Evaluation Method for Content-based Photo Retrieval

Working group report by Annelise Mark Pejtersen, Marjo Markkula, Eero Sormunen, Marius Tico and Arjen P. De Vries

The working group meeting was held in Tampere, Finland 10th and 11th August 1998. Participants*:* *Annelise Mark Pejtersen* (Risø National Laboratory, Denmark), *Marius Tico* (Tampere University of Technology, Finland), *Arjen de Vries* (University of Twente, Netherlands), *Kalervo Järvelin, Marjo Markkula, Eero Sormunen* (University of Tampere, Finland)

1 Introduction - Eero Sormunen

Past research in library and information science has seen photos as objects of conceptual indexing and concentrated on yielding theoretical frameworks for that work. Various aspects of photos that could or should be indexed have been characterised and the problems of concept-based indexing have been identified. The connection between the theoretical frameworks and professional practices has been a loose one. The developers of operational photo retrieval systems have applied traditional text retrieval techniques and domain specific indexing methods (free-text or thesaurus based).

	Academic research	**Professional Practices**
Library and information science	Theories for conceptual indexing	Collection management, indexing practices
Computer science	Algorithms for CBR	Text-based retrieval systems

Table 1. Academic Research and Professional Practices in the field of image retrieval

Past research in computer science and especially in digital image processing has seen photo retrieval as a visual matching problem. The goal has been to develop efficient algorithms for querying photos by a query image or by other visually oriented criteria (content-based retrieval - CBR methods). Although CBR has been a intensively studied area, only few commercially marketed retrieval systems have become available. Operational applications in general photo retrieval based on CBR have not been reported in the literature. However, some demo systems have become operational through the Web (e.g., Image Surfer by Excalibur in Yahoo!).

Enser concluded his literature survey by saying that above described research communities do not communicate with each other. On the other hand, neither of these research branchesseem to interact with the mainstream of the IR research (not to even mention missing links to interactive, user oriented or cognitive lines of IR research). After realising these gaps in communication it is no wonder that evaluation methods for content-based photo retrieval are not well established (as will be seen later in this report).

This report summarises the results of the MIRA working group meeting called together to generate new ideas for evaluation methods appropriate in digital photo retrieval environments. The aim was to compose an outline for a new evaluation method intended to give a more solid ground for the development of CBR methods. The first step was to look at the use of photo retrieval systems in one real work situation: illustration tasks in the newsroom (summarised in appendix 1). The second step was to discuss the state of technology in CBR (appendix 2) and the presently used evaluation methods (appendix 3). Thirdly, potential evaluation approaches were discussed. The outlined method based on the comparison of the computed similarity measures with the human perceived similarities was further developed and preliminary tested by Sormunen and Markkula.

2 Evaluation Method for Content-based Photo Retrieval

In the present digital photo collections, only text based queries are used. The lack of appropriate content-based algorithms is not the only reason for the dominance of text based retrieval. The fact is that most of the identified needs of users (at least in newspaper environment) cannot be served effectively by the CBR methods in general type of photo collections. The focus of query attributes is outside the seen image (e.g., news events, general themes) or the objects of interest are too complex for the present CBR methods (e.g., proper name queries, object types).

A promising development area for the CBR algorithms could, in the first place, be browsing rather than focused querying. The CBR algorithms could be seen as browsing tools applied in relatively large sets of thumbnail images resulting from broad text-based queries. The major consequence of seeing CBR as a browsing tool is that the algorithms are applied in restricted image spaces, i.e., in sets that are specified by textual attributes (e.g., a name of a person is associated to each of the photos).

2.1 Photo similarity

Our aim is to measure the CBR system performance in finding photos similar to the given one. This seems justified since:

- most CBR algorithms are based on similarity matching
- search for photos similar to the one already found is one of the basic browsing routines applied by the end-users

In the CBR evaluation studies the tasks have had hardly any connection to real-life tasks of image collection users and the similarity of images has often been judged by the evaluators themselves. Thus, we do not know yet if the algorithms developed and evaluated produce similarity matching that is useful to users. Our intention is to introduce a user-oriented method and ground it on the user perceived similarities.

We assume that similarity is a multidimensional concept. If photos are composed of and can be described by various attributes as have been stated in many studies, they also may be similar to each other with respect to one or more of these attributes. Similarity perceived may be basedon visual or contextual criteria.

The visual criteria might be quite concrete, such as a certain object in a photo; shooting distance, e.g., 'face photos' of a person instead of full portraits; season of the year in photo etc. The criteria might be impressions interpreted by users from visual clues such as a photo with 'a sad atmosphere' or 'action'. They may also relate to the photo itself such as the photo direction or to its aesthetic values such as composition and colour.

The contextual criteria may refer to themes (e.g. 'nuclear power') or news contexts (e.g. ÎBosnia war') with which the photos are associated. The contextual criteria may also be connected to the attributes relating of the photo itself, e.g. photo source or cost.

We are now interested in the first group above, i.e., similarity based on the visual criteria. The similarity criteria that are important to and applied by the users are likely to depend on the task in hand. Thus, the similarity assessments should be made in the context of real-life tasks.

2.2 Test collection

We assume that the query generates a large set of thumbnail images (e.g., several hundreds). One suitable photo associated to one illustration idea (an example photo) has been identified by the user. The system should find all similar photos within the set of thumbnail images.

Thus, the test collection for the content-based photo retrieval algorithms consists of

I. a set of query photos (associated with an illustration idea for a article) and
II. for each query photo

   A. a set of potentially relevant photos retrieved by a broad textual query, and
   B. the complete set of human based similarity assessments between the query photo and the photos retrieved by the textual queries (Fig. 1).

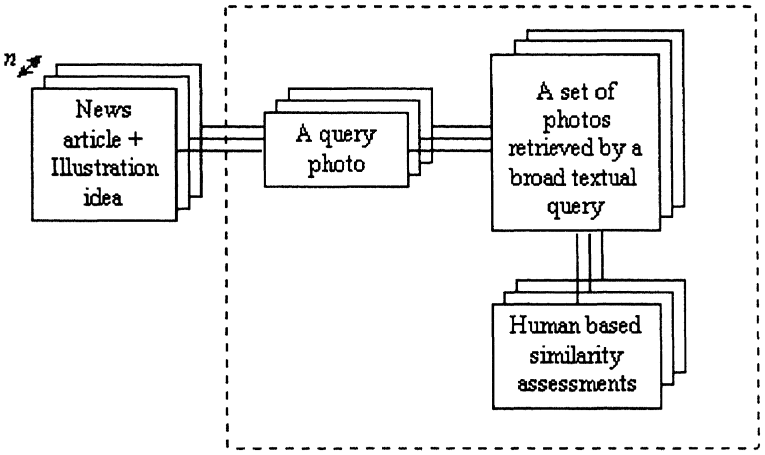
n
News article + Illustration idea
A query photo
A set of photos retrieved by a broad textual query
Human based similarity assessments

Figure 1. A test collection for content-based retrieval algorithms

The capability of a CBR algorithm in simulating human perceived similarities of photos can be measured by matching the query photo against the set of potentially relevant photos. Standard performance measures can be used: precision at the fixed recall levels (R=0.1-1.0) or after selected number of top documents (DCV=5, 10, 15, 20, ·).

2.2.1 Part one: Collecting a set of query photos

The purpose of this part of the procedure is to gather query photos for which the similarity of other photos will be assessed. The procedure is based on *simulated illustration processes.* According to the user study (see Appendix) the illustration processes of journalists embraced following components: (1) an article to illustrate, (2) illustration ideas discovered by a journalist, (3) formulating and executing a textual query, (4) browsing thumbnail images, (5) selection of candidate photos and (6) the selection of the Îbest oneâ.

As pointed out earlier, we need to give the assessor, besides the Îexample photoâ, also some context in which the similarity assessments will be made. Our hypothesis is that this context information will restrict the similarities applied to those which are important in real task situations. Our assumption is that in the illustration task, the article and the illustration idea generated for that article give appropriate context information for the assessor.

The generation of illustration ideas and finding the Îexample photosâ should be done by experts, e.g., journalist or other subjects with expertise such as students of journalism. Part 1 or the procedure is performed by simulating a real illustration process. A photo collection with concept-based indexing is available for searching. (The photo retrieval system and interface are those exploited by the journalists in their daily work. Thus, no guidance on the system usage will be needed.) A real newspaper article is given to the subject. The subject is asked to search the collection for creating illustration ideas, finding candidate photos and later the best one for illustration idea or, if many ideas, for each illustration idea. Think aloud protocol and tape recorder are exploited to gather data on the illustration ideas, selection of candidate photos and the selection criteria applied.

2.2.2 Part two: Collecting similarity assessments

Browsable photo sets of reasonable size (several hundreds) are created by the evaluators searching the base collection. The article, illustration idea recorded, the Îexample photoâ and the browsable photo set are presented to the assessor. The photo set is available through the photo retrieval system but only browsing features will be exploited. Retrieval system embrace thumbnail images, which are possible to enlarge and captions (written in photo agency and mostly in English). The assessor is asked to compare the Îexample photoâ to the photos in the set and find those photos which (s)he thinks are similar to the example. Observation and think aloud methods are used to gather deeper knowledge on the similarity perceived by users. This data should also reveal different dimensions of similarity if applied by the assessors.