

Intuitive Image Database Browsing to Support Image Annotation

DOI 10.7305/automatika.53-4.247
UDK 004.65:004.932; 004.832.2:003.6
IFAC 2.8; 1.1.8

Original scientific paper

Image annotation is typically performed manually since automatic image annotation approaches have not matured yet to be used in practice. Consequently, image annotation is a labour intensive and time consuming task. In this paper, we show how an image browsing system can be employed to support efficient and effective (manual) annotation of image databases. In contrast to other approaches, which typically present images in a linear fashion, we employ a visualisation where images are arranged by mutual visual similarity. Since in this arrangement similar images are close to each other, they can easily be selected and annotated together. Organisation on a grid layout prevents image overlap and thus contributes to a clear presentation. Large image databases are handled through a hierarchical data structure where each image in the visualisation can correspond to a cluster of images that can be expanded by the user. Experimental results indicate that annotation can be performed faster on our proposed system.

Key words: Visual information processing, Image databases, Image browsing, Image annotation, Hue sphere.

Intuitivno pretraživanje baze slike kao potpora označavanju slika. Označavanje slika obično se obavlja ručno jer automatski pristupi još nisu dovoljno kvalitetni kako bi se koristili u praksi. Zbog toga je označavanje slika u bazi vremenski zahtjevno. U ovom radu pokazat ćemo kako se sustav za pregled slika u bazi može koristiti kao učinkovita potpora ručnom označavanju slika. Za razliku od drugih pristupa, koji prikazuju slike u linearnom poretku, korištena je vizualizacija u kojoj su slike složene po međusobnoj sličnosti. Budući da su na taj način slične slike međusobno blizu jedna drugoj, lako ih je selektirati i zajednički označiti. Slike su organizirane u mrežni prikaz radi sprječavanja preklapanja i jasnije prezentacije. Velike baze podataka organizirane su u hijerarhijsku strukturu gdje svaka slika u pojedinoj vizualizaciji može pripadati skupu slika čiji prikaz korisnik po želji može proširivati. Rezultati provedenih eksperimenata pokazuju da se označavanje slika pomoću predloženog sustava može obavljati brže nego na uobičajeni način.

Ključne riječi: obrada vizualnih informacija, baze slika, pretraživanje slika, označavanje slika

1 INTRODUCTION

Information is becoming more and more visual. Consequently, methods to effectively deal with this information, in particular with images, are highly sought after. Unfortunately, textual annotations are not available for most images, while annotating images is a challenging task. Manual annotation is work intensive and difficult, while automatic annotation techniques [1] are not sufficiently mature yet to be used in practice. In this paper, we show how an image database visualisation and browsing tool [2] can be used for efficient and effective annotation of image collections. In contrast to other approaches, which typically present images in a linear fashion, we employ a visualisation where images are arranged so that visually similar images are located close to each other. Organisation on a grid layout prevents image overlap and hence contributes to a clear presentation. Large image databases are han-

dled through a hierarchical storage structure where each image in the visualisation can correspond to a cluster of images that can be expanded by the user. Experimental results show that annotation can be performed faster on our proposed system.

The remainder of the paper is organised as follows. Concepts of image database browsing are discussed in Section 2. Section 3 details the Hue Sphere Image Browser that we employ. In Section 4 we show how our image browser can be effectively utilised for image annotation, while this is backed up in Section 5 by experimental results. Finally, Section 6 concludes the paper.

2 IMAGE DATABASE BROWSING

Browsing systems provide an overview of an image database to allow for intuitive navigation through the image collection [3–8]. This is particularly the case when

images are arranged according to mutual similarity [9] so that visually similar images are located close to each other in the browsing interface. One of the main challenges of such image database navigation systems is the limited screen size. Consequently, images are typically displayed as thumbnails arranged in such a way that the user can intuitively understand what kind of images are contained in the dataset.

Visualisation methods for image repositories can be grouped into three main categories: mapping-based, clustering-based, and graph-based approaches [5]. Mapping-based techniques employ dimensionality reduction techniques to map high-dimensional image feature vectors to a low-dimensional space for visualisation. Typical examples use principal component analysis (PCA) [10, 11], multi-dimensional scaling (MDS) [12], or non-linear embedding techniques [13] to define a visualisation space onto which to place images. Clustering-based visualisations group visually similar images together, often in a hierarchical manner. Among the systems that employ such an approach are [14–19]. In graph-based navigation systems, images are the nodes of a graph structure, while the edges of the graph show relationships between images (e.g., joint keywords or visual similarity). Graph-based systems include [20–23].

Once an image set is visualised, effective tools are needed for the user to be able to interact with the dataset in an efficient and intuitive way [6]. We can distinguish between horizontal browsing which works on images of the same visualisation plane, and includes operations such as panning, zooming, magnification and scaling, and vertical browsing which allows navigation to a different level of a hierarchically organised (often clustering-based) visualisation.

3 HUE SPHERE IMAGE BROWSER

The visualisation approach we employ in this paper is based on the Hue Sphere Image Browser presented in [19, 24, 25] and further developed into an on-line application [26] to provide an accessible, efficient and effective approach to image database browsing. The reason for adopting a spherical visualisation space is that users will be familiar with the concept of a globe and how to navigate on one.

Each image is described by its median colour. However, rather than employing the standard RGB colour space, the HSV space is used [27] which humans find more intuitive. Of this, only the hue H and value (brightness) V attributes are taken, since the saturation descriptor is deemed less important for characterising image content. H and V are then transformed to longitude and latitude co-ordinates which describe the location of the image thumbnail on the sphere.

The sphere can be rotated to focus on images with different colours or tilted to look at brighter or darker images.

Since the co-ordinates are extracted directly from the images, this approach is very fast and much more efficient than methods such as multi-dimensional scaling which rely on a comparably slow and computationally expensive iterative convergence procedure. At the same time, it does not compromise the achieved image database visualisation, since, as has been shown [28], the average image colour is at least as effective a descriptor for image browsing as high-dimensional feature vectors (e.g., based on colour distributions). In addition, the features employed are intrinsically suitable for query-based image retrieval as well [29], which allows for a combination of browsing and querying.

As overlapping and occlusion of image thumbnails will have a negative impact on the browsing experience [9], images are placed on a regular grid structure so that they do not overlap each other. Clearly, and in particular for larger image sets, this will mean that more than one image can be mapped to the same cell. This is handled through a hierarchical approach of storing and accessing the image collection and is performed based directly on the grid mapping, leading to a hierarchical tree based on clustering images. However, in contrast to clustering-based visualisation methods such as [15] there is no need for computationally expensive clustering algorithms. Rather, the resolution of the grid layout represents the maximal number of clusters present at a given level (which will only be met if all cells are filled). The grid cells then directly determine which images fall into which clusters, and assigning images to clusters/grid cells can hence be performed with simple comparison operations only. A representative image for each cluster is chosen by identifying the image closest to the centre of the cell.

This procedure is adopted at the each level of the tree hierarchy, i.e. first at the root node (the initial global display) and then for each non-empty cell again in a recursive manner, where the images of each child node are again mapped

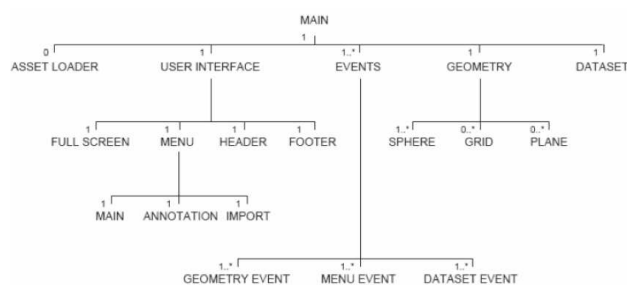


Fig. 1. High-level class diagram of the browsing application

```

<node>
  <image>
    <origURL>images/01209.tif</origURL>
    <imageURL>assets/images/ucid/01209.jpg</imageURL>
    <thumbURL>assets/thumbs/ucid/01209Thumb.jpg</thumbURL>
    <avgH>0.01788</avgH>
    <avgV>0.1725</avgV>
    <desc/>
    <gridX>2</gridX>
    <gridY>4</gridY>
  </image>
</node>
<node>
  <image>
    <origURL>images/00918.tif</origURL>
    <imageURL>assets/images/ucid/00918.jpg</imageURL>
    <thumbURL>assets/thumbs/ucid/00918Thumb.jpg</thumbURL>
    <avgH>0.0238</avgH>
    <avgV>0.1647</avgV>
    <desc/>
    <gridX>3</gridX>
    <gridY>4</gridY>
  </image>
</node>

```

Fig. 2. Extract of the XML data structure.

to a grid structure, until the complete tree is derived. The resulting interface provides an intuitive way of browsing so that the user can, starting from the initial display, successively select a representative image to refine the search. That image cluster is then expanded in the next level of the display where the user can again select an image group to navigate further into the image collection.

The grid-tree structure also provides another advantage. As the structure is fixed it can be pre-computed in completeness off-line, together with all possible grid view configurations the user can encounter. This in turn provides the user with the possibility of real-time browsing large image collections.

In order to provide an effective tool for today's networked society, we have implemented an on-line version in Adobe Flash[®] CS 4 Professional [26] that can be used on any machine with an Internet browser and Adobe Flash Player.

ActionScript 3¹, the development language for Flash applications, features a core event driven architecture, an extensive API for XML and the ability to import/export all data types to/from the local file system. Object-oriented principles were followed in the design and implementation of the application with a high-level class diagram shown in Fig. 1. Image data is stored in XML² format, making it easily readable using appropriate tools and also providing the possibility of generating database structures using external software of choice. Figure 2 shows a sample extract of the XML data structure.

¹<http://livedocs.adobe.com/flash/9.0/>

ActionScriptLangRefV3/

²<http://www.w3.org/TR/2006/REC-xml-20060816/>



Fig. 3. Navigation controls to move to neighbouring cells/clusters of the image database.

Crucially, Flash CS4 and Flash Player 10 support, alongside several other key changes, a built-in API for 2.5D. Whilst unable to manage true 3D geometry, 2D objects can be transformed to be embedded within a 3D space (hence 2.5D). The Flash 10 architecture involves each display object within the scene being distorted through matrix transformations dependent on its position and rotation in relation to the (adjustable) perspective point. This is significantly different from other approaches such as PaperVision 3D³, and combined with the fact that it is taking place natively thus taking full advantage of hardware acceleration, results in far more responsive handling of large display lists.

In addition to the original hue sphere which provides an interaction scenario similar to a globe [19, 24, 25], we also implemented a second visualisation space whereby

³<http://www.papervision3d.org/>

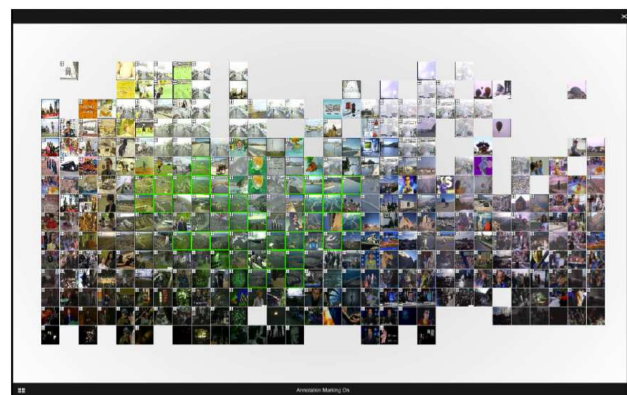


Fig. 4. Selecting multiple images and image clusters for annotation (selected images have a green border).

Table 1. Experimental results of annotation task.

	images	Windows Explorer		Browsing application	
		time taken	interactions	time taken	interactions
US flag	3	2 : 12	93	0 : 12	6
stained glass window	18	5 : 13	258	3 : 18	41
telephone box	4	2 : 42	121	1 : 01	25
average	8	3 : 22	157	1 : 30	24

the user is placed at the centre of the sphere looking outwards thus being immersed in the visualisation itself. Furthermore, a flat visualisation space is also provided where images are projected onto a plane. Examples of these three paradigms, applied to the MPEG-7 common colour dataset [30], are given in Fig. 6.

Navigation on the sphere includes the ability to rotate and tilt the sphere at the same time. Rotation around the y-axis is applied to the holder object, however tilting around the x-axis is applied to the sphere object therefore maintaining the direction of the applied tilt whilst rotating. Both rotation and tilting operations are smoothed via the introduction of an easing formula, recalculating and decelerating the rotational values every frame until they equal the newly assigned value. A limitation is applied to the tilting action to prevent the user viewing the sphere upside down.

The hierarchical data structure coupled with the corresponding organisation of images on a grid at various levels, allows for efficient and effective access to large databases. By selecting one image, the user can expand its image cluster, and in return the application will display the corresponding part of the next level of the hierarchy. While browsing through lower levels of the grid structure, the user can also immediately move to neighbouring image clusters as is illustrated in Fig. 3 for the case at the bottom level of the hierarchy (i.e., where single images are displayed). Apart from the image itself, four neighbouring images are shown corresponding to the immediate neighbours in the tree structure. Where no immediate neighbour exists (i.e., when there are no images in the corresponding grid cell), the application will search adjacent cells and

display an appropriate image for navigation.

4 IMAGE BROWSING FOR ANNOTATION

Since automatic annotation systems [1] still have a long way to go before becoming sufficiently accurate and hence useful in practice, we see image browsing tools as an attractive means for efficiently annotating image datasets. The reason for this is that, as mentioned, similar images are clustered together, or at least close to each other in the visualisation, so that annotating these images simultaneously will take fewer interactions compared to going through an image list one-by-one. Consequently, we provide annotation facilities in our application that support exactly this idea. Users can select multiple images and then attach keyword information to the selected set. Fully utilising the hierarchical data structure, by selecting an image representing an image cluster, all images in that cluster are selected at the same time, allowing for an efficient mechanism of tagging whole image sets. Figure 4 shows an example of multiple images and image clusters selected for annotation, while Fig. 5 gives an example of the actual annotation dialog.

5 EXPERIMENTAL EVALUATION

Although the usefulness of the employed image browsing tools is evident, objective evaluation of such systems has been shown to be difficult [31]. Nevertheless, we performed some user-based tests where we compared the effectiveness of our image database navigation system with a common browsing tool, namely MS Windows Explorer®. The task given to the users was to annotate a set of images. In particular, users had to find and annotate all images of a certain category (with each category consisting of between 3 and 18 images). The results of this task are listed in Table 1. From there it is clear that our similarity-based browsing application is proven to provide a faster interface for image annotation. The flags category was annotated very quickly as all 3 images are very similar and therefore very close together. However the stained glass windows took somewhat longer using the browsing application due to larger fluctuations within hue and value, spreading them over the visualisation space. It should however be noted, that the relatively fast times for Windows Explorer here are

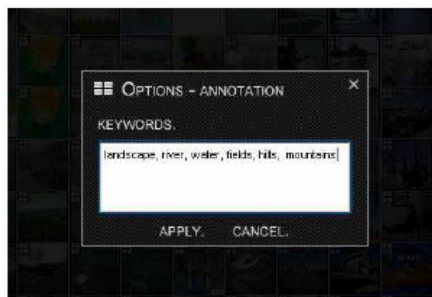


Fig. 5. Sample annotation dialog.

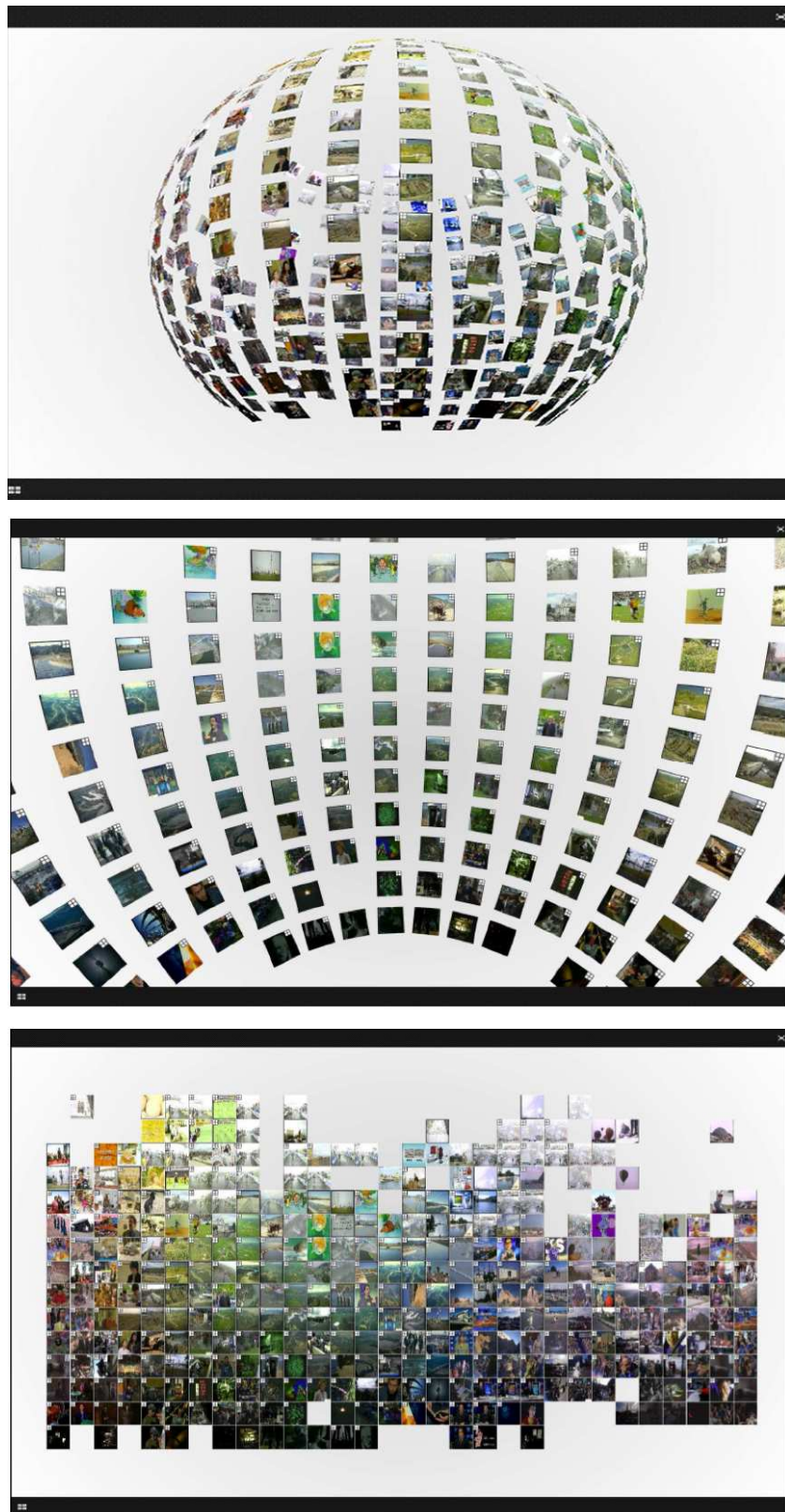


Fig. 6. Visualisation of MPEG-7 CCD database [30] projected onto hue sphere (top), with user immersed at the centre of the sphere (middle), and displayed on a flat plane (bottom).

based on the fact that the relevant images had similar file-names (or rather image numbers) and hence the reported results probably represent a best case scenario for linear image displays.

6 CONCLUSIONS

In this paper we have shown that image database browsing systems can be used to support effective and efficient annotation of image repositories. Images are projected onto either a spherical or a planar visualisation space in a similarity-based fashion where visually similar images are located close to each other in the browsing interface and hence can be easily selected together for joint annotation. Large datasets are handled through a hierarchical data structure, while images are arranged on a grid to avoid overlapping and hence user confusion. User-based experiments have shown that the browsing application is more efficient for annotation than a system based on a conventional linear display of images.

REFERENCES

- [1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, 2008.
- [2] G. Schaefer and M. Stuttard, "Image browsing for efficient image annotation," in *52nd Int. Symposium ELMAR*, pp. 123–125, 2010.
- [3] D. Heesch, "A survey of browsing models for content based image retrieval," *Multimedia Tools and Applications*, vol. 40, no. 2, pp. 261–284, 2008.
- [4] W. Plant and G. Schaefer, "Visualisation and browsing of image databases," in *Multimedia Analysis, Processing and Communications*, vol. 346 of *Studies in Computational Intelligence*, pp. 3–57, Springer, 2011.
- [5] W. Plant and G. Schaefer, "Visualising image databases," in *IEEE Int. Workshop on Multimedia Signal Processing*, pp. 1–6, 2009.
- [6] W. Plant and G. Schaefer, "Navigation and browsing of image databases," in *Int. Conference on Soft Computing and Pattern Recognition*, pp. 750–755, 2009.
- [7] G. Schaefer, "Content-based retrieval from image databases: colour, compression, and browsing," in *Int. Conference on Information Retrieval and Knowledge Management*, pp. 5–10, 2010.
- [8] G. Schaefer, "Interactive exploration of image collections," in *7th Int. Conference on Computer Recognition Systems*, vol. 95 of *Advances in Intelligent and Soft Computing*, pp. 229–238, 2011.
- [9] 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 Visualization*, pp. 36–43, 1999.
- [10] 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*, pp. 102–108, 2001.
- [11] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. Huang, "Visualization and user-modeling for browsing personal photo libraries," *Int. Journal of Computer Vision*, vol. 56, pp. 109–130, January 2004.
- [12] Y. Rubner, L. Guibas, and C. Tomasi, "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval," in *Image Understanding Workshop*, pp. 661–668, 1997.
- [13] G. Nguyen and M. Worring, "Interactive access to large image collections using similarity-based visualization," *Journal of Visual Languages and Computing*, vol. 19, no. 2, pp. 203–224, 2008.
- [14] S. Krischnamachari and M. Abdel-Mottaleb, "Image browsing using hierarchical clustering," in *IEEE Symposium on Computers and Communications*, pp. 301–307, 1999.
- [15] J. Chen, C. Bouman, and J. Dalton, "Hierarchical browsing and search of large image databases," *IEEE Trans. Image Processing*, vol. 9, pp. 442–455, March 2000.
- [16] I. Bartolini, P. Ciaccia, and M. Patella, "Adaptively browsing image databases with PIBE," *Multimedia Tools and Applications*, vol. 31, no. 3, pp. 269–286, 2006.
- [17] A. Gomi, R. Miyazaki, T. Itoh, and J. Li, "CAT: A hierarchical image browser using a rectangle packing technique," in *12th Int. Conference on Information Visualization*, pp. 82–87, 2008.
- [18] G. Schaefer and S. Ruzsala, "Image database navigation on a hierarchical MDS grid," in *Int. Conference German Pattern Recognition Society*, vol. 4174 of *Springer Lecture Notes on Computer Science*, pp. 304–313, 2006.
- [19] G. Schaefer, "A next generation browsing environment for large image repositories," *Multimedia Tools and Applications*, vol. 47, no. 1, pp. 105–120, 2010.
- [20] C. Chen, G. Gagaudakis, and P. Rosin, "Similarity-based image browsing," in *Int. Conference on Intelligent Information Processing*, pp. 206–213, 2000.
- [21] D. Heesch and S. Rürger, "NNk networks for content-based image retrieval," in *European Conference on Information Retrieval*, pp. 253–266, 2004.
- [22] M. Dontcheva, M. Agrawala, and M. Cohen, "Metadata visualization for image browsing," in *18th Annual ACM Symposium on User Interface Software and Technology*, 2005.
- [23] M. Worring, O. de Rooij, and T. van Rijn, "Browsing visual collections using graphs," in *Int. Workshop on Workshop on Multimedia Information Retrieval*, pp. 307–312, 2007.
- [24] G. Schaefer and S. Ruzsala, "Image database navigation: A globe-al approach," in *Int. Symposium on Visual Computing*, vol. 3804 of *Springer Lecture Notes on Computer Science*, pp. 279–286, 2005.

- [25] G. Schaefer and S. Ruzsala, "Hierarchical image database navigation on a hue sphere," in *Int. Symposium on Visual Computing*, vol. 4292 of *Springer Lecture Notes on Computer Science*, pp. 814–823, 2006.
- [26] G. Schaefer and M. Stuttard, "An on-line tool for browsing large image repositories," in *Int. Conference on Information Retrieval and Knowledge Management*, pp. 102–106, 2010.
- [27] J. Sangwine and R. Horne, *The Colour Image Processing Handbook*. Chapman & Hall, 1998.
- [28] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "A comparison of measures for visualising image similarity," in *The Challenge of Image Retrieval*, 2000.
- [29] M. Stricker and M. Orengo, "Similarity of color images," in *Conf. on Storage and Retrieval for Image and Video Databases III*, vol. 2420 of *Proceedings of SPIE*, pp. 381–392, 1995.
- [30] Moving Picture Experts Group, "Description of core experiments for MPEG-7 color/texture descriptors," Tech. Rep. ISO/IEC JTC1/SC29/WG11/ N2929, 1999.
- [31] W. Plant and G. Schaefer, "Evaluation and benchmarking of image database navigation tools," in *Int. Conference on Image Processing, Computer Vision, and Pattern Recognition*, vol. 1, pp. 248–254, 2009.



Gerald Schaefer gained his PhD in Computer Vision from the University of East Anglia. He worked at the Colour and Imaging Institute, University of Derby (1997-1999), in the School of Information Systems, University of East Anglia (2000-2001), in the School of Computing and Informatics at Nottingham Trent University (2001-2006), and in the School of Engineering and Applied Science at Aston University (2006-2009) before joining the Department of Computer Science at Loughborough University where he leads

the Vision, Imaging and Autonomous Systems Research Division.

His research interests are mainly in the areas of colour image analysis, image retrieval, physics-based vision, medical imaging, and computational intelligence. He has published extensively in these areas with a total publication count exceeding 250. He is a member of the editorial board of more than a dozen international journals, reviews for over 80 journals and served on the programme committee of more than 200 conferences. He has been invited as keynote or tutorial speaker to more than 30 conferences, is the organiser of various international workshops and special sessions at conferences, and the editor of several books, conference proceedings and special journal issues.



Matthew Stuttard obtained his BSc in Multimedia Computing from Aston University. He is Director and Lead Developer at MSFX Ltd and AppFx Ltd, and also a sessional Lecturer at the School of Engineering and Applied Science at Aston University.

AUTHORS' ADDRESSES

Gerald Schaefer
Department of Computer Science
Loughborough University
Loughborough, U.K.
email: gerald.schaefer@ieee.org

Matthew Stuttard
MSFX Ltd
U.K.
email: matt@msfx.co.uk

Received: 2012-04-09

Accepted: 2012-09-08