously in Section 2.5.2.1, to produce the clustering-based visualisation of query results for the two systems in our study. The 2-dimensional co-ordinates were derived as per the mapping-based visualisation approach described in the previous Section (i.e., a distance matrix of colour layout distances between query result images was reduced to just 2-dimensions via Principal Component Analysis). At the root layer (e.g., Figure 5.6a), all database images are assigned to cells based their derived co-ordinates. The root layer of the visualisation contained 108 (9×12) cells, and each of these contained the same number of sub-cells (which, in turn, could harbour further sub-cells with the same dimensions). If a root layer cell contained only a single image, this was the image that was displayed in the cell. If more than one image was mapped to a cell, however, the image whose co-ordinates were closest to the cell centre was displayed as a representative image for that image cluster. The number of images contained within that cluster overlaid the representative image (with no number being displayed if it was the only image in the cell). Then, at the next level of the hierarchy, this cluster was expanded by subdividing the cell into a set of smaller cells and performing the above procedure again (that is, mapping each image in the cluster to a sub-cell, and, should multiple images be assigned to the same sub-cell, descend into the next level of the hierarchy).



(a) Root layer

(b) Group view

Figure 5.6: Screenshots of query results in a clustering-based visualisation. The toolbar with navigation controls and browsing history is highlighted in green.

Two approaches were then employed to reduce the number of empty cells and layers within a hierarchy. For the first, if an empty cell had three or more neighbouring cells that contained images, a fraction of these images were moved to the empty cell. The images moved from each of the occupied cells were those with derived co-ordinates closest to those of the centre of the empty cell (i.e., those images closest to the borders between the two cells). The overall effect of this approach is that more cells are filled, hence making better use of the visualisation space. The second approach combined the basic, swap and bump spreading strategies proposed by Rodden et al. [73] (discussed previously in Section 2.5.1.2.1) in order to prevent the creation of an undue number of layers. When a given parent cell consisted of fewer than 25% occupied cells, cells containing more than a single image would share their images with the nearest vacant cells. If an empty cell was located in the first ring of neighbours, then the `basic' strategy was applied. If the closest vacant cell was in the second ring of neighbours, however, then the `swap' strategy was applied. Finally, should the closest cell be located within the third ring of neighbours, the `bump` strategy was applied. If no free cell was located within the three rings, a new sub-cell would be generated.

5.2.4.3.3.1 Controls To navigate into a cell of images, users could double left-click the representative image for the cell they wanted to view. The images contained within the selected cell would then be displayed to users (as shown in Figure 5.6b). The representative image double clicked by the user would then added to the ``*History*'' pane of the toolbar. Users could click images in their history to return to previously visited cells of the hierarchy. Alternatively, users could click the ``*Home*'' or ``*Back*'' buttons to return to the root layer or previously viewed cell respectively. It is important to note that, if desired, users could left click and drag representative images to mosaics (and the overlaid cell image total would be removed).

5.2.5 Procedure

As we have already outlined in Section 5.2.2, participants of the study were divided into two equally sized groups; using either three query-by-sketch or three query-by-colour systems. Recall that within each user group, the colour-based image retrieval systems visualised query results as either a ranked list (i.e., in descending order of closeness to the submitted query), a mapping-based visualisation, or as a clustering-based visualisation. The variant feature of the systems tested *between* the two groups was the query style, whilst the variant feature of the systems tested *within* groups was the manner in which query results were visualised. The Mosaic Test procedure followed was almost identical to that reported back in Section 3.3.2; the only difference being that participants in our second study *were not* required to rank their three mosaics in order of perceived closeness to the corresponding target images post test. In the following Sections, we recap the adopted procedure.

5.2.5.1 Prior to the Mosaic Tests

Prior to completing their first Mosaic Test, participants were first given written instructions explaining the concept of an image mosaic and the functionality of the Mosaic Test Tool. Each participant undertook a practice session following these written instructions, in which the participant completed an image mosaic using a small selection of relevant images. Once participants were familiar with the functionality of the Mosaic Test Tool, and the evaluator had observed each participant completing a set of training tasks listed on a check sheet (such as dragging and removing images from the colour-based image retrieval system to the Mosaic Test Tool image mosaic), participants proceeded to the Mosaic Test for the first colour-based image retrieval system they were due to evaluate. Recall that the written instructions and `training tasks' check sheet used for practise sessions in the study can be found in Appendix A.

(a) Target image 1 (b) Target image 2 (c) Target image 3

Figure 5.7: The three target images to be recreated as image mosaics by participants of the study. Each target image is predominantly comprised of three jelly beans of varying colour and orientation.

5.2.5.2 The Mosaic Tests

Participants completed a Mosaic Test for each of the colour-based image retrieval systems under evaluation (i.e., they completed three image mosaics, one per a system). Before participants created an image mosaic using each colour-based image retrieval system, they were first trained, and given an opportunity to practise with, the functionality of the colour-based image retrieval system they were about to use. The training material for the query-by-colour and query-by-sketch systems of our second user study are included in Appendices D and E respectively. When participants had indicated to the evaluator that they were satisfied with the controls of the system, they proceeded to the assessed Mosaic Test. After completing an image mosaic with a colour-based image retrieval system, participants were asked to complete a NASA-TLX [31] subjective workload assessment for the system they had just used. Before completing their first NASA-TLX assessment, participants were provided with instructions on how the assessment should be completed. This background information, along with the NASA-TLX assessment used, is included in Appendix A.

We adopted the same three jelly bean image mosaics used for our first user study (shown again in Figure 5.7). To once again ensure that results were not affected by any one of the image mosaics being more difficult to complete than the others, the order in which the target images were presented remained constant whilst the colour-based image retrieval system order was counterbalanced across participants. Furthermore, the Mosaic Test Tool and colour-based image retrieval systems used were, for each participant, run on a Sony VAIO laptop, running Windows Vista, with a 17-inch (1600 x 900 resolution) display.

5.2.5.3 After the Mosaic Tests

So that we could accurately assess which of the systems within a group was preferred by users, after creating their three image mosaics participants were asked to rank their preference for each of the tested systems on a single 20-point scale.

5.3 Results

In this Section, we present the results from our second user study comparing the query-by-colour and query-by-sketch query styles (tested *between* the user groups) and the varying query result presentations (tested *within* groups). We compare the Mosaic Test measures recorded for the tested systems, namely the time required by users to complete their image mosaics, the accuracy of users' generated mosaics, and the subjective workload experienced by users whilst creating them. Furthermore, for our comparison of query result presentations, we also compare the preference ratings that users assigned to each of the tested systems upon completing their three image mosaics.

For the statistical comparisons reported in this Section, it is important to note that we adopt a 5% significance level. If our test data could be considered normally distributed, we adopted one-way ANOVA tests for our statistical analysis. When normality could not be assured, we used the Kruskal-Wallis test instead, which is a non-parametric alternative to ANOVA (i.e., unlike ANOVA, the Kruskal-Wallis test does not require the data to be normally distributed). Shaprio-Wilk tests (which test the null hypothesis that a sample $x_1, ..., x_n$ came from a normally distributed population) were used to verify the normality of tested data.

The box-plots included throughout this Section should be interpreted as follows: the lower box line represents the first quantile (i.e., the point at which 25% of the data occurs), the bold line within the box represents the median (i.e., 50% of the data), and finally the upper box line is the third quantile (i.e., 75% of the data). The region between the first and third quantiles (i.e., the height of the box) is commonly referred to as the *interquartile range*. The maximum box-whisker length (shown as dashed lines in our plots) is 1.5 times the interquartile range. We consider any data points beyond these whiskers to be outliers (represented as circles in our box-plots).

5.3.1 Query-by-colour and Query-by-sketch Query Styles

In this Section, we report the results of the study related specifically to our comparison of the query-by-colour and query-by-sketch query styles.

5.3.1.1 Time

As part of our hypothesis that the query-by-sketch systems would be more effective for colour-based image retreival than their query-by-colour counterparts, we anticipated that users of the query-by-sketch systems would be able to generate their image mosaics faster than query-by-colour users. We were very surprised, therefore, to observe the opposite of this result. As shown in Figure 5.8, the fastest image mosaic was created using a query-by-colour system (371 seconds) whilst the slowest was created using a query-by-sketch system; both with (2,999) and without (2,572) the outliers (identified in Figure 5.8 as circles) included. Through Kruskal-Wallis tests, we found that query-by-colour users created their image mosaics significantly *faster* than users of the query-by-sketch systems; both when the outliers in the time data were included ($\chi^2_{(1)} = 7.5944, p = 0.005855$) and excluded ($\chi^2_{(1)} = 6.2158, p = 0.01266$).

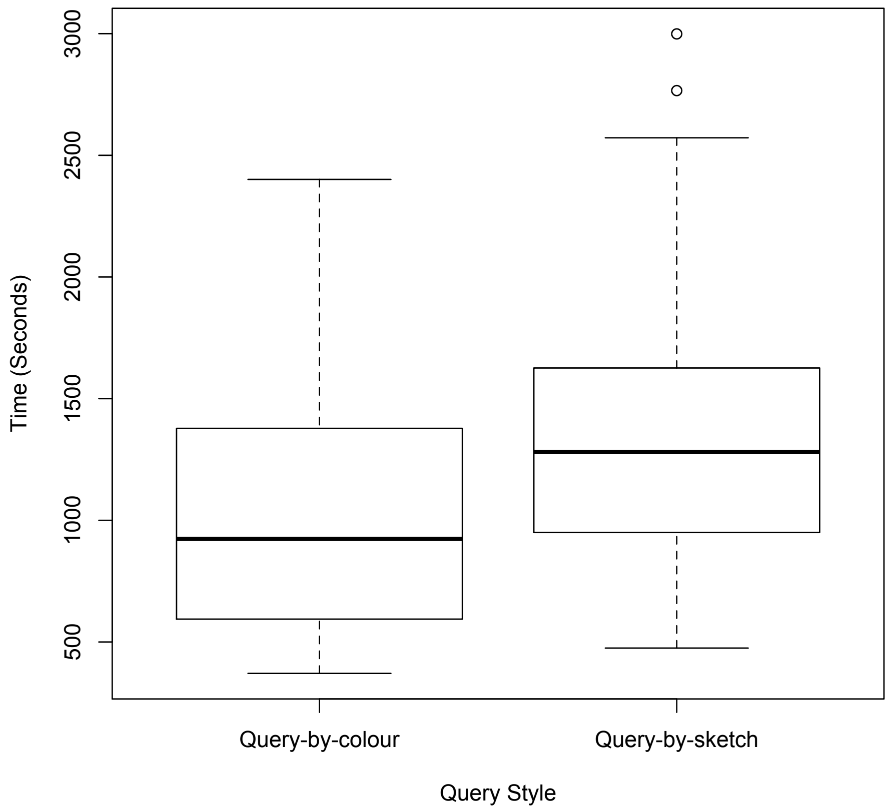


Figure 5.8: Box-plots of the times required by study participants, using either query-by-colour or query-by-sketch systems, to complete an image mosaic.

5.3.1.2 Image Mosaic Accuracy

The box-plots of the image mosaic accuracy data for the query-by-colour and query-by-sketch user groups are shown in Figure 5.9. The mean L_1 distance between the MPEG-7 colour structure descriptors of image mosaics and corresponding target images was lower in the query-by-sketch group (2,274) than the query-by-colour group (2,380). The most accurate image mosaic was generated by a participant in the query-by-sketch user group ($L_1 = 1,530$), whilst the most inaccurate mosaic was created by a user of a query-by-colour system ($L_1 = 3,095$). With the outlier identified in Figure 5.9 by a circle included, we found there to be no significant difference between the accuracy of image mosaics across the two query styles ($F_{(1,106)} = 2.85, p = 0.0943$). With the outlier excluded, we found the difference in image mosaic accuracy measures across the two query styles to be just short of significant ($F_{(1,105)} = 3.84, p = 0.0527$), based on the 5% level adopted for this research.

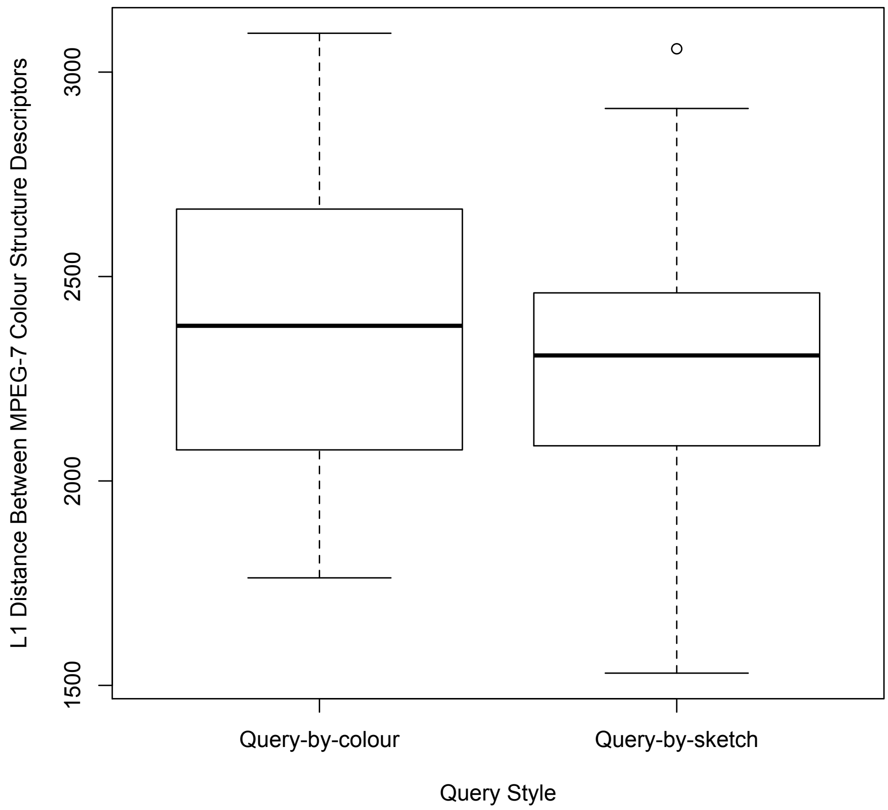


Figure 5.9: Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants using either query-by-colour or query-by-sketch systems, and their corresponding target images.

5.3.1.3 Workload

As can be seen in the box-plots in Figure 5.10, we observed no outliers in the overall workload ratings for the two user groups. As can also be seen in Figure 5.10, the mean overall workload rating was lower for users of the query-by-sketch systems (9.14) than for users of the query-by-colour systems (10.14). This supports our hypothesis that the query-by-sketch query style is more effective than query-by-colour for colour-based image retrieval. In further support of this, the lowest overall workload rating was assigned by a participant using a query-by-sketch system (3.66), whilst the highest rating was assigned by a participant using a query-by-colour system (16.5). A one-way ANOVA test on the data shows the differences in overall workload ratings between the two groups to be just short of our 5% significance level ($F_{(1,106)} = 3.6, p = 0.0605$). In the remainder of this Section, we report our findings for the physical demand, performance, and frustration, workload dimensions.

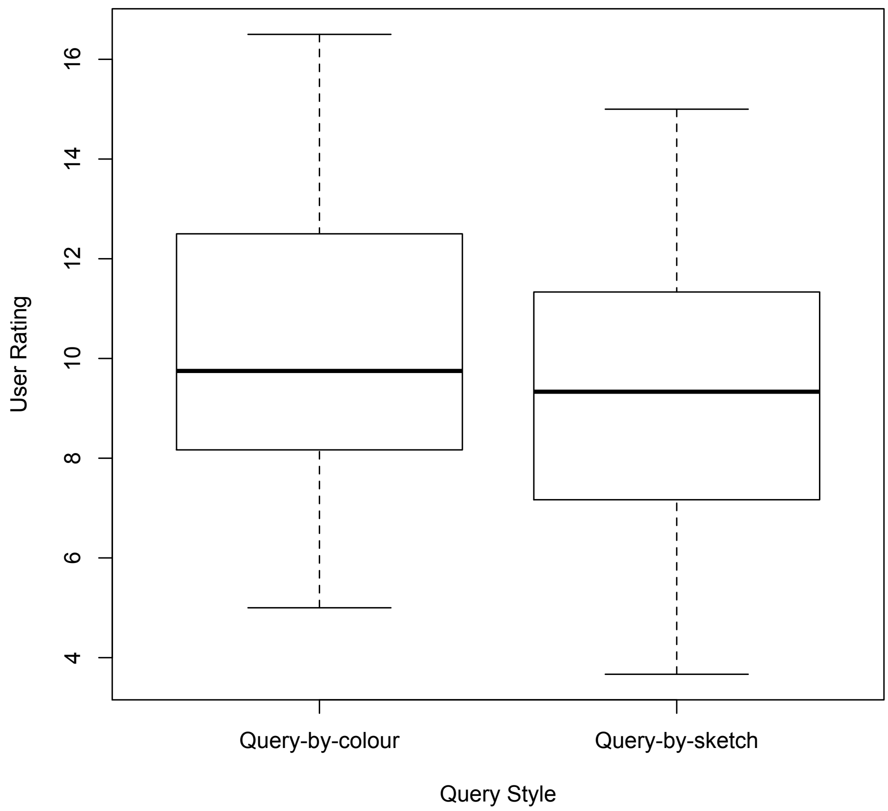


Figure 5.10: Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales), experienced by participants using either the query-by-colour or query-by-sketch systems, to create their image mosaics.

5.3.1.3.1 Physical Demand We were surprised to observe that the lowest physical demand rating in the study (0) was provided by three participants testing query-by-sketch systems. The highest physical demand rating was attributed to a query-by-colour system (19). Whilst a Kruskal-Wallis test showed the differences in physical demand to be just short of our 5% significance level when the outlier (i.e., the circle in Figure 5.11) was included in the data ($\chi^2_{(1)} = 3.512, p = 0.06093$), we *did* observe a significant difference when performing a Kruskal-Wallis test without the outlier in the query-by-sketch physical demand rating data ($\chi^2_{(1)} = 4.242, p = 0.03944$) - that is, we found the physical demand ratings of query-by-sketch users to be significantly lower than the physical demand ratings of query-by-colour users.

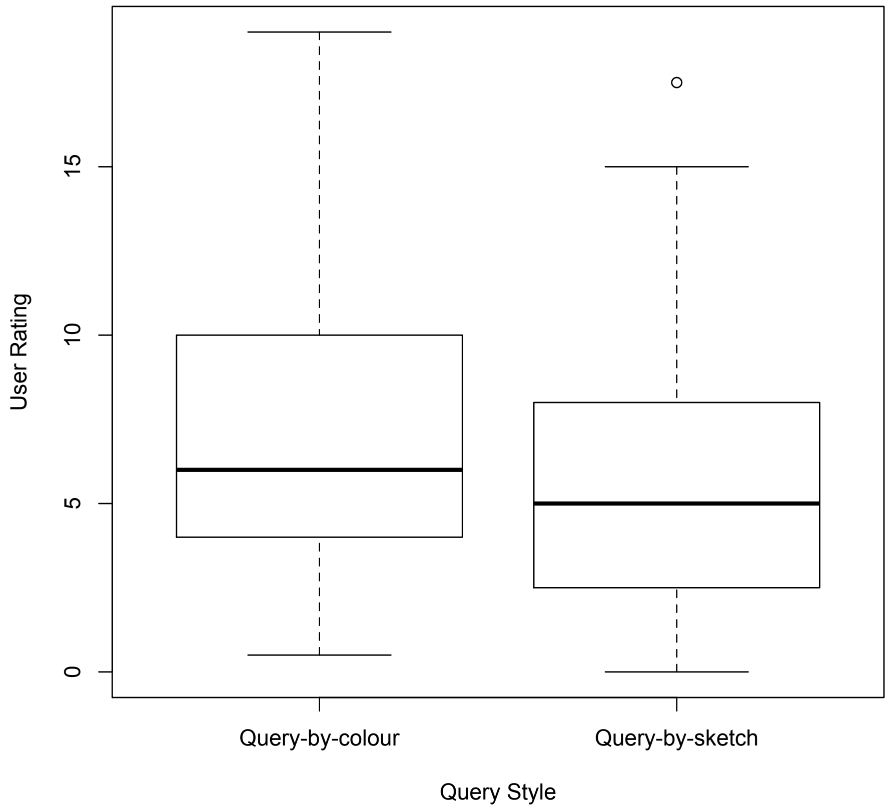


Figure 5.11: Box-plots of the physical demand ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics.

5.3.1.3.2 Performance As we reported previously in 5.3.1.2, although the differences in image mosaic accuracy were found to be just short of our 5% significance level, we observed that the mean L_1 distance between the MPEG-7 colour structure descriptors of image mosaics and corresponding target images was lower in the query-by-sketch group. This finding appears to be reflected in the perceived performance ratings of users; as well as a query-by-sketch user assigning themselves the highest possible self-perceived performance rating (20), with the two outliers (both a rating of 3) identified using circles in Figure 5.12 removed, a Kruskal-Wallis test revealed that the performance ratings assigned by query-by-sketch users were significantly lower than those attributed by query-by-colour users ($\chi^2_{(1)} = 3.8952, p = 0.04842$).

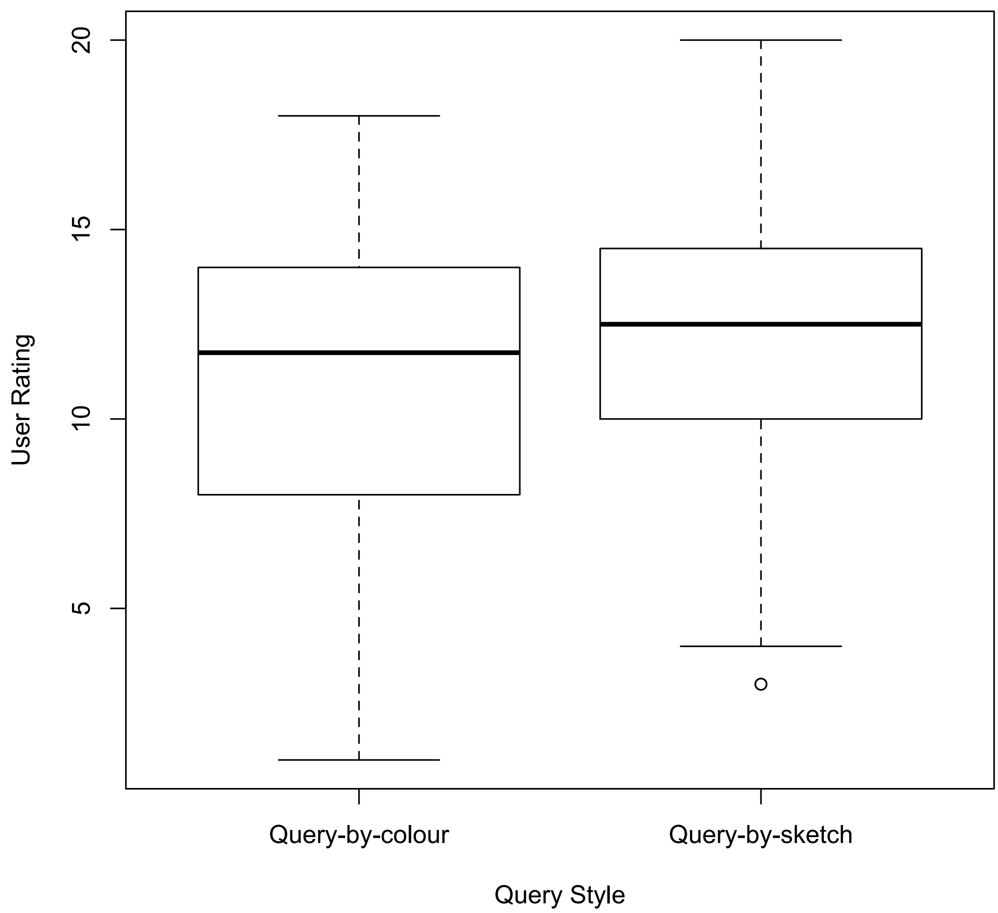


Figure 5.12: Box-plots of the performance ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics.

5.3.1.3.3 Frustration We formulated our hypothesis that the query-by-sketch query style would prove to be the most effective of the two query styles on account of participants of our first study bemoaning their inability to specify colour locations in queries. As such, we anticipated that the users of the query-by-sketch systems in our second user study would experience less frustration when creating their image mosaics than query-by-colour colours. As can be seen in Figure 5.13, the mean frustration rating of query-by-sketch users (7.6) *was* lower than the mean frustration rating of query-by-colour users (9.2). Furthermore, one participant in the query-by-sketch group even claimed to experience a frustration rating of zero. A one-way ANOVA test, however, found the differences in frustration ratings between the two groups to be just short of our 5% significance level ($F_{(1,106)} = 3.356, p = 0.0698$). It is important to note that we did not find significant differences between the ratings of the two user groups for any of the remaining workload dimensions.

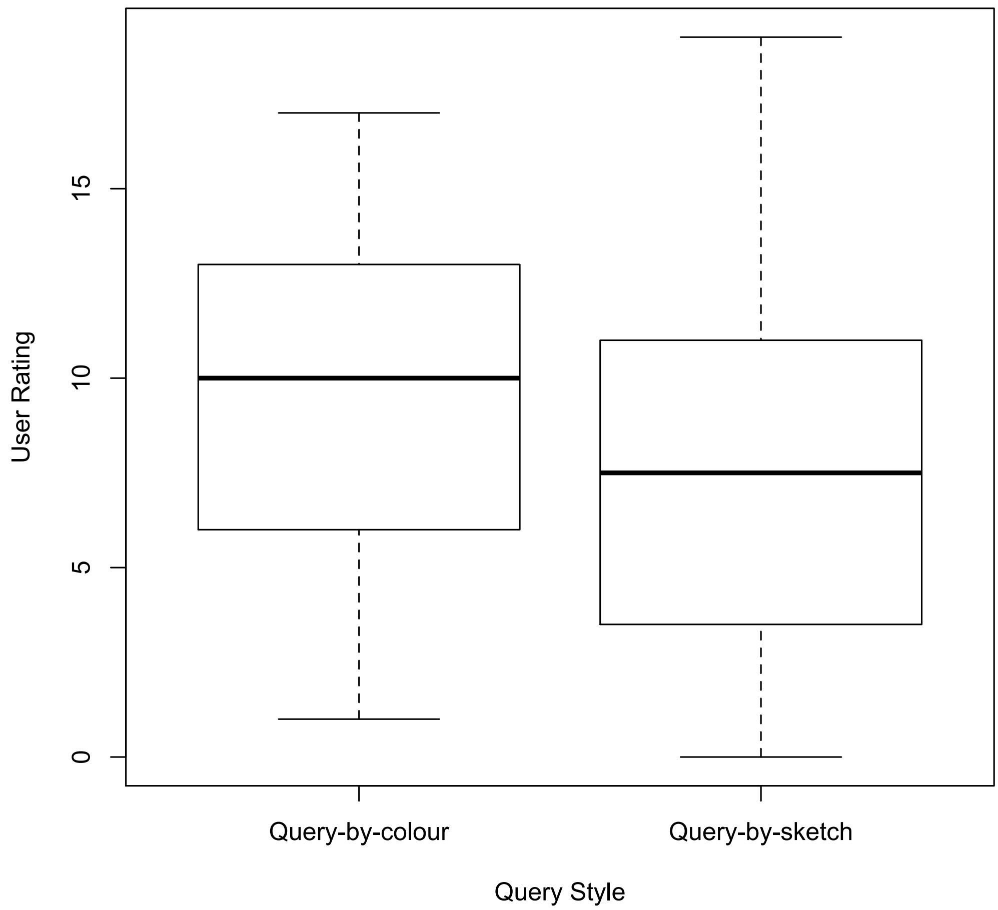


Figure 5.13: Box-plots of the frustration ratings, assigned by participants using either the query-by-colour or query-by-sketch systems, for creating their image mosaics.

5.3.2 Presentation of Query Results

In this Section, we report the results of our study related specifically to our comparison of the three tested query result visualisations; namely a ranked list, a mapping-based visualisation, or a clustering-based visualisation.

5.3.2.1 Time

As shown in Figure 5.14, the times taken by participants to generate their image mosaics ranged from a mere 371 seconds using a ranked list to 2,999 seconds (almost 50 minutes) using a clustering-based visualisation. As can also be seen in Figure 5.14, in complete contrast to the findings of Liu *et al.*, with the outliers (displayed as circles) removed, we observed that the slowest time was recorded by a participant using a mapping-based visualisation (2,572 seconds, or approximately 43 minutes). Through Kruskal-Wallis tests, we found there to be no difference between the times recorded across the query result presentations both with ($\chi^2_{(2)} = 1.1837, p = 0.5533$) and without ($\chi^2_{(2)} = 1.8341, p = 0.3997$) the identified outliers included.

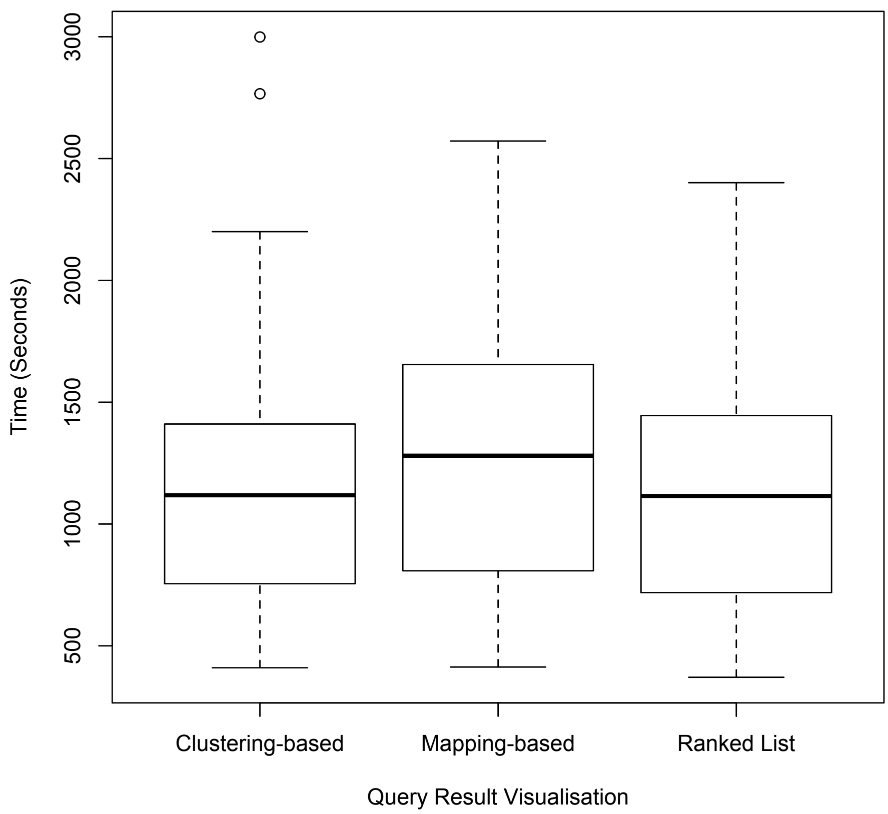


Figure 5.14: Box-plots of the times required by study participants to complete an image mosaics using each of the query result visualisations.

5.3.2.2 Image Mosaic Accuracy

As shown in Figure 5.15, the most accurate (or relevant) image mosaic was generated using a system presenting query results as a mapping-based visualisation ($L_1 = 1,530$). The least accurate image mosaic, on the other hand, was created by a participant using a clustering-based visualisation of query results ($L_1 = 3,095$). A one-way ANOVA test found there to be no significant difference between the accuracy of image mosaics across the three query result presentations ($F_{(2,105)} = 0.368, p = 0.693$).

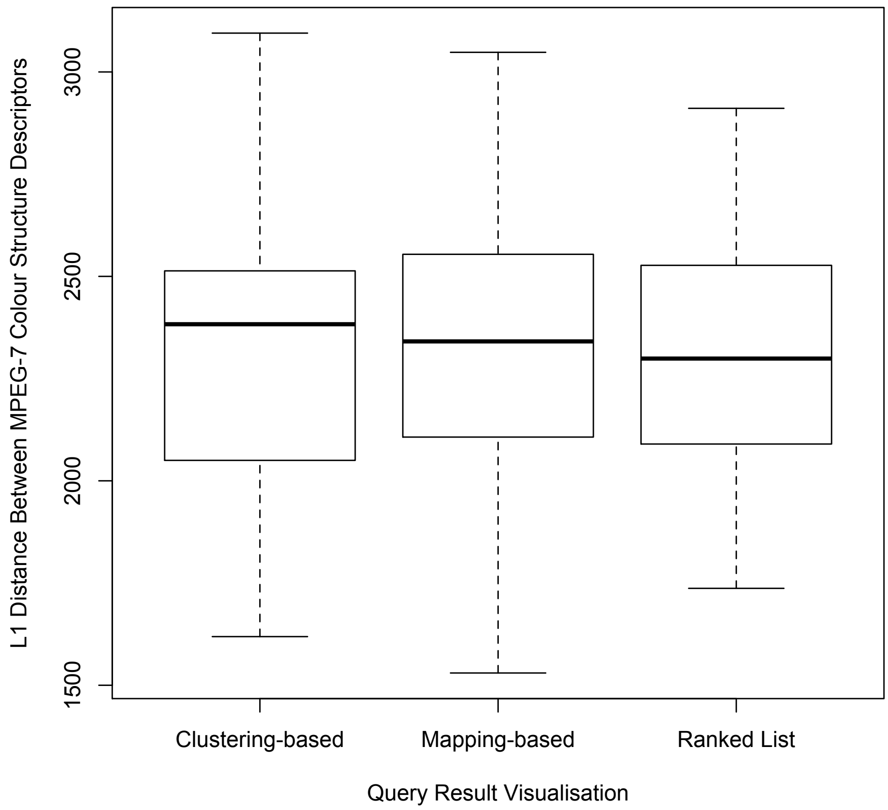


Figure 5.15: Box-plots of the L_1 distance between the MPEG-7 colour structure descriptors of image mosaics, created by participants using each of the query result visualisations, and their corresponding target images.

5.3.2.3 Workload

Box plots of the overall workload experienced by users across the query result presentations are shown in Figure 5.16. Contrary to what would be expected based on Liu *et al.*'s findings, the joint highest overall workload ratings of two participants were recorded after using a mapping-based visualisation ($\mu = 14.33$). The least overall workload experienced by a participant was recorded when using a system presenting query results as a ranked list ($\mu = 3.66$). It is important to note, however, that a one-way ANOVA test did not find the differences in overall workload across the query result presentations to be significant ($F_{(2,105)} = 0.465, p = 0.63$). We did not find any significant differences in the ratings of users across the query result visualisations for any of the individual workload dimensions.

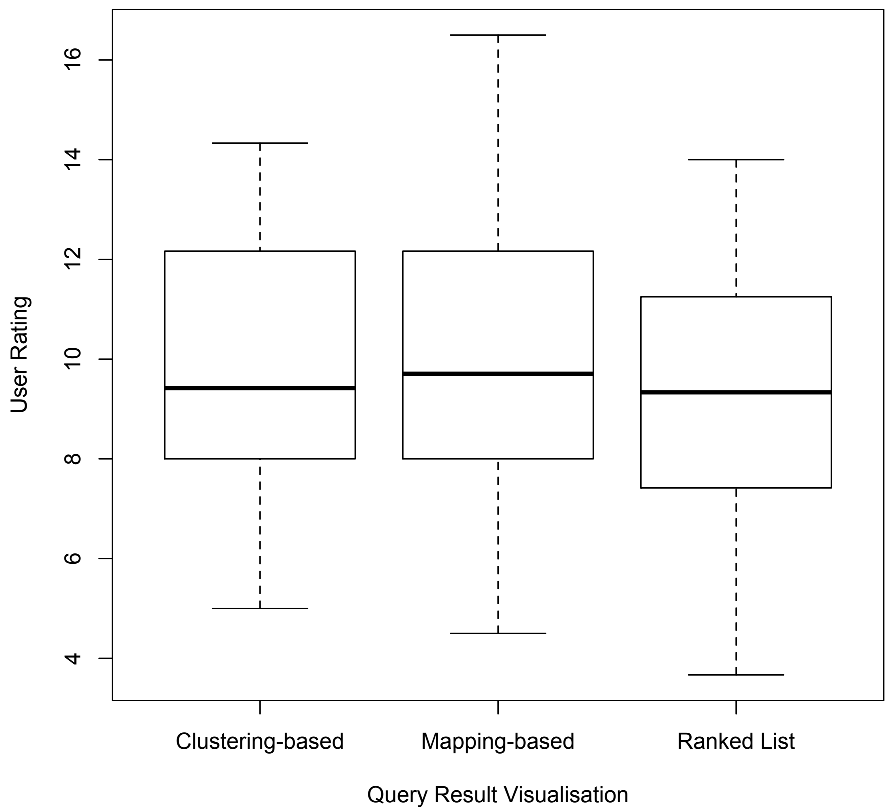


Figure 5.16: Box-plots of the overall workload (mean ratings of the six NASA-TLX subjective Workload assessment scales) experienced by participants using each of the query result presentations to create their image mosaics.

5.3.2.4 Preference

Whilst we did not find any significant differences between any of the Mosaic Test effectiveness measures across the three query result presentations, upon performing a Kruskal-Wallis test we did observe a significant difference in the preference ratings assigned by participants ($\chi^2_{(2)} = 6.1956, p = 0.045$). A post-hoc test Wilcoxon Signed-rank test - used to assess whether the mean ranks of populations differ - showed that the preference ratings assigned to the clustering-based visualisation of query results received significantly lower preference ratings than the ranked list presentation ($p = 0.021$). Interestingly, one participant of the study opted to assign a preference rating of zero to the mapping-based visualisation system they had tested.

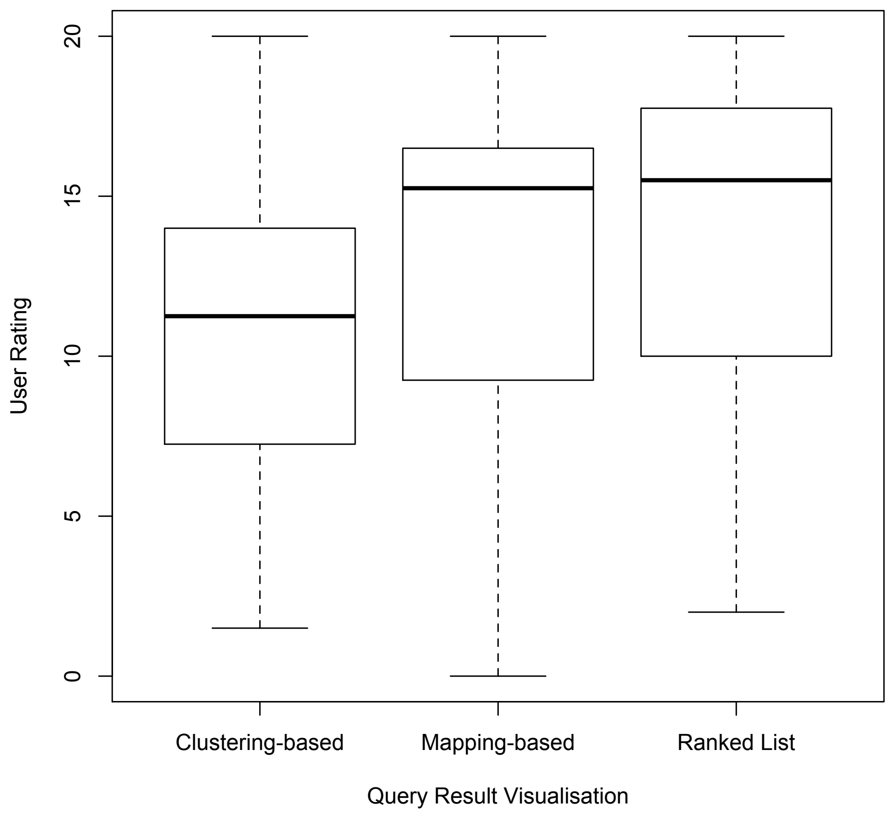


Figure 5.17: Box-plots of the preference ratings assigned to each of the query result presentations by participants for creating their image mosaics.

5.4 Discussion

In this Section, we discuss the results of our second user study. In addition, we discuss observations of test sessions and the aspects of our post-test discussions with study participants which we believe contributed towards the results reported above.

5.4.1 Query-by-colour and Query-by-sketch Query Styles

Recall from Section 5.2.1 our hypothesis that the query-by-sketch systems tested in the study would prove to be more effective for colour-based image retrieval than those adopting the query-by-colour query style. We formulated this hypothesis on account of the number of users in our first user study bemoaning their inability to specify the location of colours in queries. The statistical comparison of the frustration ratings between the two user groups - just short of our 5% significance level ($p = 0.0698$) - coupled with users of the query-by-colour systems expressing to us their frustration at not being able to specify colour locations in queries, appear to support our hypothesis.

The only anomaly in the results with regards to our hypothesis is the comparison of the times taken by users for creating their image mosaics - with users of the query-by-sketch systems taking significantly longer to create their mosaics than the users of the query-by-colour systems. A possible explanation for this result is related to a finding from our first study - that is, users spending more time searching for images with a given system due to having more confidence in its ability to find the most suitable images in a database. Given that users of the query-by-sketch systems assigned themselves significantly higher performance ratings than those using the query-by-colour systems, coupled with the image mosaics of query-by-sketch users being (almost significantly) more accurate than query-by-colour users, we believe this to be a plausible explanation.

We did not expect to observe users of the query-by-colour systems experience significantly more physical demand when creating their image mosaics than those using the query-by-sketch systems. On the contrary, due to the added interactions involved in sketching image requirements, we would have anticipated the opposite. One suggestion for this result occurring is that during test sessions, we observed users of the query-by-colour systems perform less queries than users of the query-by-sketch systems; if an image close to their requirements was not apparent immediately, query-by-sketch users would almost immediately amend their queries. Users of the query-by-colour systems, on the other hand, spent more time browsing through query results. The query-by-colour users may have experienced more physical demand than query-by-sketch users on account of the highly interactive nature of similarity-based visualisations - that is, the query-by-colour users would have performed many more zoom and pan operations in the mapping-based visualisation system, and more layer navigation operations in the clustering-based visualisation system, than query-by-sketch users.

5.4.2 Presentation of Query Results

In the study of Liu *et al.* [55], it was observed that users could find images fastest using tested systems presenting query results in the form of mapping and clustering-based visualisations (as opposed to the ranked list presentation). Whilst the search times recorded for the systems adopting mapping and clustering-based visualisations were not found to be significantly different, users did assign higher preference ratings to the mapping-based visualisation system - claiming it to be more intuitive, interesting, and convenient than the ranked list and clustering-based visualisation systems for searching and comparing images. The results of our study, however, appear to contradict these findings and instead side with our hypothesis that the ranked list presentation would prove to be the most effective of the three for colour-based image retrieval.

We found there to be no significant differences between the times required by users across the three colour-based image retrieval systems used (i.e., across the tested query result visualisations). Furthermore, we also found no significant differences in the accuracy of images mosaics created or workload ratings assigned by participants. We did, however, observe a significant difference between the recorded preference ratings. We observed that the systems presenting query results as a ranked list received significantly higher preference ratings than the systems presenting query results via a clustering-based visualisation. Post-test discussions with participants revealed that many found the groupings in the clustering-based visualisation to be unintuitive; with test users stating that the representative images of groups did not accurately reflect the images contained within them, and that many groups contained very dissimilar images. This reflects our criticism of clustering-based visualisations stated previously in Section 2.5.2.2 - that is, the visualisation style is heavily reliant upon the underlying features and metric used in defining similarity. Should the definition of similarity not match human perception well, an image can be placed within clusters that users would not intuitively think to inspect in order to retrieve it.

Whilst we did not find the difference in preference ratings between the ranked list and mapping-based visualisation query result visualisations to be significant, on average the ranked list approach did receive higher ratings from users. Many participants complained that the magnitude of visual information in the mapping-based visualisations of query results was `overwhelming' when searching for images. This was optimised by a participant assigning a zero preference rating to the mapping-based visualisation system they had tested.

5.5 Conclusion

In this Chapter, we have described our second user study which was undertaken to fulfil the final aim of this research; *to compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists*. Recall that Liu *et al.* [55] conducted a user study comparing three systems; presenting keyword query results either as a ranked list (where images ranked by the system as most similar to the query appear at the head of the list), a mapping-based visualisation, or a clustering-based visualisation. The results of Liu *et al.*'s study revealed that, of the three systems tested, users not only preferred the mapping-based visualisation system, but could also retrieve images significantly faster with the two similarity-based visualisation systems as opposed to when using the system presenting query results as a ranked list. Since Google Images [44] and Bing Images [17] currently adopt a ranked list for presenting both keyword and colour query results, we wanted to evaluate what effect changing the visualisation of results would have on the large number of users searching by colour (as opposed Liu *et al.*'s investigation into keywords).

Another motivation for conducting our second user study was to fulfil one more aspect of the research aim; *to evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval*. Given that many colour-based image retrieval systems typically support either the query-by-colour query style *or* the query-by-sketch query style [47, 42, 22, 44], but rarely both, and that no previous research had been conducted directly comparing the query styles, we wanted to provide designers creating future systems with guidance regarding which of the query styles must be included to best support users undertaking colour-based image retrieval.

For our second user study, we recruited 36 participants to compare six colour-based image retrieval systems using the Mosaic Test. Participants in our second user study were divided into two equally sized groups, the first of which created image mosaics using three systems adopting the query-by-colour query style, whilst the second group generated their mosaics using three query-by-sketch query systems. The results of our study suggest that query-by-sketch is more effective for colour based image retrieval than query-by-colour. Users of the query-by-sketch systems in our study experienced less frustration and physical demand than the users of the query-by-colour systems when creating their image mosaics. Whilst users of the query-by-colour systems were able to generate their image mosaics more quickly than the query-by-sketch users, we believe that users of the query-by-sketch systems spent more time searching for images using their systems due to having more confidence in their system's ability to find the most suitable images in a database. Given that users of the query-by-sketch systems assigned themselves significantly higher performance ratings than those using the query-by-colour systems, coupled with the image mosaics of query-by-sketch users being (almost significantly) more accurate than query-by-colour users, we believe this to be a plausible explanation.

An important point we must raise regarding our results comparing the query-by-colour and query-by-sketch query styles is that, for the Mosaic Test, users are searching for images with colours in specific locations. This, however, may not always be the case in the real world. For example, an interior designer may require images for a mood board that merely contain the colour(s) they require, with the location of the requisite colour(s) being inconsequential. In such a scenario, we suspect the query-by-colour query style could be more effective than query-by-sketch, on account of the easier query formulation (i.e., fewer interactions required). Nevertheless, the results of our user study have shown that when colour location is of importance to users, the query-by-sketch query style is the more effective of the two approaches. Consequently, we recommend that colour-based image retrieval systems should include both a query-by-sketch *and* a query-by-colour facility, to support those users for whom colour location is important, as well as those for whom it is not.

Within the two user groups of our second study, users tested three colour-based image retrieval systems visualising query results as either a ranked list, a mapping-based visualisation, or as a clustering-based visualisation. In contrast to those produced via Liu *et al.*'s study, we observed no significant difference in the times required by users across the tested query result visualisations (i.e., across the three colour-based image retrieval systems tested). Further to this, we also found no significant difference in the accuracy of images mosaics produced or workload ratings assigned by participants across the visualisations. We did, however, observe a significant difference between preference ratings. We observed that the systems presenting query results via a ranked list received significantly higher preference ratings than the systems presenting query results as a clustering-based visualisation. Whilst participants of Liu *et al.*'s study heralded mapping-based visualisations for presenting query results, participants of our second user study - as per participants of our first user study - complained that the magnitude of visual information in the mapping-based visualisation was overwhelming when searching for images. This was optimised by a participant assigning a zero preference rating to the mapping-based visualisation system they had tested. Consequently, these findings lead us to conclude that colour-based image retrieval query results should continue to be presented as ranked lists.

Chapter 6

Conclusions

Prior to our research, very few evaluations of colour-based image retrieval systems had ever been conducted. This, in our opinion, is a contributory factor as to why colour-based image retrieval systems today adopt the same query-based approaches first devised nearly two decades ago - originally designed to index image databases much smaller in magnitude than those available today. This, coupled with our discovery of a large demographic of users searching images databases by colour, provided the fundamental motivation for this research - to discover more effective approaches for colour-based image retrieval. By fulfilling our research aims, we believe we have made a significant contribution to the existing knowledge of colour-based image retrieval.

The main focus of this research was to explore the potential of similarity-based visualisations for colour-based image retrieval, due in main to them not having been evaluated previously specifically for searching image databases by colour. Whilst the findings of our research (summarised in the next section) cast doubt on the value of similarity-based visualisations for colour-based image retrieval (i.e., they appear to be no more effective than the currently adopted query-based approaches of query-by-colour and query-by-sketch), we must stress that, due to time and resource limitations, we have only been able to test a limited number of visualisations. It may be that using a different combination of feature vectors, spreading strategies, dimension reduction techniques (for mapping-based visualisations), or visualisation styles, could produce a more effective similarity-based visualisation for colour-based image retrieval which may have resulted in a different outcome to our research. Nevertheless, we have contributed to the field the first reported evaluations of similarity-based visualisations specifically for colour-based image retrieval.

Another contribution of this research is our comparison of the query-by-colour and query-by-sketch query styles for colour-based image retrieval undertaken as part of our second user study. Given that many colour-based image retrieval systems typically support either the query-by-colour query style *or* the query-by-sketch query style [47, 42, 22, 44], but rarely both, and that no previous research had been conducted directly comparing the query styles, we wanted to provide designers creating future systems with guidance as to which of the query styles must be included to best support users undertaking colour-based image retrieval. Our research has shown that, when colour location is of importance to users, query-by-sketch is the more effective of the two query styles. Consequently, we have contributed evidence to the field that colour-based image retrieval systems supporting only one of the colour-based image retrieval query styles are not suited to all users searching image databases by colour. We therefore recommend that colour-based image retrieval systems should now include both a query-by-sketch *and* a query-by-colour facility, to support those users for whom colour location is important as well as those for whom it is not.

The most significant contribution of this research, however, is the Mosaic Test; providing a starting point for ourselves and others in the field to undertake evaluations of systems and approaches for colour-based image retrieval. Prior to the research documented in this thesis, the field lacked a reliable method for systematically and meaningfully evaluating colour-based image retrieval systems. It is for this reason that we believe very few evaluations of systems and approaches for colour-based image retrieval have ever been conducted, which, as stated earlier, we believe to be a contributory factor as to why systems in use today adopt query-based approaches first devised nearly two decades ago. In this thesis, as well as proving that the Mosaic Test overcomes the two fundamental drawbacks of existing strategies for evaluating colour-based image retrieval systems and approaches, we have clearly documented how the Mosaic Test can be used to conduct future user-based evaluations to further advance the field towards more effective colour-based image retrieval.

In the following Section, we summarise the main findings of our research and show how we accomplished our research aims.

6.1 Summary of Main Findings

In this thesis, we have described our two user studies exploring the effectiveness of similarity-based visualisations for colour-based Image retrieval. In order for us to evaluate and compare the effectiveness of similarity-based visualisations and query-based approaches for colour-based Image retrieval, our first task was to overcome the lack of a reliable method for systematically and meaningfully evaluating colour-based image retrieval systems. Since no standardised image database with associated ground-truth existed for use in colour-based image retrieval evaluations, the precision and recall measures could not be adopted. Furthermore, the interactive target and category search task, alongside precision and recall, do not support *reflection-in-action*, the iterative image selection process often adopted by creative individuals. As a solution, we devised the Mosaic Test - a user-based evaluation approach in which evaluation study participants complete an image mosaic of a predetermined target image using the colour-based image retrieval system under evaluation. The time and users' perception of the workload required to complete this creative task, as well as the visual accuracy of their image mosaics (in comparison with the initial target images), are used to assess the effectiveness of the system being tested. To substantiate our claim that the Mosaic Test overcomes the major drawbacks associated with existing evaluation methods, we aimed to:

- Show that that the Mosaic Test provides a reliable mechanism by which to meaningfully evaluate and compare colour-based image retrieval systems.

We fulfilled this research aim in Chapter 3. From the findings of our first user study, in which we evaluated the Mosaic Test using twelve `expert' (i.e., users with experience working in the graphic design industry) and twelve `non-expert' participants (i.e., users with no prior graphic design experience), we concluded that the Mosaic Test overcomes the major drawbacks associated with previous evaluation methods: in addition to providing valuable effectiveness data relating to efficiency and user workload (by recording the task time and asking users to complete the NASA-TLX workload assessment tool), the Mosaic Test enables participants to reflect on the relevance of retrieved images within the context of their image mosaic (i.e., to perform reflection-in-action [84]), and automatically measures the relevance of retrieved images by computing MPEG-7 colour structure descriptors (from the user-generated image mosaics and target images quantised to 256 possible colours) and calculating the L_1 distance between them.

In Chapter 4, we detailed how the 24 participants of our first user study tested three colour-based image retrieval systems indexing 25,000 images. The first of these systems adopted the query-by-colour query style to facilitate colour-based image retrieval. The second system adopted a mapping-based visualisation which users could interactively explore using zooming and panning tools. These two systems were included in the study so that we could compare a mapping-based visualisation with the query-by-colour query style, contributing to the following research aim:

- Evaluate and compare the effectiveness of query-based approaches and similarity-based visualisations for colour-based image retrieval.

Participants of our first user study assigned higher preference ratings to the query-by-colour system than for the system featuring a mapping-based visualisation, as well as registering higher overall workload ratings for the latter. Analysis of the participant's subjective workload ratings revealed that users experienced significantly more effort, mental demand and frustration when creating image mosaics using the mapping-based visualisation systems than when using the query-by-colour system. Post-test discussions revealed that the frustration experienced and extra effort required by many participants when creating image mosaics using the mapping-based visualisation system - due to the seemingly random organisation of images causing users to accidentally revisit and review areas of the visualisation multiple times - as well as the mental demand required to process the large amount of visual information, impacted negatively on their preference ratings and workload ratings.

As well as the query-by-colour and mapping-based visualisation systems, participants of our first user study also tested a third colour-based image retrieval system. The third system tested featured the same mapping-based visualisation as the second system. For this third system, however, we removed the zoom facility so that users could only pan the visualisation to search for images. Removing this functionality from the third system enabled us to meet another of our research aims:

- Assess the importance of zooming for colour-based image retrieval from mapping-based visualisations of large image databases.

Our results show that participants of the study created their image mosaics significantly faster using the pannable-only mapping-based visualisation system than when using the zoomable system. We believe this could be due to the fact that fewer interactions (i.e., mouse operations) were required in order to navigate to an area of interest within the visualisation in the pannable-only system. We found there to be no significant difference between the other Mosaic Test measures recorded across the two systems. We did, however, also observe a significant difference in the users' preference ratings, with participants assigning higher ratings to the zoomable system. Many of the study participants suggested that they had felt limited whilst using the pannable-only system, preferring instead the ability to customise the number of visualisation images displayed on screen at one time via zooming. On the basis of this finding, we conclude that systems facilitating colour-based image retrieval via a mapping-based visualisation should include a zoom facility to improve the overall user experience.

Liu *et al.* [55] undertook a user study comparing three image retrieval systems which presented keyword query results either in the form of a ranked list (where images ranked by the system as most similar to the query appear at the head of the list), a mapping-based visualisation, or a clustering-based visualisation. It was observed that users could find images fastest using the systems presenting query results in the form of mapping-based and clustering-based visualisations (as opposed to a ranked list presentation). Furthermore, users assigned higher preference ratings to the mapping-based visualisation system. Since Google Images [44] and Bing Images [17] currently adopt a ranked list for presenting both keyword and colour query results, we wanted to evaluate what effect changing the visualisation of results would have on the large number of users searching by colour. It was for this reason that we added the research aim:

- Compare the effectiveness of colour-based image retrieval query result visualisations - namely, mapping and clustering-based visualisations against ranked lists.

For our second user study, we recruited 36 participants to compare six colour-based image retrieval systems using the Mosaic Test. Participants in our second user study were divided into two equally sized groups, the first of which created image mosaics using three systems adopting the query-by-colour query style, whilst the second group generated their mosaics using three query-by-sketch systems. Within the two user groups of our second study, users tested three colour-based image retrieval systems visualising query results as either a ranked list, a mapping-based visualisation, or as a clustering-based visualisation, enabling to satisfy the above research aim. In contrast to those produced via Liu *et al.*'s study, we observed no significant difference in the times required by users across the tested query result visualisations (i.e., across the three colour-based image retrieval systems tested). Further to this, we also found no significant difference in the accuracy of images mosaics produced or workload ratings assigned by participants across the visualisations. We did, however, observe a significant difference between preference ratings. We observed that the systems presenting query results via a ranked list received significantly higher preference ratings than the systems presenting query results as a clustering-based visualisation. Whilst participants of Liu *et al.*'s study heralded mapping-based visualisations for presenting query results, participants of our second user study - as per participants of our first user study - complained that the magnitude of visual information in the mapping-based visualisation was overwhelming when searching for images. This was optimised by a participant assigning a zero preference rating to the mapping-based visualisation system they had tested. Consequently, these findings lead us to conclude that colour-based image retrieval query results should continue to be presented as ranked lists.

The results of our second study suggest that query-by-sketch is more effective for colour based image retrieval than query-by-colour. Users of the query-by-sketch systems in our study experienced less frustration and physical demand than the users of the query-by-colour systems when creating their image mosaics. Whilst users of the query-by-colour systems were able to generate their image mosaics more quickly than the query-by-sketch users, we believe that users of the query-by-sketch systems spent more time searching for images using their systems due to having more confidence in their system's ability to find the most suitable images in a database. Given that users of the query-by-sketch systems assigned themselves significantly higher performance ratings than those using the query-by-colour systems, coupled with the image mosaics of query-by-sketch users being (almost significantly) more accurate than query-by-colour users, we believe this to be a plausible explanation. It is based on the above findings that we recommend colour-based image retrieval systems should include both a query-by-sketch *and* a query-by-colour facility.

6.2 Future Work

As a further contribution to knowledge, in this Section we outline several avenues for future work into colour-based image retrieval derived from this research. If conducted, we believe these channels of future work have the potential to further advance the field towards more effective colour-based image retrieval.

6.2.1 Increasing the Number of Mosaic Test Participants

First and foremost, we aim to make the Mosaic Test documentation and Mosaic Test Tool software available on-line so that others may systematically and meaningfully evaluate the effectiveness of colour-based image retrieval systems and approaches. To obtain accurate and reliable Mosaic Test effectiveness measures, it is imperative that any future studies evaluating colour-based image retrieval systems or approaches using the Mosaic Test recruit as many participants as is possible. Recruiting a large number of volunteers to participate in any user study, however, can prove to be a difficult, time-consuming and financially expensive operation. In the following Sections, we describe two approaches for overcoming these difficulties.

6.2.1.1 Crowd Sourcing

One way in which a much larger scale Mosaic Test could be conducted in the future would be to adapt the Mosaic Test Tool, as well as other Mosaic Test elements such as the size of the image mosaic participants are asked to generate, in order to make use of *crowd-sourcing* - whereby large-scale tasks are divided into much smaller sub-tasks which are then distributed to, and completed by, a large group of people for a small financial reward. For testing colour-based image retrieval systems or approaches using the Mosaic Test via crowd-sourcing, members of a crowd-sourcing community (such as Amazon's Mechanical Turk [2]) could be asked to complete a single cell (as opposed to the 16 cells as per our user studies) of a very large image mosaic using the system under evaluation. Upon retrieving an image from the database, the participant could then complete the NASA-TLX scales required for evaluating subjective workload for a Mosaic Test. The time taken by all participants to locate their image, the accuracy of the selected images (in comparison with the corresponding target image section) and the NASA-TLX ratings could then be used to reliably evaluate the effectiveness of the tested colour-based image retrieval system.

6.2.1.2 Mobile Application

Using the approach above will enable much larger Mosaic Tests to be conducted than those undertaken for this research. The number of users that can participate in any study, however, is inevitably capped by the financial budget allocated for participant remuneration. A solution to this limitation is to create a mobile application (or `app'), converting the Mosaic Test into a freely downloadable mini game for mobile platforms. In order to analyse the touch behaviour of smartphone users, Henze *et al.* [35] published a game entitled ``Hit It!'' on October 31st, 2010. In the game, users were required to touch target(s) on the screen within a set timeframe. If users touched the target(s) within the set time, they were advanced to the next level. Levels were made increasingly difficult by adding extra targets and/or reducing the time permitted to complete the task. If users failed to touch the target(s) in the allotted time they received a penalty point. Should users receive three penalty points, they were forced to start back at the first level. From their free `app', Henze *et al.* [35] collected more than 120 million touch events from 91,731 installations and used the data to determine the error rate for different target sizes and screen positions.

A similar approach could be undertaken to perform a much more comprehensive analysis of image mosaic accuracy measures than that implemented in our research. An `app' could be developed with which users can create their own image mosaics and have the accuracy of them rated by other users. These user-generated mosaics, coupled with the accuracy ratings assigned by users, could then be analysed in order to uncover which feature vector and associated distance measure is truly best suited for measuring the accuracy of image mosaics produced during a Mosaic Test.

6.2.2 Incorporating Keyword Search into the Mosaic Test

As reported previously in Section 2.4, Markkula and Sormunen [60] observed that colour is an important aesthetic criteria for journalists during the final selection phase of a typical retrieval session. In its current state, the Mosaic Test measures the effectiveness of colour-based image retrieval systems for users at this final stage of retrieval (or users not interested in the semantic concept of images). In most cases, however, users will have performed a keyword search (or multiple keyword searches) prior to this in order to find images with a particular semantic concept, before making his or her final selections on the basis of colour. Google Images [44] and Bing Images [17] are two examples of systems supporting this type of search. Whilst the Mosaic Test in its current form does not incorporate the keyword search stage of retrieval, as well as asking users to find suitably coloured images for their image mosaics, an added caveat could be that users are only permitted to use images semantically relevant to the target image. Users would then be required to undertake keyword searches to filter database images semantically before making their final selections on the basis of colour. To prevent users from submitting mosaics with semantically irrelevant images, an annotated database could be used (e.g., the MIR-FLICKR image collections [40] used in this research) and the Mosaic Test Tool modified to check that database images dragged to cells during a Mosaic Test share at least one keyword with those assigned to the target image (i.e., the Mosaic Test Tool would reject a database image during the test if it did not share at least one keyword with the target image).

6.2.3 Similarity-based Visualisation Variations

As discussed previously, the findings of this research cast doubt on the value of similarity-based visualisations for colour-based image retrieval (i.e., they appear to be no more effective than the currently adopted query-based approaches of query-by-colour and query-by-sketch). In our experiments, however, we adopted just one dimensional reduction technique (Principal Component Analysis) and only two feature vectors (namely the colour histogram [93] and the MPEG-7 colour layout descriptor [59] for our first and second user studies resepctively). As we have described previously in Chapter 2, there are countless combinations of feature vectors (and associated distance measures), dimensional reduction techniques (e.g., multi-dimensional scaling [52], Fastmap [24], etc.), spreading strategies, and visualisation styles, that could yield more effective similarity-based visualisations for colour-based image retrieval than those produced for this study. Of course, the effectiveness of these alternative similarity-based visualisations could be assessed using our Mosaic Test.

6.2.4 The Query-by-colour and Query-by-sketch Query Styles

The findings from our second user study suggest that whilst query-by-colour is fastest for searching image databases by colour, the query-by-sketch query style enables users to find the database images most suited to their colour requirements (i.e., users could produce more accurate image mosaics). As per the similarity-based visualisations used for this research, however, there are several modifications that could be made to both the query-by-colour and query-by-sketch systems evaluated as part of our research which may make them more effective for colour-based image retrieval. For instance, adding a colour proportion selection tool (with which users can specify the quantities of selected colours) to the query-by-colour system may enable users to find database images more suited to their colour requirements (i.e., users could produce more accurate image mosaics in a Mosaic Test). For the query-by-sketch system, improving the drawing tool in the query panel could potentially reduce the amount of time it takes users to retrieve images from a database. Changing the feature vectors and distance metrics used for our colour-based image retrieval systems could also enhance their effectiveness.

The design of the colour palette used in all of our query-based systems was identical to that of the MultiColr [42] system, mainly due to the positive reviews it had received online (e.g., [26]). As reported previously in Section 2.6.2, it would appear that Broek *et al.* [98] is the only example of research into the design of colour palettes for colour-based image retrieval, but is only a theoretical examination (based on a heuristic incorporating Fitt's Law - a mathematical model of human movement) of graphical colour palettes taken from a variety of different systems. As such, future work could be undertaken to empirically compare various colour palette designs using the Mosaic Test; to discover which truly is most effective for colour-based image retrieval.

Publications

Aspects of the work described in this thesis feature in the following publications:

- Plant, W., Lumsden, J., and Nabney, I.T., "The Mosaic Test: Measuring the Effectiveness of Colour-based Image Retrieval" in *Multimedia Tools and Applications*, p.1-22, 2012.
- Plant, W., Lumsden, J., and Nabney, I.T., (2011), *The Mosaic Test: Benchmarking Colour-based Image Retrieval Systems Using Image Mosaics*, in Proceedings of the First European Workshop on Human-Computer Interaction and Information Retrieval, p. 23-26, Newcastle, U.K., 4 July.
- Plant, W., and Schaefer, G., (2011), Visualisation and Browsing of Image Databases, in *Multimedia Analysis, Processing and Communications*, Ed. W. Lin et al., Springer Berlin / Heidelberg, Vol. 346, p. 3-57.

Bibliography

[1] M. Abdel-Mottaleb, S. Krischnamachari, and N.J. Mankovich. Performance Evaluation of Clustering Algorithms for Scalable Image Retrieval. In *IEEE Computer Society Workshop on Empirical Evaluation of Computer Vision Algorithms*, 1998.

[2] Amazon.com, Inc. Amazon Mechanical Turk. World Wide Web Page: `https://www.mturk.com/mturk/welcome`, accessed 21st September 2012.

[3] D. Atsmon and A. Atsmon. Picitup. World Wide Web Page: `http://www.picitup.com/`, accessed 21st September 2012.

[4] S. Ayache and G. Quéenot. Video Corpus Annotation using Active Learning. In *European Conference on Information Retrieval (ECIR)*, pages 187--198, Glasgow, Scotland, mar 2008.

[5] K.U. Barthel. Automatic Image Sorting using MPEG-7 Descriptors. In *Workshop On Immersive Communication And Broadcast Systems*, Berlin, Germany, October 2005.

[6] B. Bederson. Quantum Treemaps and Bubblemaps for a Zoomable Image Browser. In *14th ACM Symposium on User Interface Software and Technology*, pages 71--80, Orlando, Florida, USA, Novemeber 2001.

[7] D. Borth, C. Schulze, A. Ulges, and T. Breuel. Navidgator - Similarity Based Browsing for Image and Video Databases. In *31st German Conference on Advances in Artificial Intelligence*, pages 22--29, Kaiserslautern, Germany, September 2008.

[8] R. Cabral. ImgSeek. World Wide Web Page: `http://www.imgseek.net/`, accessed 21st September 2012.

[9] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1026--1038, 2002.

[10] Cerious Software, Inc. ThumbsPlus. `http://www.cerious.com/`, accessed 21st September 2012.

[11] H. C. Chan and Y. Wang. Human Factors in Color-based Image Retrieval: An Empirical Study on Size Estimate Accuracies. *Journal of Visual Communication and Image Representation*, 15(2):113--131, 2004.

[12] C. Chen. *Information Visualization*. Springer-Verlag, New York, USA, 2004.

[13] C. Chen, G. Gagaudakis, and P. Rosin. Similarity-based Image Browsing. In *International Conference on Intelligent Information Processing*, pages 206--213, Beijing, China, August 2000.

[14] Y. Chen and A. Butz. PhotoSim: Tightly Integrating Image Analysis into a Photo Browsing UI. In *8th International Symposium on Smart Graphics*, pages 224--231, Rennes, France, August 2008.

[15] P. Clough, H. Müller, and M. Sanderson. Seven years of image retrieval evaluation. In H. Müller, P. Clough, T. Deselaers, and B. Caputo, editors, *ImageCLEF*, volume 32 of *The Information Retrieval Series*, pages 3--18. Springer Berlin Heidelberg, 2010.

[16] T. Combs and B. Bederson. Does Zooming Improve Image Browsing? In *4th ACM Conference on Digital Libraries*, pages 130--137, California, USA, 1999.

[17] Microsoft Corp. Bing Images. World Wide Web Page: `http://www.bing.com/images`, accessed 21st September 2012.

[18] Corvida. Search Flickr Images by Color With the Multicolr Search Lab. World Wide Web Page: `http://www.readwriteweb.com/archives/search_flickr_images_by_color_with_multicolr_search_lab.php`, accessed 21st September 2012.

[19] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 40(2):1--60, 2008.

[20] A. Del-Bimbo, M. Mugaini, P. Pala, and F. Turco. Visual Querying by Colour Perceptive Regions. *Pattern Recognition*, 31(9):1241--1253, 1998.

[21] H. Ding, J. Liu, and H. Lu. Hierarchical Clustering-based Navigation of Image Search Results. In *16th ACM International Conference on Multimedia*, pages 741--744, Vancouver, Canada, October 2008.

[22] Exalead. Chromatik. World Wide Web Page: `http://chromatik.labs.exalead.com/`, accessed 21st September 2012.

[23] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber. Efficient and Effective Querying by Image Content. *Journal of Intelligent Information Systems*, 3:231--262, 1994.

[24] C. Faloutsos and K. Lin. FastMap: A Fast Algorithm for Indexing, Datamining and Visualization of Traditional and Multimedia Datasets. In *ACM SIGMOD International Conference on Management of Data*, pages 163--174, San Jose, California, May 1995.

[25] R. Feinson. A Brief History of Mosaics. World Wide Web Page: `http://www.royfeinson.com/doubletakeimages/briefhistory.html`, accessed 21st September 2012.

[26] M. Frauenfelder. Multicolr - Stunningly Elegant Flickr Tool. World Wide Web Page: `http://boingboing.net/2008/10/27/multicolr-stunningly.html`, accessed 21st September 2012.

[27] D. Gasienica. OpenZoom. World Wide Web Page: `http://www.openzoom.org/`, accessed 21st September 2012.

[28] A. Gomi, R. Miyazaki, T. Itoh, and J. Li. CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique. In *12th International Conference on Information Visualization*, pages 82--87, London, UK, July 2008.

[29] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as Essence for Photo Browsing through Personal Digital Libraries. In *2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 326--335, Portland, Oregon, USA, July 2002.

[30] A. Gupta and R. Jain. Visual Information Retrieval. *Communications of the ACM*, 40(5):70--79, 1997.

[31] S. G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. In *Human Factors and Ergonomics Society 50th Annual Meeting*, pages 904--908, San Francisco, USA, October 2006.

[32] D. Heesch. A Survey of Browsing Models for Content Based Image Retrieval. *Multimedia Tools and Applications*, 40(2):261--284, 2008.

[33] D. Heesch and S. Rüger. NN^k Networks for Content-Based Image Retrieval. In *26th European Conference on Information Retrieval*, pages 253--266, Sunderland, UK, April 2004.

[34] D. Heesch and S. Rüger. Three Interfaces for Content-Based Access to Image Collections. In *3rd International Conference on Image and Video Retrieval*, pages 491--499, July, Dublin, Ireland 2004.

[35] N. Henze, E. Rukzio, and S. Boll. 100,000,000 taps: Analysis and improvement of touch performance in the large. In *Proceedings of 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2011)*, pages 133--142, Stockholm, Sweden, September 2011.

[36] O. Hilliges. Finding the Unknown - Serendipitous Discovery in Co-Located Consumption of Digital Photo Collections. In *Workshop Collocated Social Practices Surrounding Photos in conjunction with the ACM CHI*, Florence, Italy, April 2008.

[37] O. Hilliges, P. Kunath, A. Pryakhin, A. Butz, and H.P. Kriegel. Browsing and Sorting Digital Pictures using Automatic Image Classification and Quality Analysis. In *12th International Conference on Human-Computer Interaction*, pages 882--891, Beijing, China, July 2007.

[38] J. Huang, S. Ravi Kumar, M. Mitra, and W. Zhu. Spatial Color Indexing and Applications. *International Journal of Computer Vision*, 35(3):245--268, 1999.

[39] J. Huang, S. Ravi Kumar, M. Mitra, W. Zhu, and R. Zabih. Image Indexing Using Color Correlograms. In *Computer Vision and Pattern Recognition*, pages 762--768, 1997.

[40] M. J. Huiskes and M. S. Lew. The MIR Flickr Retrieval Evaluation. In *1st ACM International Conference on Multimedia Information Retrieval*, pages 39--43, Vancouver, Canada, October 2008.

[41] A. Ibrahim, A. Al-Zou'bi, R. Sahawneh, and M. Makhadmeh. Fixed Representative Colors Feature Extraction Algorithm for Moving Picture Experts Group-7 Dominant Color Descriptor. *Computer Science*, 5(11):773--777, 2009.

[42] idée Inc. idée MultiColr Search Lab. World Wide Web Page: `http://labs.ideeinc.com/multicolr`, accessed 21st September 2012.

[43] Imagekind Inc. Shop Art by Color. World Wide Web Page: `http://www.imagekind.com/shop/ColorPicker.aspx`, accessed 21st September 2012.

[44] Google Inc. Google Images. World Wide Web Page: `http://images.google.com/`, accessed 21st September 2012.

[45] Yahoo! Inc. Flickr. `http://www.flickr.com/`, accessed 21st September 2012.

[46] iStockPhoto LP. iStockPhoto.com. World Wide Web Page: `http://www.istockphoto.com/`, accessed 21st September 2012.

[47] C. E. Jacobs, A. Finkelstein, and D. H. Salesin. Fast Multiresolution Image Querying. In *22nd International Conference on Computer Graphics and Interactive Techniques*, pages 277--286, Los Angeles, California, USA, August 1995.

[48] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., New Jersey, USA, 1988.

[49] Y. Jing and S. Baluja. VisualRank: Applying PageRank to Large-Scale Image Search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1877--1890, 2008.

[50] I. Keller, T. Meiers, T. Ellerbrock, and T. Sikora. Image Browsing with PCA-Assisted User-Interaction. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 102--108, Kauai, Hawaii, December 2001.

[51] S. Krischnamachari and M. Abdel-Mottaleb. Image Browsing using Hierarchical Clustering. In *4th IEEE Symposium Computers and Communications*, pages 301--307, Red Sea, Egypt, July 1999.

[52] J.B. Kruskal and M. Wish. *Multidimensional Scaling*. Sage, California, USA, 1978.

[53] T. K. Lau and I. King. Montage: An Image Database for the Fashion, Textile, and Clothing Industry in Hong Kong. In *3rd Asian Conference on Computer Vision*, pages 410--417, Hong Kong, January 1998.

[54] S. Lim, R. Smith, and G. Lu. i-Map: An Interactive Visualisation and Navigation System of an Image Database for Finding a Sample Image to Initiate a Visual Query. In *Conference of the Australian Computer-Human Interaction Special Interest Group (CHISIG) of the Human Factors and Ergonomics Society of Australia (HFESA)*, Wollongong, Australia, November 2004.

[55] H. Liu, X. Xie, X. Tang, Z. W. Li, and W. Y. Ma. Effective Browsing of Web Image Search Results. In *6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 84--90, New York, USA, October 2004.

[56] G. Loy and J-O. Eklundh. A Review of Benchmarking Content Based Image Retrieval. In *MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation*, Vienna, Austria, September 2005.

[57] W. Y. Ma and B. S. Manjunath. NeTra: A Toolbox for Navigating Large Image Databases. *Multimedia Systems*, 7(3):184--198, 1999.

[58] L.W. MacDonald. Using Color Effectively in Computer Graphics. *IEEE Computer Graphics and Applications*, 19(4):20--35, 1999.

[59] B. S. Manjunath, J.R. Ohm, V. V. Vasudevan, and A. Yamada. Color and Texture Descriptors. *IEEE Transaction on Circuits and Systems for Video Technology*, 11(6):703--715, 2001.

[60] M. Markkula and E. Sormunen. End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. *Information Retrieval*, 1:259--285, 2000.

[61] S. McDonald and J. Tait. Search Strategies in Content-based Image Retrieval. In *ACM Conference on Research and Development in Information Retrieval*, pages 80--87, 2003.

[62] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. Huang. Visualization and User-Modeling for Browsing Personal Photo Libraries. *International Journal of Computer Vision*, 56(1/2):109--130, 2004.

[63] H. Müller, S. Marchand-Maillet, and T. Pun. The Truth About Corel - Evaluation in Image Retrieval. In *International Conference on Image and Video Retrieval*, pages 38--49, London, UK, July 2002.

[64] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet, and T. Pun. Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals. *Pattern Recognition Letters*, 22(5):593--601, 2001.

[65] S. Nakade and P.T. Karule. Mosaicture: Image Mosaic Generating System Using CBIR Technique. In *International Conference on Computational Intelligence and Multimedia Applications*, pages 339--343, Sivakasi, India, December 2007.

[66] M. Nakazato, L. Manola, and T. Huang. ImageGrouper: A Group-Oriented User Interface for Content-Based Image Retrieval and Digital Image Arrangement. *Journal of Visual Language and Computing*, 14(4):363--386, 2003.

[67] G.P. Nguyen and M.Worring. Interactive Access to Large Image Collections using Similarity Based Visualization. *Journal of Visual Languages and Computing*, 19:203--224, 2008.

[68] T. Osman, D. Thakker, G. Schaefer, and P. Lakin. An Integrative Semantic Framework for Image Annotation and Retrieval. In *IEEE / WIC / ACM International Conference on Web Intelligence*, pages 366--373, California, USA, November 2007.

[69] Z. Pecenovic, M. Do, M. Vetterli, and P. Pu. Integrated Browsing and Searching of Large Image Collections. In *4th International Conference on Advances in Visual Information Systems*, pages 279--289, Lyon, France, November 2000.

[70] A. Pentland, W.R. Picard, and S. Sclaroff. Photobook: Content-Based Manipulation of Image Databases. *International Journal of Computer Vision*, 18(3):233--254, 1996.

[71] C. Plaisant, D. Carr, and B. Shneiderman. Image Browsers: Taxonomy, Guidelines, and Informal Specifications. *IEEE Software*, 12:21--32, 1995.

[72] J. Platt, M. Czerwinski, and B. Field. PhotoTOC: Automatic Clustering for Browsing Personal Photographs. Technical report, Microsoft Research, 2002.

[73] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Evaluating a Visualisation of Image Similarity as a Tool for Image Browsing. In *IEEE Symposium on Information Visualisation*, pages 36--43, San Francisco, California, USA, October 1999.

[74] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. A Comparison of Measures for Visualising Image Similarity. In *3rd UK Conference on Image Retrieval*, Brighton, United Kingdom, May 2000.

[75] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does Organisation by Similarity Assist Image Browsing? In *SIGCHI Conference on Human Factors in Computing Systems*, pages 190--197, Seattle, Washington, USA, April 2001.

[76] K. Rodden and K. Wood. How Do People Manage Their Digital Photographs? In *SIGCHI Conference on Human Factors in Computing Systems*, pages 409--416, Florida, USA, April 2003.

[77] Y. Rubner, L. J. Guibas, and C. Tomasi. The Earth Movers Distance, Multi-dimensional Scaling, and Color-based Image Retrieval. In *APRA Image Understanding Workshop*, pages 661--668, May 1997.

[78] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40(2):99--121, 2000.

[79] M. Sarkar and M. Brown. Graphical Fisheye Views. *Communications of the ACM*, 37(12):73--83, 1994.

[80] G. Schaefer and S. Ruszala. Image Database Navigation: A Globe-al Approach. In *International Symposium on Visual Computing*, pages 279--286, Nevada, USA, December 2005.

[81] G. Schaefer and S. Ruszala. Image Database Navigation on a Hierarchical Hue Sphere. In *International Symposium on Visual Computing*, pages 814--823, 2006.

[82] G. Schaefer and S. Ruszala. Image Database Navigation on a Hierarchical MDS Grid. In *28th Symposium of the German Association for Pattern Recognition*, pages 304--313, Berlin, Germany, September 2006.

[83] G. Schaefer and M. Stich. UCID - An Uncompressed Colour Image Database. In *Storage and Retrieval Methods and Applications for Multimedia*, pages 472--480, San Jose, California, USA, January 2004.

[84] D. A Schön. *The Reflective Practitioner: How Professionals Think in Action*. Temple Smith, London, 1983.

[85] S. Schroeder. Google Adds Color Filter to Image Search. World Wide Web Page: `http://mashable.com/2009/04/08/google-adds-color-filter-to-image-search/`, accessed 21st September 2012.

[86] N. Sebe, M. Lew, and D. P. Huijsmans. Which ranking metric is optimal? with applications in image retrieval and stereo matching. In *International Conference on Pattern Recognition*, pages 265--271, 1998.

[87] T. Sikora. The MPEG-7 Visual Standard for Content Description - An Overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):696--702, 2001.

[88] R. Silvers. Photomosaics: Putting Pictures in their Place. Master's thesis, Massachusetts Institute of Technology, 1996.

[89] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349--1380, 2000.

[90] J. R. Smith and S. Chang. VisualSEEk: A Fully Automated Content-based Image Query System. In *4th ACM International Conference on Multimedia*, pages 87--98, Boston, USA, November 1996.

[91] R. Stehling, M. Nascimento, and A. Falcao. *Multimedia Mining: A Highway to Intelligent Multimedia Documents*, chapter Techniques for Color-Based Image Retrieval, pages 61--80. Kluwer Academic Publishers, 2002.

[92] M. Stricker. Bounds for the discrimination power of color indexing techniques. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, volume 2185, pages 15--24, 1994.

[93] M. Swain and D. Ballard. Color Indexing. *International Journal of Computer Vision*, 7(1):11--32, 1991.

[94] D. Tanguay. Search the Rainbow. World Wide Web Page: `http://googleblog.blogspot.co.uk/2009/04/search-rainbow.html`, accessed 21st September 2012.

[95] Michael Terry and Elizabeth D. Mynatt. Recognizing Creative Needs in User Interface Design. In *4th Conference on Creativity and Cognition*, pages 38--44, Loughborough, UK, October 2002.

[96] N. Tran. Generating Photomosaics: An Empirical Study. In *ACM Symposium on Applied Computing*, pages 105--109, Texas, USA, March 1999.

[97] E.L. van den Broek, L.G. Vuurpijl, and P. Kisters. Content-based Image Retrieval: Color-selection Exploited. In *3rd Dutch-Belgian Information Retrieval Workshop*, pages 37--46, Leuven, Belgium, 2002.

[98] E.L. van den Broek, L.G. Vuurpijl, P. Kisters, and J.C.M. von Schmid. Design Guidelines for a Content-Based Image Retrieval Color-Selection Interface. In *Conference on Dutch Directions in HCI*, pages 14--17, Amsterdam, Netherlands, June 2004.

[99] M. Vijfwinkel. CG Textures. World Wide Web Page: `http://www.cgtextures.com/`, accessed 21st September 2012.

[100] J. Wang, X.S. Hua, and Y. Zhao. Color-Structured Image Search. Technical Report MSR-TR-2009-82, Microsoft Research, 2009.

[101] S. Westman, A. Lustila, and P. Oittinen. Search Strategies in Multimodal Image Retrieval. In *2nd International Symposium on Information Interaction in Context*, pages 13--20, London, UK, October 2008.

[102] C. Yang. Content-based Image Retrieval: A Comparison Between Query by Example and Image Browsing Map Approaches. *Journal of Information Science*, 30(3):254--267, 2004.

[103] Y. Zhang, M.A. Nascimento, and O.R. Zaiane. Building Image Mosaics: An Application of Content-based Image Retrieval. In *International Conference on Multimedia and Expo*, pages 317--320, Baltimore, USA, July 2003.

Appendix A

Mosaic Test Documents

This appendix contains the background questionnaire, Mosaic Test instructions, and NASA-TLX workload assessment, used for the two studies described in this thesis.

Participant #: ________

Some Background Information

To let us assess how relevant elements of people's life experience and/or physical makeup impact on the way they search for images based on colour, we would ask you to answer the following series of questions. **Please note, this information will remain anonymous; you are not required to complete any question that you do not wish to answer.** If you have any queries, please do not hesitate to ask the researcher.

1. Please indicate your age by ticking the box next to the age-range that applies to you:

☐ 15 – 20 years	☐ 36 – 40 years	☐ 56 – 60 years
☐ 21 – 25 years	☐ 41 – 45 years	☐ 61 – 65 years
☐ 26 – 30 years	☐ 46 – 50 years	☐ 66 + years
☐ 31 – 35 years	☐ 51 – 55 years	

2. Please indicate your gender by ticking the appropriate box:

☐ Male ☐ Female

3. Please indicate how often you search or browse a collection of images, e.g. looking for a particular picture or theme:

☐ Never **(Go to Question 5)**

☐ Daily
☐ At least once a week
☐ 2-3 times a month
☐ Less than once a month
} **(Go to Question 4)**

4. Please specify the name of any systems or websites you may use to search or browse a collection of images:

____

____

____

____

5. Please specify the highest formal qualification you possess in Art or a related discipline (leave blank if you have no such qualifications):

____

____

Figure A.1: Pre-study questionnaire.

Introduction

You are being asked to help evaluate 3 computer systems that allow users to search a large collection of images (or photographs) to find a specific or desired image. A collection of images is often referred to as an **image database**. Traditionally, the most common way of searching for an image in an image database is by entering a set of descriptive words (known as keywords) into a search system. The computer systems you will be testing today do not need a set of keywords to search for images; instead they allow you to search for images based on colours. We refer to such systems as **Search By Colour Systems**.

The search by colour systems you will be testing today enable graphic designers or others in creative industries to search a large collection of images using colours. **Please remember that it is the technology that is being evaluated, not you!**

Aston University

Illustration removed for copyright restrictions

Figure 1 - An example of a Photomosaic.

In this study, you will be required to complete 3 Photomosaics. A Photomosaic (or photograph mosaic – see Figure 1 for an example) is an image that is made up of a collection of lots of smaller, identically sized images. The images making up the mosaic are chosen on the basis of their colours. When a Photomosaic is viewed from a distance, the smaller images collectively appear to form the target image.

IMPORTANTLY, the closer the match between the colours in the small mosaic images and the section of the larger image they are used to fill, the more accurate and visually effective the Photomosaic will be!

Figure A.2: Mosaic Test Tool instructions (*page one*).

Figure 2 – Target image (left) and photomosaic (right).

Today, you will be asked to complete a series of photomosaics of jelly beans using different image search systems to find the images you need. In each case, you will be shown the target image (see Figure 2 – left). You will also see an empty photomosaic (see Figure 2 – right) of the target image. Each of the photomosaics you will be completing today consist of 16 cells that need to be filled in with images from the image database. Your task is to search the image database using a search by colour system to find 16 images with the right colours to complete the 16 cells in the mosaic and insert those images into the photomosaic. **Remember that the closer the match between the colours of the selected images and the corresponding colours in the target image, the more accurate and visually effective the Photomosaic will appear.**

The researcher will determine the order in which you use the different search by colour systems. You will be given some training on how to use each search by colour system, and allowed to practise using the system before completing each photomosaic. Once you have completed and submitted your photomosaic, you will be asked to complete a series of ratings concerning how difficult you felt the task was to complete. After you have completed the three photomosaics using the three different search by colour systems, you will be asked to rate your order of preference for the three search by colour systems you have used.

Figure A.3: Mosaic Test Tool instructions (*page two*).

Mosaic Test Tool :: Training Session

Before you begin to test the three search by colour systems, we first need to explain how to complete a mosaic using our **Mosaic Test Tool** (see Figure 3). This Mosaic Test Tool (which will always appear on the left of the screen) has been developed specifically to support the creation of photomosaics. To let you get comfortable with creating photomosaics, we have set up a sample search by colour system (see Figure 3 - right) that will allow you to try out the Mosaic Test Tool without worrying about what your actions might do.

Please feel free to ask questions at any point during the training session – it is important that you are comfortable with the Mosaic Test Tool before you proceed!

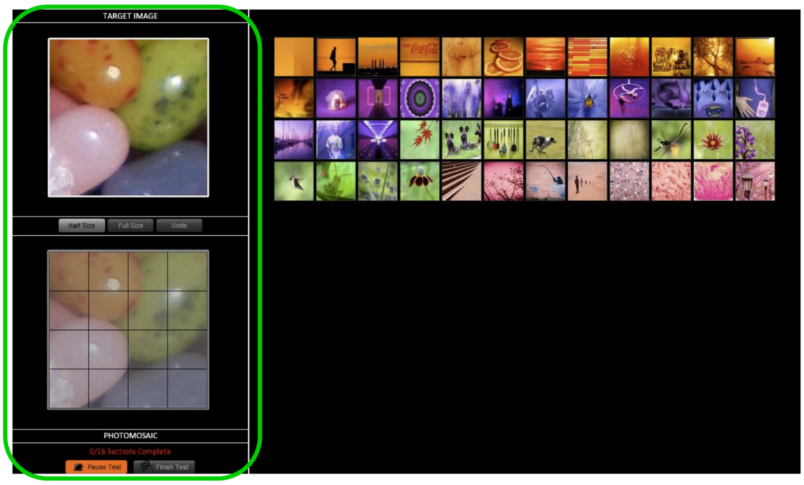


Figure 3 – The Mosaic Test Tool (left – circled in green) and the sample search by colour system (right).

You should have in front of you a laptop (with mouse attached) displaying the Mosaic Test Tool and the sample search by colour system as shown in Figure 3. The screen will currently be blacked out with a message box in the centre. The Mosaic Test Tool consists of 2 main parts. First is the **target image**, located in the top portion of the tool. This is the image you are aiming to replicate using smaller images in the photomosaic. Below the target image, and a small control panel (which we will describe later), is the initially empty **photomosaic**. You are required to fill in the photomosaic by dragging images from a search by colour system.

The 16 incomplete cells of the photomosaic have as a background a faded version of the corresponding area in the target image (see Figure 4). You should use this as a guide for searching for images within the database. Hovering the mouse cursor over a photomosaic cell will outline the corresponding area within the target image in **BLUE** (see Figure 4). Feel free to try this now. So, for example, the cell with a **BLUE** outline in the photomosaic in Figure 4 would require you to search for and insert an **ORANGE** image to match the corresponding area in the target image.

Figure A.4: Mosaic Test Tool instructions (*page three*).

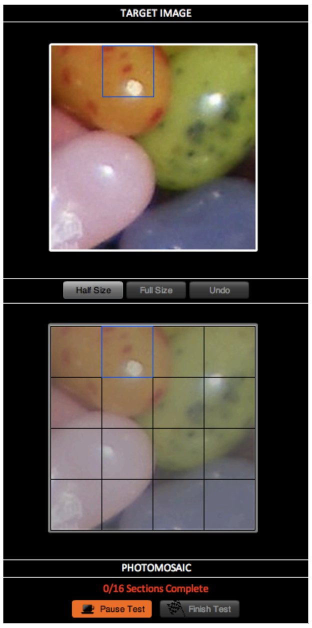


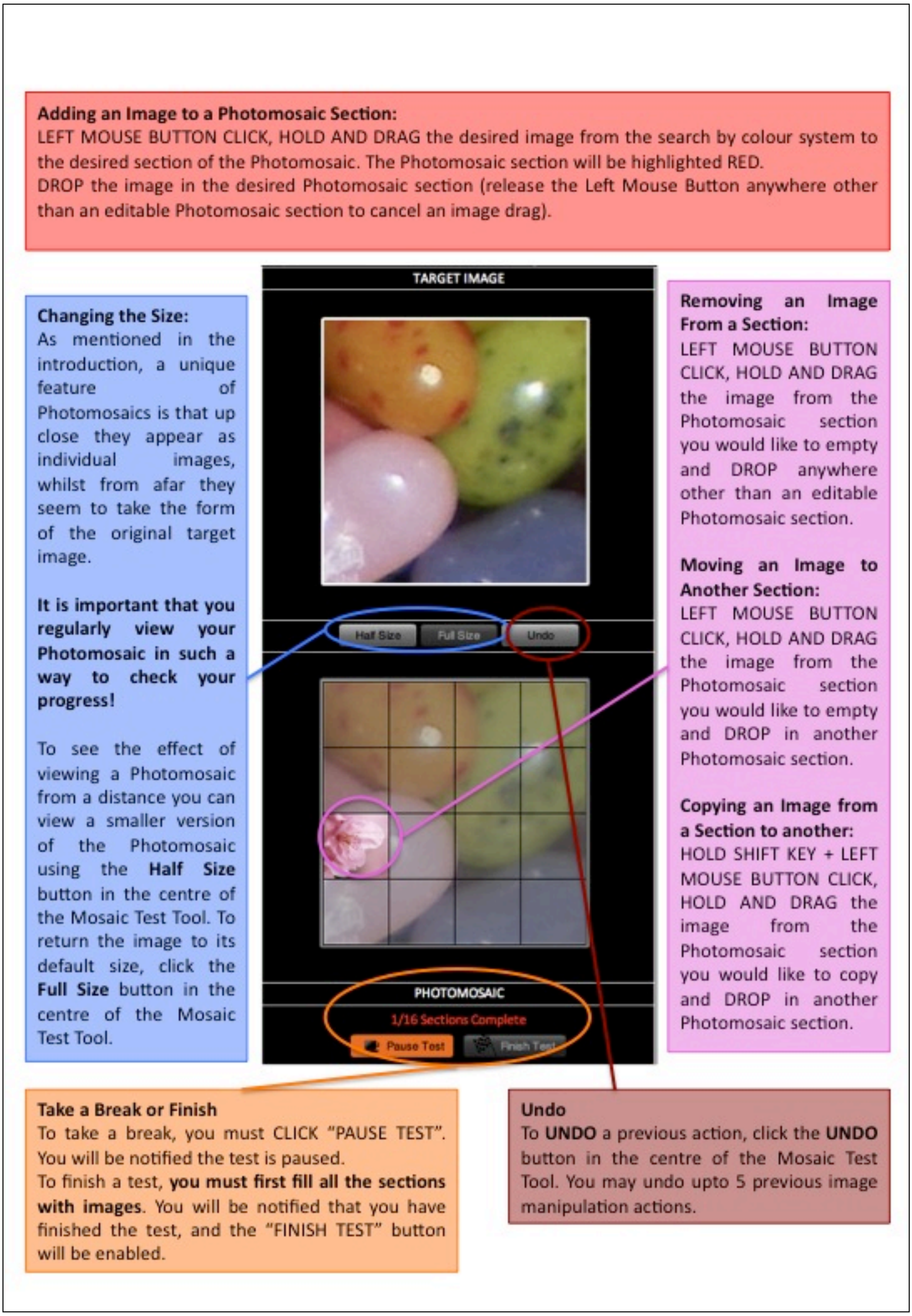
Figure 4 – The incomplete cells of the photomosaic.

So now you know what you are required to do, you need to know how to add images you find in the database to your photomosaic, as well as moving, copying or deleting images in the photomosaic. You will also need to know the various ways in which you can manually assess the quality of you current photomomosaic. These actions are all explained on the following page.

Finally, it is important that you understand how to submit a completed photomosaic or take a break from the experiment. The next page also explains these actions too.

You may use the following page as a guide throughout this study.

Figure A.5: Mosaic Test Tool instructions (*page four*).



Adding an Image to a Photomosaic Section:
LEFT MOUSE BUTTON CLICK, HOLD AND DRAG the desired image from the search by colour system to the desired section of the Photomosaic. The Photomosaic section will be highlighted RED.
DROP the image in the desired Photomosaic section (release the Left Mouse Button anywhere other than an editable Photomosaic section to cancel an image drag).

Changing the Size:
As mentioned in the introduction, a unique feature of Photomosaics is that up close they appear as individual images, whilst from afar they seem to take the form of the original target image.

It is important that you regularly view your Photomosaic in such a way to check your progress!

To see the effect of viewing a Photomosaic from a distance you can view a smaller version of the Photomosaic using the **Half Size** button in the centre of the Mosaic Test Tool. To return the image to its default size, click the **Full Size** button in the centre of the Mosaic Test Tool.

Removing an Image From a Section:
LEFT MOUSE BUTTON CLICK, HOLD AND DRAG the image from the Photomosaic section you would like to empty and DROP anywhere other than an editable Photomosaic section.

Moving an Image to Another Section:
LEFT MOUSE BUTTON CLICK, HOLD AND DRAG the image from the Photomosaic section you would like to empty and DROP in another Photomosaic section.

Copying an Image from a Section to another:
HOLD SHIFT KEY + LEFT MOUSE BUTTON CLICK, HOLD AND DRAG the image from the Photomosaic section you would like to copy and DROP in another Photomosaic section.

Take a Break or Finish
To take a break, you must CLICK "PAUSE TEST". You will be notified the test is paused.
To finish a test, **you must first fill all the sections with images**. You will be notified that you have finished the test, and the "FINISH TEST" button will be enabled.

Undo
To **UNDO** a previous action, click the **UNDO** button in the centre of the Mosaic Test Tool. You may undo upto 5 previous image manipulation actions.

Figure A.6: Mosaic Test Tool instructions (*page five*).

Participant #: ________

Mosaic Test Tool :: Training Task

So that you are comfortable with everything you will be asked to do during the study, please click **START and** complete the photomosaic of jelly beans using images found using the sample search by colour system. Use the scroll bar or mouse wheel to scroll through all the images in the system.

Remember:

- **The closer the match between the colours of the selected images and the corresponding target image sections, the more accurate and visually effective the photomosaic will appear.**

To familiarise yourself fully with the Mosaic Test Tool, please check off and complete the following tasks:

- ☐ Drag an image from the search by colour system to a photomosaic cell (Left Click, Hold and Drag image and Drop in desired photomosaic cell).
- ☐ Move an image from one photomosaic section to another (Left Click, Hold and Drag image and Drop in desired photomosaic cell).
- ☐ Copy an image within a photomosaic cell and place the duplicate in another photomosaic cell (Shift Key + Left Click, Hold and Drag image to be copied and Drop in another photomosaic cell).
- ☐ Remove an image from the photomosaic (Left Click, Hold and Drag image and Drop somewhere outside of the red border area).
- ☐ Pause the Mosaic Test Tool (Left Click "Pause Test" button).
- ☐ Make the photomosaic half size and full size.
- ☐ Submit a completed photomosaic (Left Click "Finish Test" button when all cells of the photomosaic contain an image and you are happy with your end product).

Feel free to ask the researcher for assistance at any point. You may repeat the whole photomosaic if you would like further practice, or progress to the actual experiment whenever you feel are ready – just ask the researcher!

Figure A.7: Mosaic Test Tool `training tasks' check sheet.

Workload Assessment - Introduction

We would like you to complete some rating scales designed to find out about your experiences regarding the image search by colour system you have just used to create your photomosaic. In particular, we would like to know about the "workload" you experienced. The factors that influence your perception of workload associated with interacting with the system to complete the photomosaic may come from the system itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The workload contributed by these factors may be different for the different search by colour systems. The physical parts of workload are easy to measure but the mental ones are harder.

Since workload is something that is experienced individually by each person, we need to measure it by asking each person to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually. This set of 6 scales was developed for you to use in evaluating your experiences in different tasks.

Please read the definitions of the scales carefully. **If you have a question about any of the scales please ask the researcher**. It is extremely important that they be clear to you.

Thinking carefully about the image search by colour system, we would like you to fill in the 6 scales. **You should evaluate the session by marking each scale at the point that matches your experience**. **Each line has a description at each end: please consider each scale individually**. **Please consider your responses carefully**. Your ratings will play an important role in the evaluation being conducted so your active participation is essential to the success of this experiment and is greatly appreciated. Please remember that you are free to terminate your contribution to this study at any time and you do not have to complete any of the scales if you do not wish to. If you have any questions, please do not hesitate to ask the researcher.

Figure A.8: NASA-TLX workload assessment [31] background information.

Participant #: ________

NASA-TLX Subjective Workload Assessment

MENTAL DEMAND How much mental, visual, and auditory activity was required? (e.g. thinking, deciding, calculating, remembering, looking, searching, listening, scanning, looking etc.)

MENTAL DEMAND

Low **High**

PHYSICAL DEMAND How much physical activity was required ? (e.g. pushing, pulling, turning, controlling, activating etc.)

PHYSICAL DEMAND

Low **High**

TEMPORAL DEMAND (TIME PRESSURE) How much time pressure did you feel due to the rate or pace at which the task elements occurred? (e.g. slow, leisurely, rapid, frantic)

TEMPORAL DEMAND

Low **High**

PERFORMANCE LEVEL ACHIEVED How successful do you think you were in accomplishing the task(s) set by the experimenter? How satisfied were you with your performance? Don't just think of your 'score', but how you felt you performed.

PERFORMANCE

Poor **Good**

EFFORT EXPENDED How hard did you have to work (mentally and physically) to accomplish your level of performance?

EFFORT

Low **High**

FRUSTRATION LEVEL How much frustration did you experience? (e.g. were you relaxed, content, stressed, irritated, discouraged?)

FRUSTRATION

Low **High**

Figure A.9: NASA-TLX workload assessment [31].

Participant #: ________

Preference Assessment

You have now tested all the search by colour systems. We would like you to tell us which system (A, B and C) you preferred. Below is a scale identical to those that you have been using to indicate your workload. We would like you to complete it as before but this time put an 'A' on the scale to indicate your preference for system A, a 'B' on the scale to indicate your preference for system B, and 'C' to indicate your preference for system C. So, for example, if you preferred system A to the system B, you would put the 'A' nearer the High end of the scale than the 'B'.

Please now indicate your preferences on the following scale:

OVERALL PREFERENCE Please rate your preference for the systems using the following identifiers:
(A) for System A
(B) for System B
(C) for System C

Low **High**

We would like to know which of the photomosaics you created you feel was your most accurate (or the one you think was best). In the list below, rank your photomosaics according to how good you think each was: put 1 in the box next to the photomosaic you were happiest with, and so on. If necessary, you may score 2 or indeed all 3 photomosaics with the same rank.

Please now indicate your photomosaic preferences:

☐ First Photomosaic
☐ Second Photomosaic
☐ Third Photomosaic

We would also like to know in what situations or for which tasks, if any, you think it might be beneficial to use a search by colour system such as those you have tested today. Please list any suggestions below:

Finally, if you have any other comments regarding this experiment, please take a moment to note them here:

You have now completed this experiment. Thank you for your participation.

Figure A.10: System preference assessment.

Appendix B

First Study System Instructions

This appendix contains the system instructions provided to participants of the first user study.

System A :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool. We will now introduce you to the **System A** for searching for images according to their colour(s). Before you begin to complete the photomosaic using system A, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using System A:**

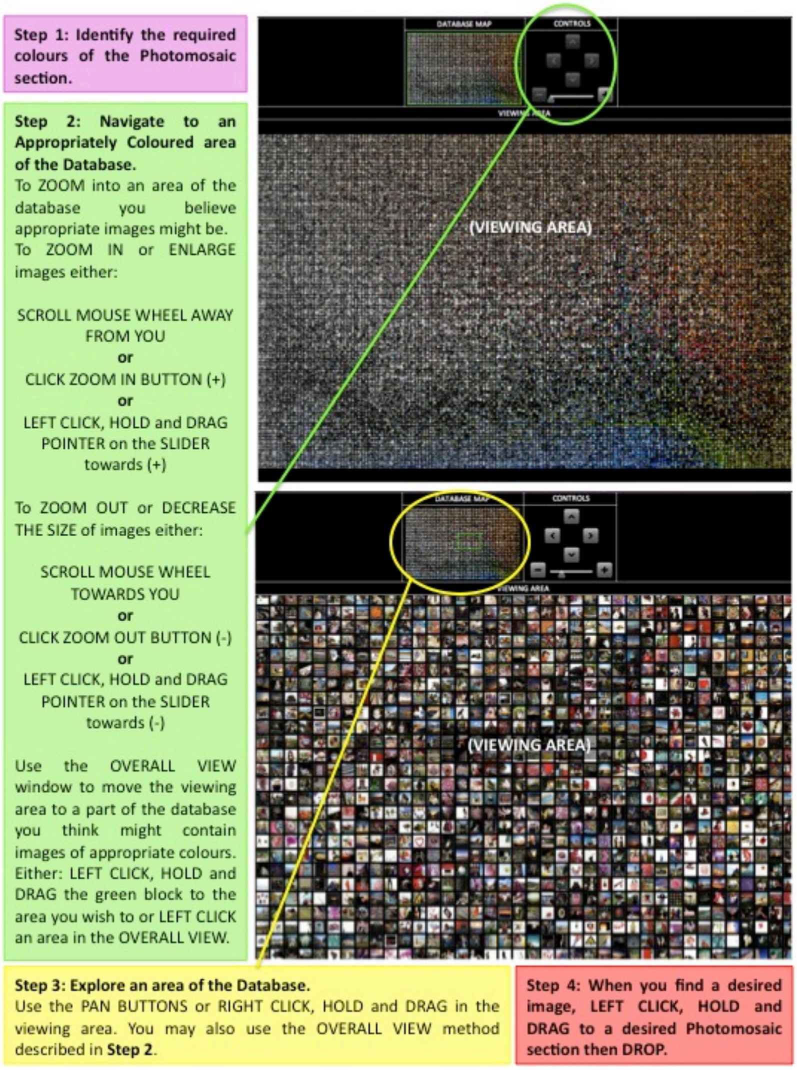


Figure B.1: Zoomable mapping-based visualisation system training instructions.

System B :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool. We will now introduce you to the **System B** for searching for images according to their colour(s). Before you begin to complete the photomosaic using system A, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using System B:**

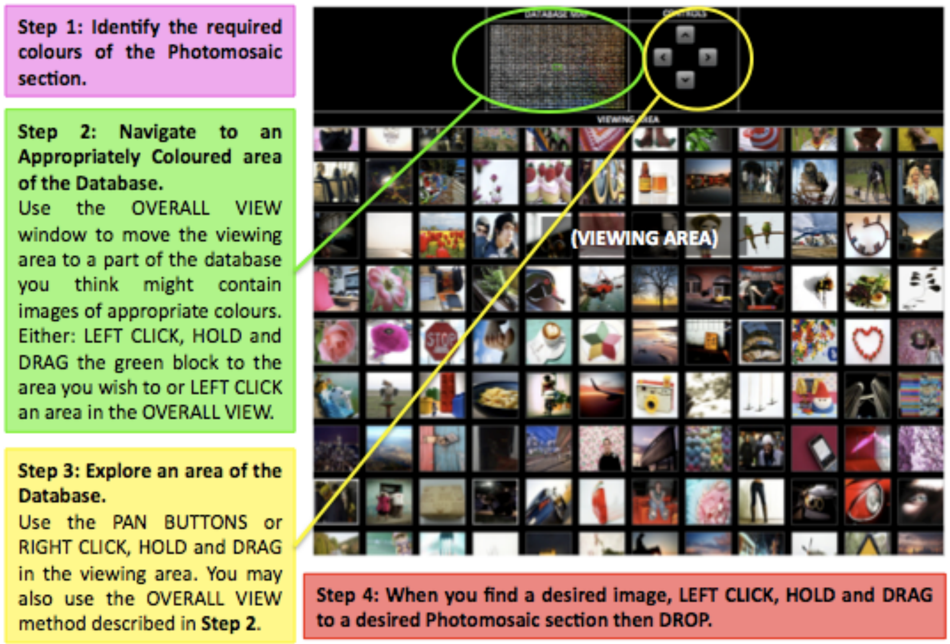


Figure B.2: Pannable-only mapping-based visualisation system training instructions.

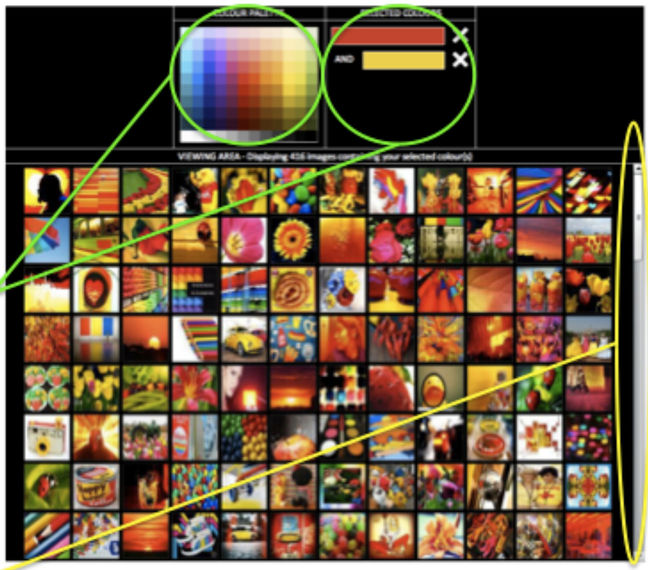
System C :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool. We will now introduce you to **System C** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system C, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system C:**

Step 1: Identify the required colours of the Photomosaic section (in this case, Orange and Yellow)

Step 2: Choose and Click the colours on the Colour Palette. If you require twice as much of a colour as other colours in your image, select the same colour twice! Your current colour combination is shown to the right of the colour palette. Click the **WHITE** cross to remove a colour from you search.

Step 3: Navigate through the results. Use the scrollbar or mouse wheel to scroll through the results of your search. The number of image results is displayed below the colour palette.



Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired Photomosaic section then DROP.

You now have the opportunity to have a practise with system C without worrying about what your actions might do! Please feel free to ask questions at any point whilst practising – it is important that you are comfortable with searching for images using the system C before you proceed to completing the test photomosaic.

When you are ready to begin your practise session with System C, please click "START" on the message box on the computer screen. When you are happy to proceed to the experimental session, please inform the researcher.

Figure B.3: Query-by-colour system training instructions.

Appendix C

Image Mosaic Measure Accuracy Rankings

This appendix contains a full list of the Spearman's rank correlation coefficients (r_s) between the image mosaic distance rankings made by participants and the rankings generated by the tested image mosaic accuracy measures.

Rank	Descriptor	Quantisation	Metric	r_s
8	APP	N/A	L_2	0.275
16	MLCH	64	L_2	0.165
23	MPEG-7 CLD	N/A	DEFAULT	0.132
15	MPEG-7 CST	32	L_1	0.167
5	MPEG-7 CST	64	L_1	0.355
2	**MPEG-7 CST**	**128**	$\mathbf{L_1}$	**0.418**
1	**MPEG-7 CST**	**256**	$\mathbf{L_1}$	**0.528**
11	MPEG-7 CST	32	L_2	0.22
4	MPEG-7 CST	64	L_2	0.363
7	MPEG-7 CST	128	L_2	0.33
6	MPEG-7 CST	256	L_2	0.341
12	MPEG-7 SCD	16	L_1	0.209
31	MPEG-7 SCD	32	L_1	0.078
22	MPEG-7 SCD	64	L_1	0.133
41	MPEG-7 SCD	128	L_1	-0.11
18	MPEG-7 SCD	256	L_1	0.154
9	MPEG-7 SCD	16	L_2	0.253
40	MPEG-7 SCD	32	L_2	0.011
14	MPEG-7 SCD	64	L_2	0.176
30	MPEG-7 SCD	128	L_2	0.088
3	**MPEG-7 SCD**	**256**	$\mathbf{L_2}$	**0.407**
27	LOCAL SCD	16	L_1	0.11
32	LOCAL SCD	32	L_1	0.077
37	LOCAL SCD	64	L_1	0.022
27	LOCAL SCD	128	L_1	0.11
27	LOCAL SCD	256	L_1	0.11
23	LOCAL SCD	16	L_2	0.132
36	LOCAL SCD	32	L_2	0.055
37	LOCAL SCD	64	L_2	0.022
23	LOCAL SCD	128	L_2	0.132
18	LOCAL SCD	256	L_2	0.154
37	ACC	16	L_1	0.022
12	ACC	32	L_1	0.209
18	ACC	64	L_1	0.154
16	ACC	128	L_1	0.165
18	ACC	256	L_1	0.154
34	ACC	16	L_2	0.066
23	ACC	32	L_2	0.132
9	ACC	64	L_2	0.253
34	ACC	128	L_2	0.066
32	ACC	256	L_2	0.077

Table C.1: Spearman's rank correlation coefficients (r_s) between the image mosaic distance rankings made by participants and the rankings generated by the tested image mosaic accuracy measures.

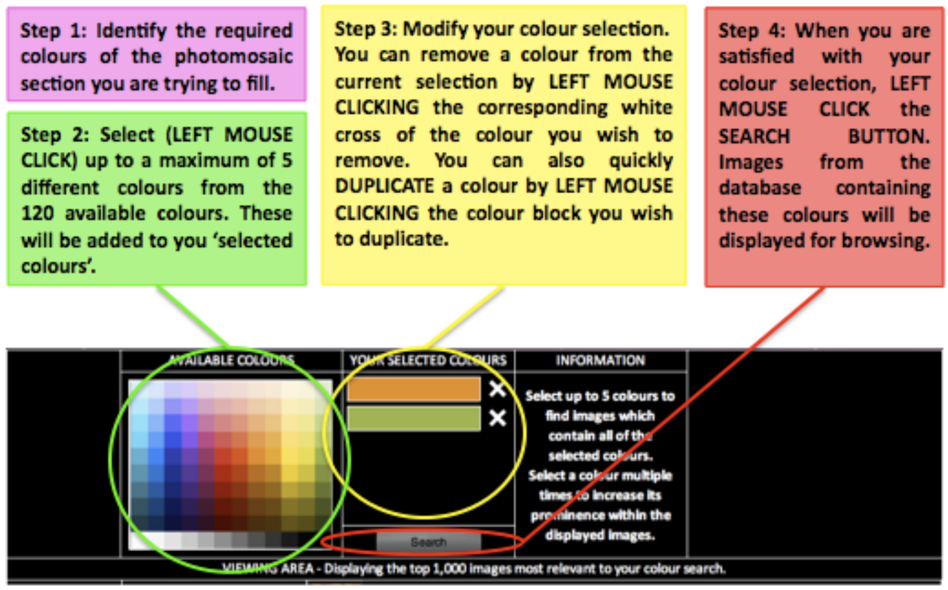
Appendix D

Second Study Query-by-colour Systems Instructions

This appendix contains the instructions provided to participants testing the query-by-colour systems in the second user study.

Colour-Picker Toolbar :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool. We will now introduce you to the **COLOUR-PICKER TOOLBAR,** which is included on each of the systems you will be using today and will always appear at the top of the system. Before you begin to complete photomosaics using each of the 3 image search by colour systems, we first need to explain how you should use the **COLOUR-PICKER TOOLBAR** to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using the COLOUR-PICKER TOOLBAR:**



You should now be familiar with searching for images using the **COLOUR-PICKER TOOLBAR**. Don't worry, you will get a chance to practise with the **COLOUR-PICKER TOOLBAR** when you begin training with the first of the 3 image search by colour systems. If you do have any questions, do not hesitate to ask your researcher now.

Figure D.1: Query-by-colour query panel instructions.

System D :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-PICKER TOOLBAR**. We will now introduce you to **System D** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system D, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system D:**

Step 1: Identify the required colours of the photomosaic section (in this case, Orange and Yellow)

Step 2: Use the COLOUR-PICKER TOOLBAR to enter your colour requirements, then LEFT MOUSE CLICK SEARCH to find similarly coloured images.

Step 3: Navigate through the results.
Use the scrollbar or mouse wheel to scroll through the results of your search.

Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired photomosaic section then DROP.

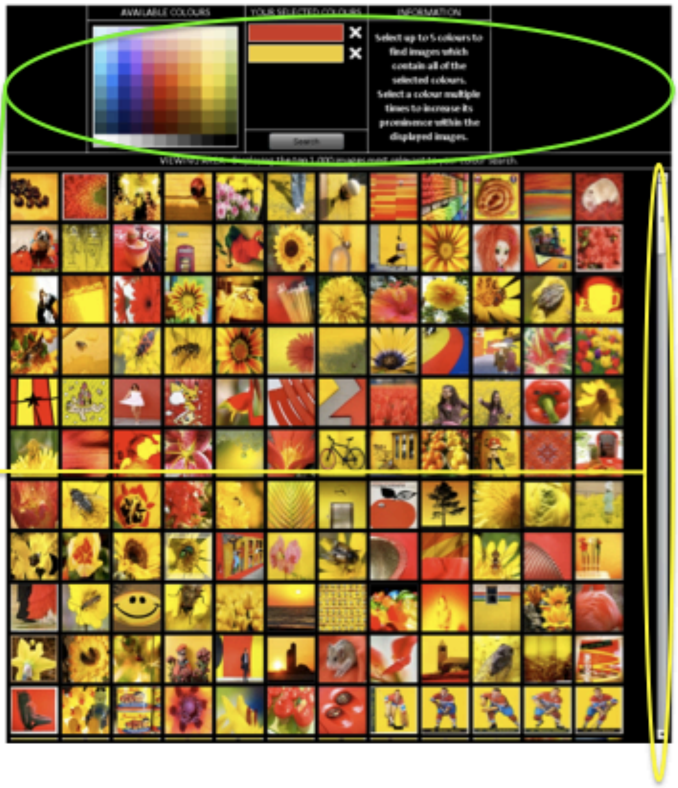


Figure D.2: Ranked list query-by-colour system instructions.

System E :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-PICKER TOOLBAR**. We will now introduce you to **System E** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system E, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system E:**

Step 1: Identify the required colours of the photomosaic section (in this case, Orange and Yellow)

Step 2: Use the COLOUR-PICKER TOOLBAR to enter your colour requirements, then LEFT MOUSE CLICK SEARCH to find similarly coloured images.

Step 3: Navigate through the results.

To ZOOM into an area of the results you believe appropriate images might be.
To ZOOM IN or ENLARGE images either:

SCROLL MOUSE WHEEL AWAY FROM YOU
or
CLICK ZOOM IN BUTTON (+)
or
LEFT CLICK, HOLD and DRAG POINTER on the SLIDER towards (+)

To ZOOM OUT or DECREASE THE SIZE of images either:

SCROLL MOUSE WHEEL TOWARDS YOU
or
CLICK ZOOM OUT BUTTON (-)
or
LEFT CLICK, HOLD and DRAG POINTER on the SLIDER towards (-)

Use the PAN BUTTONS or RIGHT CLICK, HOLD and DRAG in the viewing area to MOVE AROUND the search results.

Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired photomosaic section then DROP.

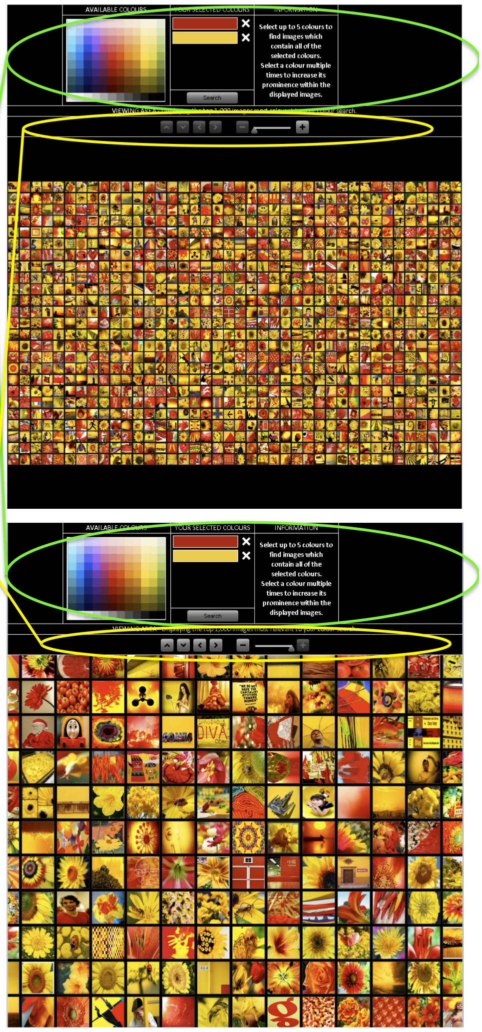


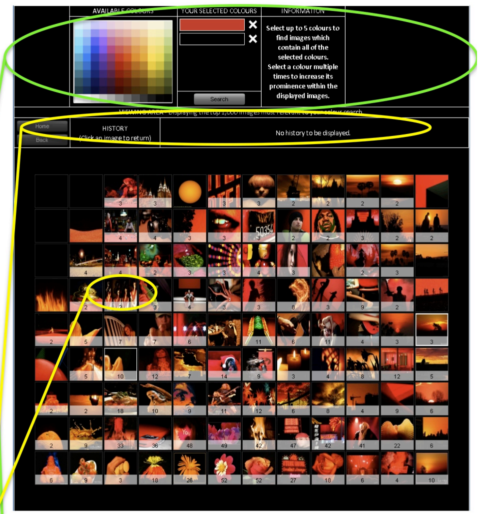
Figure D.3: Mapping-based visualisation query-by-colour system instructions.

System F :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-PICKER TOOLBAR**. We will now introduce you to **System F** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system F, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system F:**

Step 1: Identify the required colours of the photomosaic section.

Step 2: Use the COLOUR-PICKER TOOLBAR to enter your colour requirements, then LEFT MOUSE CLICK SEARCH to find similarly coloured images.

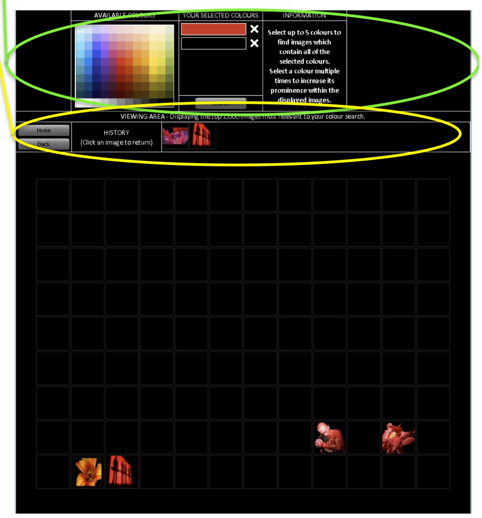


Step 3: Navigate through the results.

Image search results are grouped together based on their similarity. Each group has a representative image. The number overlaying the representative image is the number of images contained within that group. If there is only a single image in the group, no number is displayed.

To view a group of images, DOUBLE CLICK the LEFT MOUSE BUTTON on the representative image group (note this will not work on a single image).

You will notice that once the group images have been displayed, the representative image of the group (the image you **Double Clicked**) is added to the **History Bar**. You can click the image to return back to the group it was in if needed. The **HOME** and **BACK** buttons also enable you to navigate between groups.



Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired photomosaic section then DROP. This can be done on single images OR images representing a group (with a number).

Figure D.4: Clustering-based visualisation query-by-colour system instructions.

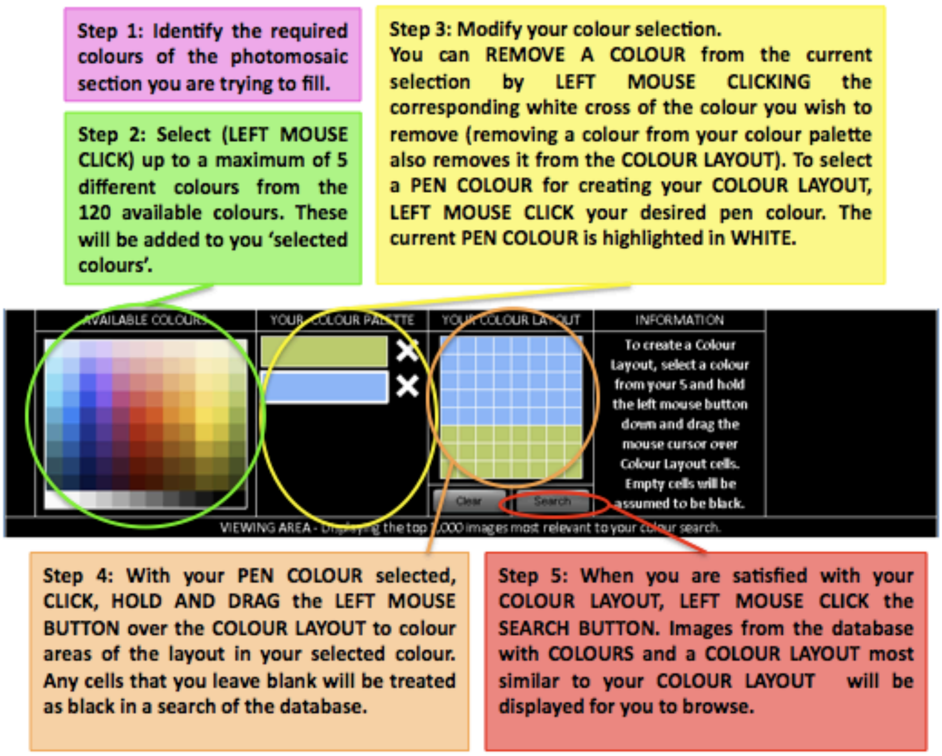
Appendix E

Second Study Query-by-sketch Systems Instructions

This appendix contains the instructions provided to participants testing the query-by-sketch systems in the second user study.

Colour-Layout Toolbar :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool. We will now introduce you to the **COLOUR-LAYOUT TOOLBAR**, which is included on each of the systems you will be using today and will always appear at the top of the system. Before you begin to complete your photomosaics using each of the 3 image search by colour systems, we first need to explain how you should use the **COLOUR-LAYOUT TOOLBAR** to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using the COLOUR-LAYOUT TOOLBAR:**



You should now be familiar with searching for images using the **COLOUR-LAYOUT TOOLBAR**. Don't worry, you will get a chance to practise with the **COLOUR-LAYOUT TOOLBAR** when you begin training with the first of the 3 image search by colour systems. If you do have any questions, do not hesitate to ask your researcher now.

Figure E.1: Query-by-sketch query panel instructions.

System G :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-LAYOUT TOOLBAR**. We will now introduce you to **System G** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system G, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system G:**

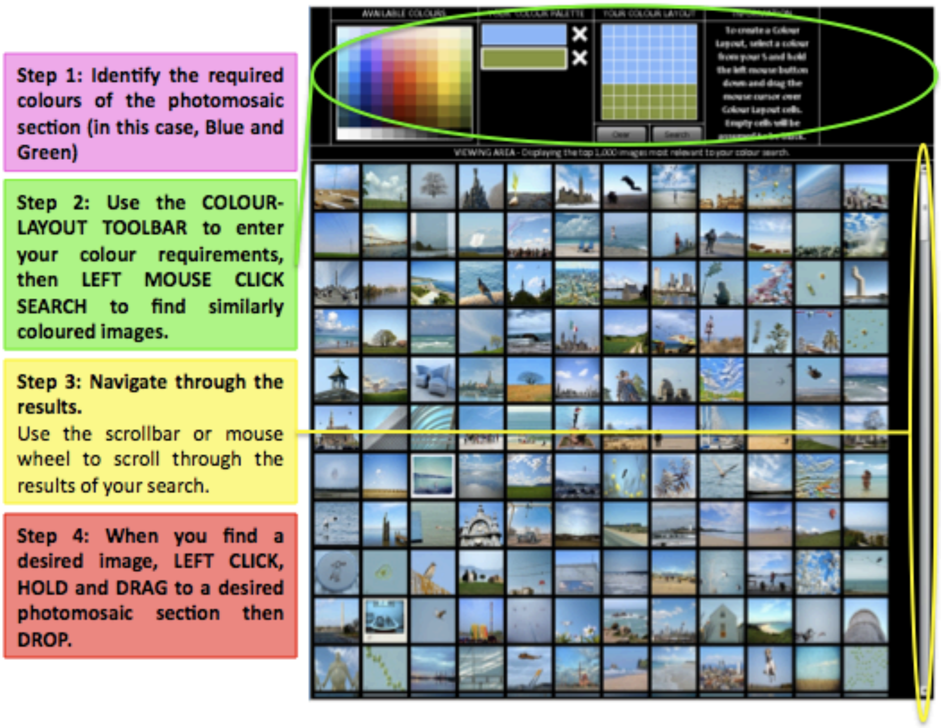


Figure E.2: Ranked list query-by-sketch system instructions.

System H :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-LAYOUT TOOLBAR**. We will now introduce you to **System H** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system H, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system H:**

Step 1: Identify the required colours of the photomosaic section (in this case, Blue and Green)

Step 2: Use the COLOUR-LAYOUT TOOLBAR to enter your colour requirements, then LEFT MOUSE CLICK SEARCH to find similarly coloured images.

Step 3: Navigate through the results.

To ZOOM into an area of the results you believe appropriate images might be.
To ZOOM IN or ENLARGE images either:

SCROLL MOUSE WHEEL AWAY FROM YOU
or
CLICK ZOOM IN BUTTON (+)
or
LEFT CLICK, HOLD and DRAG POINTER on the SLIDER towards (+)

To ZOOM OUT or DECREASE THE SIZE of images either:

SCROLL MOUSE WHEEL TOWARDS YOU
or
CLICK ZOOM OUT BUTTON (-)
or
LEFT CLICK, HOLD and DRAG POINTER on the SLIDER towards (-)

Use the PAN BUTTONS or RIGHT CLICK, HOLD and DRAG in the viewing area to MOVE AROUND the search results.

Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired photomosaic section then DROP.

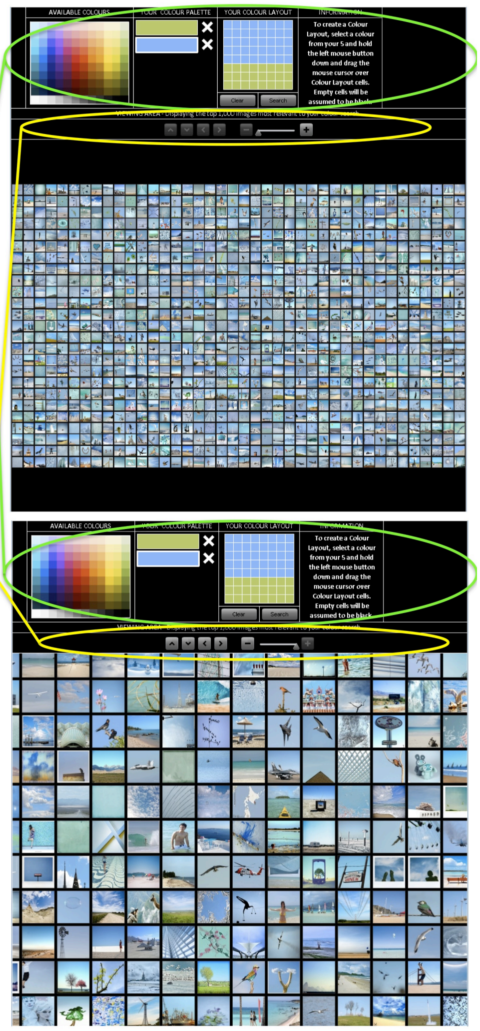


Figure E.3: Mapping-based visualisation query-by-sketch system instructions.

System I :: Training Session

You should now be familiar with how to complete a photomosaic using the Mosaic Test Tool, and how to use the **COLOUR-LAYOUT TOOLBAR**. We will now introduce you to **System I** for searching for images according to their colour(s). Before you begin to complete the photomosaic using the system I, we need to explain how you should use the system to search for images by colour. **Below is an outline of the steps and actions required to search for images in the database using system I:**

Step 1: Identify the required colours of the photomosaic section.

Step 2: Use the COLOUR-LAYOUT TOOLBAR to enter your colour requirements, then LEFT MOUSE CLICK SEARCH to find similarly coloured images.

Step 3: Navigate through the results.

Image search results are grouped together based on their similarity. Each group has a representative image. The number overlaying the representative image is the number of images contained within that group. If there is only a single image in the group, no number is displayed.

To view a group of images, DOUBLE CLICK the LEFT MOUSE BUTTON on the representative image group (note this will not work on a single image).

You will notice that once the group images have been displayed, the representative image of the group (the image you **Double Clicked**) is added to the **History Bar**. You can click the image to return back to the group it was in if needed. The **HOME** and **BACK** buttons also enable you to navigate between groups.

Step 4: When you find a desired image, LEFT CLICK, HOLD and DRAG to a desired photomosaic section then DROP. This can be done on single images OR images representing a group (with a number).

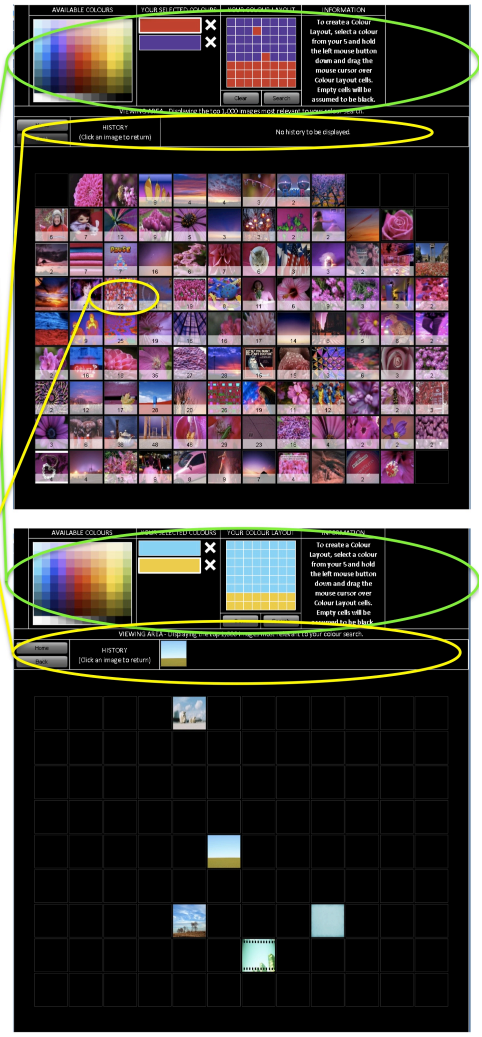


Figure E.4: Clustering-based visualisation query-by-sketch system instructions.

Appendix F

User Study Consent Forms

This appendix contains the consent forms that participants were required to sign prior to taking part in our user studies.

CONSENT FORM
for
participation in the research study entitled:

EVALUATING THE PERFORMANCE OF IMAGE SEARCH BY COLOUR SYSTEMS USING PHOTOMOSAICS

Please read this consent form carefully and ask all the questions you might have before deciding to participate or not in this research study. You are free to ask questions before, during, or after your participation in this research.

Dear participant,

The purpose of this consent form is to seek your free and informed consent to participate in a research study entitled "Evaluating the Performance of Image Search by Colour Systems Using Photomosaics". This research is a study into the comparative effectiveness of different image search methods for finding or retrieving images in an image database based on their colour(s). This study is designed to (a) determine and compare the effectiveness of the different search methods, and (b) determine which of the search methods is preferred by users. The results of this study will influence the creation of a future system for searching images in an image database according to their colour(s). The study is being conducted by William Plant as part of his PhD research with the Aston University, School of Engineering and Applied Science.

If you agree to take part in this study, you will be asked to provide some background information regarding your experience of searching for images in an image database and your artistic background. Due to the nature of the study – i.e., the focus on colour use – **we also ask you to self disclose if you are colour blind**. **Unfortunately you will not be able to participate in this study if you do have any form of colour blindness.**

You will be asked to complete 3 photomosaics using 3 different search by colour systems to find images to complete each of the photomosaics. A photomosaic (or photograph mosaic) is an image that has been created out of a series of lots of other small, identically sized images (or photos). The images making up the mosaic are chosen on the basis of their colours. When a photomosaic is viewed from a distance, the smaller images collectively appear to form the intended larger image. To create photomosaics we need to search for images of specific colours in an image database. Search by colour systems enable you to search a database of images on the basis of their colours. In this experiment, you will have access to a database of 25,000 images. This image database is a freely downloadable collection containing images uploaded by users of the photo-sharing community website Flickr. **The images in the database are uncensored: as a result, we cannot make any assurances as to the content of the images contained within the database. In the unlikely event that you do encounter an image that is, to you, offensive, we would like to assure you that no offence was intended**.

After completing each photomosaic, you will be asked some questions about your opinion of the workload associated with the search by colour system you have used. After you have used all 3 systems, you will be asked to rate your preferences for them. We will record the images you select from the database, the time it takes you to complete each photomosaic, and also your interactions with each search by colour system. **The computer screen will also be video recorded**, with the sole

Figure F.1: First study consent form (*page one*).

intention of monitoring your on-screen actions so we can review your search actions at a later date, and potentially identify any recurring search patterns across participants in the study.

The whole study session will take about 1 hour 30 minutes to complete. Your participation in this study is entirely voluntary and you may end your participation at any time or for any reason without penalty.

If you have any questions about this study, please contact William Plant or his supervisors Dr. Joanna Lumsden or Prof. Ian Nabney as follows:

William Plant
PhD Student
School of Engineering & Applied Science
Aston University
e-Mail: plantwr1@aston.ac.uk

Dr. Joanna Lumsden
Lecturer and Aston Interactive Media (AIM) Lab Manager
School of Engineering & Applied Science
Aston University
e-Mail: j.lumsden@aston.ac.uk

Prof Ian Nabney
Head of Computer Science
School of Engineering & Applied Science
Aston University
e-Mail: i.t.nabney@aston.ac.uk

Risks

The risks associated with participating in this study are minimal.

Confidentiality and Data Storage

All data collected during this study will be kept confidential. Your responses to the study questions will remain anonymous. No one other than members of the research team will have access to the data collected. Individuals will not be identifiable from the data and will not be identified in any publications related to this research. The data will be stored in a secure fashion and will be destroyed 5 years from the date of this study.

Figure F.2: First study consent form (*page two*).

Consent

I, the undersigned, acknowledge that I have been given sufficient time to consider my participation in this study. I confirm that I have received, read, and understood all the information above and give my full and informed consent to participate in the study.

I confirm I have understood the information provided regarding the image database – specifically, that the content of images within the database is uncensored and that the research team take no responsibility for the content of images (and any potential offense caused by images).

I give my permission for my interactions with the computer during the study to be video recorded for future analysis.

I understand that my participation in this study is entirely voluntary and that I am free to end my participation in the study at any time or for any reason without penalty. I also understand that any member of the research team can end my participation in the study for financial, scientific, or ethical reasons at any time.

I understand that by signing this form, I give my full and informed consent to the research team to use the data collected for the purpose of this research and any related research that follows.

Print Name: ____

Signature: ____

Date: __________________

If you would like to receive information about the outcome of this study, please enter details of the most appropriate means by which this information can be communicated to you:

Figure F.3: First study consent form (*page three*).

CONSENT FORM
for
participation in the research study entitled:

COMPARING METHODS FOR PRESENTING SEARCH RESULTS IN IMAGE SEARCH BY COLOUR SYSTEMS

Please read this consent form carefully and ask all the questions you might have before deciding to participate or not in this research study. You are free to ask questions before, during, or after your participation in this research.

Dear participant,

The purpose of this consent form is to seek your free and informed consent to participate in a research study entitled "*Comparing Methods for Presenting Search Results in Image Search By Colour Systems*". This research is a study into the comparative effectiveness of different methods for displaying and presenting images retrieved from an image database based on their colour(s). This study is designed to (a) determine and compare the effectiveness of different search result presentations, and (b) determine which of the search result presentations is preferred by users. The results of this study will influence the creation of future systems for searching images in an image database according to their colour(s). The study is being conducted by William Plant as part of his PhD research with the Aston University, School of Engineering and Applied Science.

If you agree to take part in this study, you will be asked to provide some background information regarding your experience of searching for images in an image database and your artistic background. Due to the nature of the study – i.e., the focus on colour use – **we also ask you to undertake a short colour blindness test**. **Unfortunately you will not be able to participate in this study if you do have any form of colour blindness.**

You will be asked to complete 3 photomosaics using 3 different search by colour systems to find images to complete each of the photomosaics. A photomosaic (or photograph mosaic) is an image that has been created out of a series of lots of other small, identically sized images (or photos). The images making up the mosaic are chosen on the basis of their colours. When a photomosaic is viewed from a distance, the smaller images collectively appear to form the intended larger image. To create photomosaics we need to search for images of specific colours in an image database. Search by colour systems enable you to search a database of images on the basis of their colours. In this experiment, you will have access to a database of 1,000,000 images. This image database is a freely downloadable collection containing images uploaded by users of the photo-sharing community website Flickr. **The images in the database are uncensored: as a result, we cannot make any assurances as to the content of the images contained within the database. In the unlikely event that you do encounter an image that is, to you, offensive, we would like to assure you that no offence was intended**.

After completing each photomosaic, you will be asked some questions about your opinion of the workload associated with the search by colour system you have used. After you have used all 3 systems, you will be asked to rate your preferences for them. We will record the images you select from the database and the time it takes you to complete each photomosaic. **The computer screen will also be video recorded**, with the sole intention of monitoring your on-screen actions so we can

Figure F.4: Second study consent form (*page one*).

review your search actions at a later date, and potentially identify any recurring search patterns across participants in the study.

The whole study session will take about 1 hour 30 minutes to complete. Your participation in this study is entirely voluntary and you may end your participation at any time or for any reason without penalty.

If you have any questions about this study, please contact William Plant or his supervisors Dr. Joanna Lumsden or Prof. Ian Nabney as follows:

William Plant
PhD Student
School of Engineering & Applied Science
Aston University
e-Mail: plantwr1@aston.ac.uk

Dr. Joanna Lumsden
Lecturer and Aston Interactive Media (AIM) Lab Manager
School of Engineering & Applied Science
Aston University
e-Mail: j.lumsden@aston.ac.uk

Prof Ian Nabney
Head of Computer Science
School of Engineering & Applied Science
Aston University
e-Mail: i.t.nabney@aston.ac.uk

Risks

The risks associated with participating in this study are minimal.

Confidentiality and Data Storage

All data collected during this study will be kept confidential. Your responses to the study questions will remain anonymous. No one other than members of the research team will have access to the data collected. Individuals will not be identifiable from the data and will not be identified in any publications related to this research. The data will be stored in a secure fashion and will be destroyed 5 years from the date of this study.

Figure F.5: Second study consent form (*page two*).

Consent

I, the undersigned, acknowledge that I have been given sufficient time to consider my participation in this study. I confirm that I have received, read, and understood all the information above and give my full and informed consent to participate in the study.

I confirm I have understood the information provided regarding the image database – specifically, that the content of images within the database is uncensored and that the research team take no responsibility for the content of images (and any potential offense caused by images).

I confirm I have understood that I am required to complete a short colour blindness test as part of this study, the results of which will be kept completely confidential. I also understand that the results of the colour blindness test do not act as a clinical diagnosis for colour blindness.

I give my permission for my interactions with the computer during the study to be video recorded for future analysis.

I understand that my participation in this study is entirely voluntary and that I am free to end my participation in the study at any time or for any reason without penalty. I also understand that any member of the research team can end my participation in the study for financial, scientific, or ethical reasons at any time.

I understand that by signing this form, I give my full and informed consent to the research team to use the data collected for the purpose of this research and any related research that follows.

Print Name: ______________________________

Signature: ______________________________

Date: ____________

If you would like to receive information about the outcome of this study, please enter details of the most appropriate means by which this information can be communicated to you:

____
____
____
____

Figure F.6: Second study consent form (*page three*).