Translating Journalists' Requirements into Features for Image Search

Julian Stöttinger, Jana Banova, Thomas Pönitz
Institute of Computer-Aided Automation, PRIP Group
Vienna University of Technology
Favoritenstraße 9/1832
A-1040 Vienna, Austria
julian@prip.tuwien.ac.at

Nicu Sebe
Facoltà di Scienze Cognitive
University of Trento
Via Belenzani, 12
38122 Trento
Italy

Allan Hanbury
IR Facility
Palais Eschenbach
Eschenbachgasse 11
1010 Vienna
Austria

***Abstract*—This paper illustrates how taking advantage of user studies highlighting the user requirements can lead to the selection of suitable visual features in image search systems. The results of a study to identify pertinent visual features to enhance a text-based press photo search system used by journalists are presented. A requirement was that the visual features should be intuitively understandable by the journalists. This feature selection task is approached by first determining the journalists' photo searching requirements based on a published user study. These requirements are then mapped to suitable visual features. The emphasis was on identifying suitable and intuitive low-level features, as these can be rapidly implemented in the existing text-based image search system. Results demonstrating the use of the selected features are shown.**

I. Introduction

Journalists have access to huge databases of photographs for illustrating their articles. They would like online searches to give fast access to the most suitable photos satisfying a requirement. At present, the majority of search interfaces to these databases rely on pure text search, making use of meta-data entered by photographers when they upload their photos, or by archivists.

This paper describes the results of a study carried out for the owner of a large database of press photographs in Austria. They were interested in adding the possibility of doing visual search to their current text retrieval system, and wished to find out which visual features would be the most suitable. An important pre-requisite was that the end users should be able to tune the image similarity criteria based on their image needs. This implies that the features to be used for a specific search should be selectable, but also that the features should be intuitively understandable for the end users. The features should also enable the users to improve the returned results.

The main contribution of this paper is to illustrate how taking advantage of user studies highlighting the user requirements can lead to the selection of suitable features in image search systems. We approached this task by first carrying out an analysis of journalists' photo searching requirements by further analysing the results of a published user study (Section II). These requirements were then mapped to suitable visual features (Section III). The emphasis was on identifying suitable and intuitive low-level features, as these can be rapidly implemented in the existing text-based image search system. The identified features are described in Section IV. In contrast to selecting a set of generic image features, such an approach to feature selection based on user requirement analysis should lead to better acceptance and use of these features by the end users.

II. Analysis of Journalists' Requirements

We based the selection of suitable features on a study by Markkula and Sormunen [1], [2] of journalists' photo needs at a large newspaper in Finland. In this study, an observer interactively followed the work of eight journalists and interviewed them. Amongst other aspects considered, the search topics and the selection criteria used for photos were analysed. As the results of the user study are summarised in narrative form, we begin by extracting the search requirements obtained in this analysis in a more succinct form. In the next section, visual features suitable for meeting some of the requirements are discussed.

The analysis in [1], [2] showed that the following categories of search topics were important[1]:

- *Concrete objects and people:* Most frequently animals, vehicles and people.
- *Documentary photos of events:* Mostly recent news events.
- *Photos of places:* such as "Rauma old town".
- *Themes or Topics:* such as "holidays in the south", "child care at home".
- *Subjectively interpretable aspects:* such as "young love" or "a child's anxiety"

It is also mentioned that more than half of the requests were refined, with the following being the most common criteria:

- *Colour/B&W:* old black and white photos are seldom wanted.
- *Shooting distance:* close-ups were often desired.
- *Time expressions:* such as shooting year or season.
- *Attributes of objects:* such as "judge with a wig".
- *Action taking place:* such as "champagne cork popping".

[1]In [2], distinction is made between requests typed into an online search system and requests sent to archivists. We have ignored this distinction here.

Having obtained the retrieval set, the journalists applied the following common selection criteria:

- *Topicality:* how well the photo fits the story.
- *Technical and biographical criteria:* the photo should be technically good. The photo source, publishing information and publishing history are also considered.
- *Impression to be conveyed:* e.g. thought-provoking, dramatic, surprising, shocking, funny. Non-typical photos were also often sought.
- *Passport photos/formal portraits:* a photograph of a single face.
- *Aesthetic criteria:* colour and composition played the most important role in the final selection phase.

According to [1], the *aesthetic criteria* play an important role in the final stage of selecting photos, at which time a group of candidate photos has been chosen. In particular, photos already chosen for a page restrict the possibility of using other similar photos on the same page. In order to balance the photos on the page, photos of different types and with different visual features should be used.

III. Mapping Requirements to Features

In [2], the authors state pessimistically that "the results of this study were not very favourable for applying content-based indexing and retrieval methods in newspaper type of photo archives". However, recent research on high level visual processing has the potential to contribute towards meeting some of the requirements. In particular, the object categorisation research evaluated by the PASCAL VOC [3] can contribute towards the *concrete objects and people* requirements (many of the object classes in the challenge correspond to the examples listed in the requirement). The *Photos of places* requirement should be met to a large extent by the recent expansion of the automated geo-tagging of photos, although work on the visual recognition of landmark buildings [4] could also contribute. Work on the automated judgement of the technical quality of photos [5], [6] could be used to rank potentially technically good photos higher. Finding *passport photos* or *formal portraits* can be done using the Viola-Jones object detector [7]. By measuring the size and number of detected faces, portrait and group photos can be found. Exploratory work on automated emotion classification in images [8] has the potential to contribute towards the *Impression* and *Subjective interpretation* requirements. Finally, the ability to find images that are visually almost identical has an application for the *biographical criteria* requirement — a journalist could attempt to find an almost identical image that has a less stringent copyright or lower usage fee.

In [2] it is also stated that "low-level visual features were not expressed as the main search criteria in any of the search topics analysed". Nevertheless, it is possible to identify three requirements to which low-level features can contribute. *Colour/B&W* is an obvious low-level feature. *Shooting distance* is related to the *depth of field* (DOF), as close-up photos are often taken with a low depth of field, as illustrated in Figure 1. This can, for recent photos, usually be obtained from the F-number field of the EXIF data provided with the photo. For older photographs, and for those without EXIF information, the depth of field can be estimated by image analysis. Finally, many of the traditional CBIR features based on colour and texture can be used to describe *aesthetic criteria*. For example, sorting photos by visual similarity in order to find photos which are visually different enough. The advantage of using low-level features related to these criteria is that journalists have an intuitive understanding of them. In addition, they can be easily combined to find images satisfying a set of requirements. A set of low-level features that could contribute to satisfying the requirements are described in the following section.

The remaining requirements must at present be met through the use of image meta-data, such as restricting the time frame for documentary photos of events. Others can at present be met based only on text search on meta-data, or through judgement of the results by the users.

IV. Features

The low-level features satisfying the requirements specified in the previous section are described.

Two colour spaces are used: the RGB space, and a cylindrical coordinate colour representation in terms of hue, saturation and lightness. Although there are many such cylindrical coordinate spaces available (HSV, HLS, HSI, etc.), it can be shown [9] that a "unified" set of cylindrical colour coordinates suitable for image analysis can be derived. The coordinates are calculated as:

$$i = \frac{1}{3}(R+G+B) \quad (1)$$

$$s = \max(R,G,B) - \min(R,G,B) \quad (2)$$

$$h = \arctan\left(\frac{\sqrt{3}(G-B)}{2R-G-B}\right) \quad (3)$$

A. Colour/B&W

This is expressed in a single feature, the average grey-value of the saturation channel (s) of an image represented in the cylindrical coordinate colour space above.

B. Shooting Distance

For close-up and macro shots, photographers often reduce the depth of field (DOF) to blur the background. They thereby reduce the complexity of the image and make the subject of the photograph stand out. Hence close-up images are often characterised by having a blurred region surrounding the main object in the image. This is demonstrated in Figure 1, showing a low DOF image compared to an image with a higher DOF. For the DOF-indicator, we use the feature suggested by Datta et al. [5]. A three-level Daubechies wavelet transform [10] is computed for each colour channel in the cylindrical coordinate representation. The image is then divided into 16 equally sized rectangles arranged as a 4×4 grid. Finally, the ratio of the sum of the high frequency wavelet coefficients (level 3) of the four inner rectangles with respect to the

(a) (b)

Figure 1. (a) Low and (b) High DOF image.

whole image is calculated. This is done for each channel (h, s, i) separately, producing a vector of three features. As an alternative, one could consider using an algorithm to detect the regions of an image that are in focus, as presented for example in [11].

C. Colour and Composition

Many features traditionally used for content-based image retrieval fall into this category.

1) Colour: Colour histograms were the first features used in content-based image retrieval, however as shown by Deselaers et al. [12], they perform well for many tasks in image retrieval. In order to allow a more semantically based approach to colour similarity, we introduced a *colour names histogram*, based on work by Van de Weijer et al. [13].

In English, eleven basic colour terms have been defined based on a linguistic study [14]. These are: black, blue, brown, green, grey, orange, pink, purple, red, white, yellow. Van de Weijer et al. [13] used images downloaded from the Internet to learn the mapping of these colour names to RGB coordinates, thereby creating a lookup table of each RGB coordinate to one of eleven colour names[2]. Examples of the resulting mapping to eleven colours are shown in Figure 2. From these reduced colour images, eleven-bin colour histograms are created.

Differences in image acquisition conditions, such as lighting changes, can have a noticeable effect on the resulting histograms. This is particularly noticeable for faces (see Figure 2).

In addition, single bins of the colour names histogram are used as single features. This means that one can locate images with a similar percentage of each of the basic colours individually or in selected combinations, allowing the layout to be planned based on specific colour combinations. As an example, searching a database of Austrian press photographs for those containing similar amounts of red and black to the query photograph (leftmost image in Figure 3) produced the results shown in Figure 3 as well as the image in Figure 2(c). Such a search puts more emphasis on the important colours than a search based on a full colour histogram.

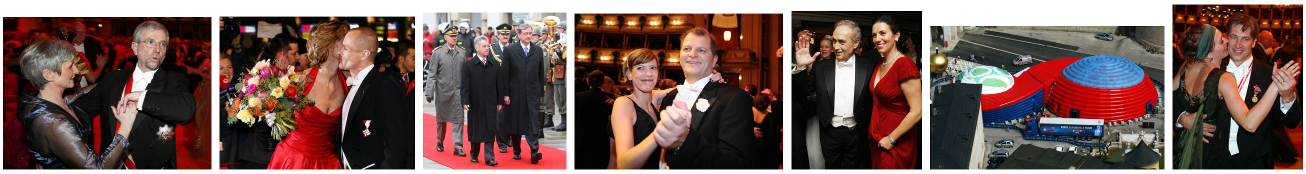
2) Composition: There are many constituents contributing to the composition of a photo. Here we consider only the spatial organisation constituent, and characterise it using low-level image segmentation, based on work in [5]. A simple composition feature is the complexity or *level of detail* of an image. This can be characterised by the number of regions resulting from a segmentation, as this number depends on the spatial complexity of the image. This is illustrated in Figure 4, where the less complex image on the left produces fewer regions. However, instead of the segmentation by clustering in CIELUV space used in the [5], we made use of a waterfall segmentation. The advantage of the waterfall segmentation is that it takes spatial information as well as colour information into account, resulting in regions that are more contiguous.

The Waterfall [15] is a Watershed performed considering only the low pass points separating regions. In order not to take into account the value of other pixels, each region is filled with the value of the smallest pass point of its frontier (see Figure 5). A morphological reconstruction [16] may be used for this purpose. This operation establishes a hierarchy among the frontiers produced by the Watershed. The process may be iterated until a single region covers the whole image. In the example of Figure 5 the Watershed lines are indicated by arrows and only solid line arrows will be preserved by the Waterfall.

[2]This lookup table is available here: http://lear.inrialpes.fr/people/vandeweijer/color_names.html

(a) (b) (c) (d)

Figure 2. Examples showing the assignment of eleven basic colours. (a) and (c) original images. (b) and (d) basic colours assigned.



(a) (b) (c) (d) (e) (f) (g)

Figure 3. (a) Query image, (b)–(g) Results based on a similar amount of *red* and *black*.

(a) (b) (c) (d)

Figure 4. (a) and (c) Original images. (b) Waterfall segmentation of (a) produces 25 regions. (d) Waterfall segmentation of (c) produces 189 regions.

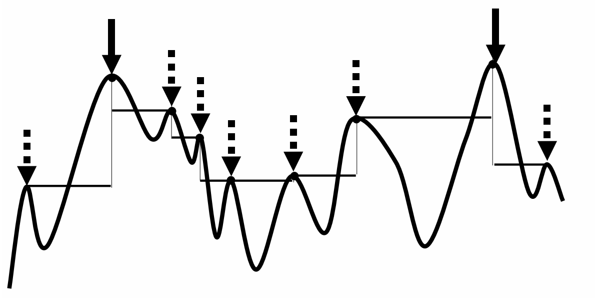


Figure 5. Waterfall principle.

An efficient graph-based Waterfall algorithm is presented in [17]. The Waterfall segmentation at level 2 for two images of differing levels of detail are shown in Figure 4 (after applying a leveling of size 3).

The waterfall algorithm is carried out on the gradient of the colour image. We use the gradient found to give the best results in a morphological waterfall segmentation in [18]. This is the *saturation weighing-based colour gradient* applied in the cylindrical coordinate colour space. This gradient gives a larger weight to the differences in hue when the saturation is high, and a larger weight to differences in luminance when the saturation is low. In order to simplify the image before segmenting it, thereby eliminating small regions, we make use of the morphological *leveling* [19]. The filter used to produce the marker for the leveling operator is the morphological alternating sequential filter [16], where the size of the filter refers to the number of subsequent opening and closing operations. In order to apply these filters to colour images, we apply the filter separately to each colour component. We pre-process with a leveling of size 3 and use level 2 of the waterfall hierarchy, as these parameters result in large contiguous regions. The level of detail feature is not very sensitive to the values of these parameters, as long as the same parameters are used for every image (and a hierarchy level that does not produce a single region covering the whole image is used).

Further segmentation composition features that remain to be investigated include the colours, shapes and positions of the five largest regions obtained by the segmentation, as proposed in [5]. Through the combination of the image segmentations with the colour names histogram, it should be possible to obtain a simple representation of the colour distribution that can be queried both by visual example and text description.

V. Conclusion

This paper demonstrates how the results of a user study can be translated into visual features that are designed to: (i) meet the requirements of the users, and (ii) be intuitively understandable for the users. Based on a published study on the image search requirements of journalists, a collection of both high-level and low-level visual features were identified. At this stage of the project, the selected features have been presented to a group of end users, who indicated that they were satisfied with the search capabilities made possible by the features. The low-level features are currently being implemented as supplemental features in the (currently text only) search system used by the journalists. Given the possibility to choose single visual features or combinations, it is important to design a suitable user interface. A study on the effectiveness of the features and their acceptance and use by the journalists is planned.

Acknowledgment

This work was partly supported by Austrian Press Agency Informations Technologie GmbH[3], the Austrian Research Promotion Agency (FFG)[4] project 815994, and the CogVis[5] Ltd. However, this paper reflects only the authors' views; the APA-IT, FFG or CogVis Ltd. are not liable for any use that may be made of the information contained herein.

References

[1] M. Markkula and E. Sormunen, "Searching for photos - journalistic practices in pictorial IR," in *The Challenge of Image Retrieval*, 1998.

[2] ——, "End-user searching challenges indexing practices in the digital newspaper photo archive," *Information Retrieval*, vol. 1, no. 4, pp. 259–285, 2000.

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[4] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Study aesthetics in photographic images using a computational approach," in *Proceedings of the European Conference On Computer Vision (ECCV)*, 2006, pp. 288–301.

[6] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 419–426.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[8] W. Wang and Q. He, "A survey on emotional semantic image retrieval," in *Proc. IEEE Conf. Image Processing*, 2008, pp. 117–120.

[9] A. Hanbury, "Constructing cylindrical coordinate colour spaces," *Pattern Recognition Letters*, vol. 29, no. 4, pp. 494–500, 2008.

[10] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[11] L. Kovács and T. Szirányi, "Focus area extraction by blind deconvolution for defining regions of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1080–1085, 2007.

[12] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Information Retrieval*, vol. 11, pp. 77–107, 2008.

[13] J. van de Weijer, C. Schmid, and J. Verbeek, "Learning color names from real-world images," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2007.

[14] B. Berlin and P. Kay, *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.

[15] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical Morphology and its Applications to Image Processing, Proc. ISMM'94*, 1994, pp. 69–76.

[16] P. Soille, *Morphological Image Analysis*, 2nd ed. Springer, 2002.

[17] B. Marcotegui and S. Beucher, "Fast implementation of waterfall based on graphs," in *Mathematical Morphology and its Applications to Image Processing, Proc. ISMM'05*, 2005, pp. 177–186.

[18] J. Angulo and J. Serra, "Color segmentation by ordered mergings," in *Proc. of the Int. Conf. on Image Processing*, vol. II, 2003, pp. 125–128.

[19] F. Meyer, "Levelings, image simplification filters for segmentation," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 59–72, 2004.

[3] http://www.apa-it.at/
[4] http://www.ffg.at/
[5] http://www.cogvis.at/