Augmenting Navigation for Collaborative Tagging with Emergent Semantics

Melanie Aurnhammer[1], Peter Hanappe[1], and Luc Steels[1 2]

[1] Sony Computer Science Laboratory, Paris, France
[2] Vrije Universiteit Brussel, Brussels, Belgium
{melanie,hanappe}@csl.sony.fr, steels@arti.vub.ac.be

Abstract. We propose an approach that unifies browsing by tags and visual features for intuitive exploration of image databases. In contrast to traditional image retrieval approaches, we utilise tags provided by users on collaborative tagging sites, complemented by simple image analysis and classification. This allows us to find new relations between data elements. We introduce the concept of a navigation map, that describes links between users, tags, and data elements for the example of the collaborative tagging site Flickr. We show that introducing similarity search based on image features yields additional links on this map. These theoretical considerations are supported by examples provided by our system, using data and tags from real Flickr users.

1 Introduction

Collaborative tagging is a form of social software that has recently attracted a huge number of users. Web sites like Flickr, del.icio.us, Technorati, CiteULike, Buzznet, and Last.fm, to name but a few, encourage users to share photos, URLs, blogs, article references, and music titles. These data objects are associated with tags, common words freely chosen by the user. They describe a data item in a subjective and often associative way. It is an intuitive and effective method to organise and retrieve data. Tags are used to organise personal data, and are made public so that other users can access and browse them.

Tagging addresses the problem of providing meta-data for Web resources very differently from the method proposed by the Semantic Web initiative [1]. In the latter approach, Web content creators annotate their work using an ontology that was defined a priori by a group of experts. With tagging, Internet users describe resources using their own labels. This bottom-up approach is an instance of Semiotic Dynamics[2, 3], in which the uncoordinated actions of many users lead to the emergence of partially shared taxonomies. It resonates with earlier studies that used computational models to investigate the emergence of a shared lexicon by a population of autonomous agents [4, 5].

The effect of public exposure of tags is twofold. First, it creates an incentive for people to tag their data items in order to make them accessible to others. Second, the motivation of a high exposure of their data encourages people to align their tags with those of other users. Indeed, it has been observed that over time, the relative frequency of tags used to annotate a data element stabilises [2]. Thus, collaborative or social tags are not completely arbitrary and hence provide an interesting means to search databases. Especially domains like images or music, where semantic retrieval is an extremely hard problem, benefit from the tagging approach.

However, using tags alone for searching and browsing databases clearly has its limitations. First, people make mistakes while tagging, such as spelling mistakes, or accidental tagging with the wrong tag. Second, there is no solution to cope with homonymy, i.e. to distinguish different meanings of a word. Third, synonymy or different languages can only be handled by tagging data explicitly with all terms. One possible approach to solve the synonymy problem is to translate the local taxonomies into a global taxonomy which is used for querying and information exchange. The translation could be aided by mediators [6] and achieved through automated schema matching [7, 8]. However, this approach requires a one-to-one mapping of taxonomies, which is not always possible.

Our approach to tackle the shortcomings of collaborative tagging is to employ content-based image retrieval techniques. The combination of social tagging and data analysis provides the user with an intuitive way to browse databases, and allows him to experience and explore interesting new relationships between data, tags, and users, which we summarise in an augmented navigation map. It is a way to achieve emergent semantics because it can potentially ground the meaning of tags into the data [9–12]. To allow seemless integration beween tagging and data analysis, an adequate interface is obviously a crucial factor for user acceptance. Although there has recently been some effort to design tools that apply simple image analysis algorithms to Flickr images [13, 14], these tools work separately and are not employed for integrated navigation. Previous approaches for combining textual and image information such as [15] have been concentrating on recognising objects, which is an extremely hard problem. Using social tags has the advantage that semantic information is already provided by the user. Instead of attempting automatic semantic interpretation, we thus restrict ourselves on extracting global, low-level features.

In the following section, we describe first the user interface, including navigation possibilities and the tag visualisation. In Section 3 we present our method and give technical details of our implementation. Our concept of a navigation map is explained in Section 4. An example in Section 5 illustrates the improvement of the navigation map by introducing data analysis. We then conclude our work and present plans for future work.

2 The Interface

The interface of our system provides an intuitive way to combine collaborative tagging and content-based image retrieval. Data can be explored in different ways, either according to tags or using visual features. The application follows the user actions without enforcing a particular interaction pattern on him. The visualisation of tags is easy to understand and intuitive to use. The interface can be used to assemble a collection that can e.g. be shown as a slide show, or printed as photo album.

2.1 Tag Search

The entry point to assemble a collection is to search with a tag. When the user enters a tag, images annotated by this tag are shown in the *suggestion display area* at the lower right part of the screen (see Figure 1). The user has the possibility to perform a search on tags at any time in the process.

2.2 Suggestion Display

The images displayed in the suggestion display can be considered propositions from the archive. These images are either selected according to a common tag (see above), or by using similarity search. If images have been proposed according to a tag, the user can select one or more of these images and add them to his collection (*user collection area*, see Figure 1).

As soon as an image has been added to the user collection, two more functionalities are available: the *tag visualisation* (see section 2.4) and the search engine. The search can be started by choosing one or more examples of the collection in order to find visually similar images (see Section 3.2). The images proposed by the system are again shown at the lower right part of the screen. The user has the possibility to refine his search by simply selecting more images as positive examples, or others as negative examples, from the results.

2.3 User Collection Area

This area can be found in the top right part of the screen. The images selected by the user are displayed here and form his collection. Within this area, the user can start a search for visually similar images, or move images forward or backward inside his collection to change the order. In addition, he can display his images in a slide show.

2.4 Tag Visualisation

The visualisation of sets is shown in the upper left part of the screen (see Figure 1 and 2). The white circle in the centre of this area represents the collection of the user. The related tags – and their corresponding sets of photos – are shown as circles arranged around this centre. The visualisation feature has two modes. The first mode shows only the visualisation of tags related to the user collection. Before a search has been performed, the displayed circles are filled with a single colour. For clarity reasons, the number of displayed sets is restricted to the 32 largest sets. The size of each circle indicates the number of the photos contained in the set. The distance of the sets from the inner circle denotes the overlap of the images in the collection with the set, estimated by the number of common

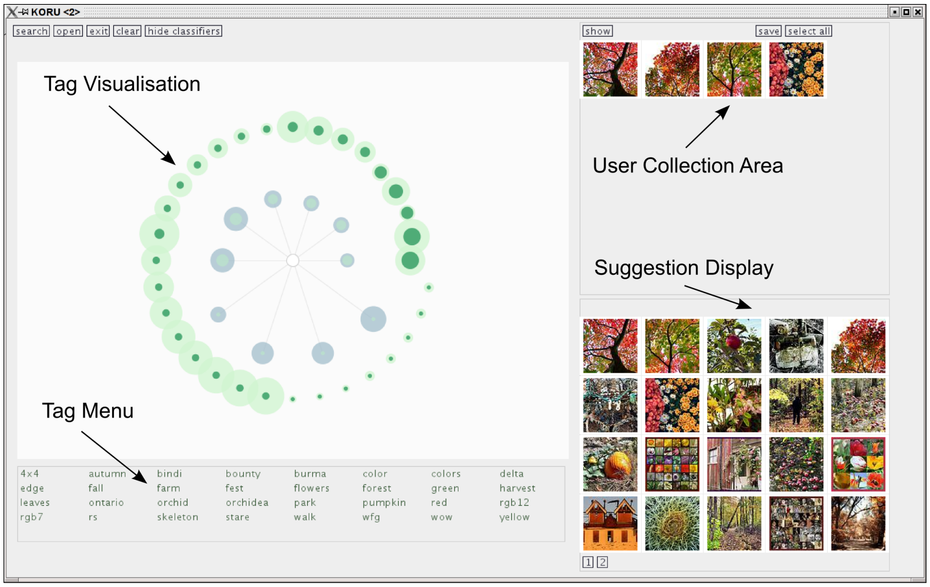


Fig. 1. Screenshot of the Interface. User Collection: images selected by the user to the theme "fall". Suggestion Display: Search results returned for first image of User Collection.

photos. When the user click on any of the circles, the corresponding images are shown in the suggestion area. We call the visualisation of these sets the *tag-sphere*.

The second mode is shown when the user applies a similarity search. The tag-sphere gets in the background and a second sphere of sets is drawn in a different colour. Each circle represents a tag that is related to the search results. The newly-added circles consist of two different colours. The inner, darker colour represents those images, which have been suggested by the classifier-engine. The outer, lighter circle represents images possessing the same tag, i.e. belonging to the same set, but those were not found by the classifier. A higher proportion of the inner circle indicates a relation between visual features and the tag related to the images in the set. The number of displayed sets is again restricted to the 32 largest sets. We will call the visualisation of these sets the *classifier-sphere*.

When images are added from the search results to the collection, the tag-sphere changes as well. Lighter circles inside some of the displayed circles get visible. These inner circles denote images found by similarity search that belong to the displayed set (i.e. they possess the same tag). The higher the proportion of the inner circle, the more likely is a relation between the visual similarity of the search image(s) and the tag. Every time an image is added to the user collection, the tag-sphere is updated. New sets are appended at the end of the sphere and thus a kind of tag history is represented. When an image is selected, the sets in which it is represented are highlighted and the corresponding tags displayed.

Below the tag visualisation is the *tag menu* (see Figure 5), where the tags of the current sphere are shown in alphabetic order. This gives the user the additional possibility to access the sets by going through the tags menu.

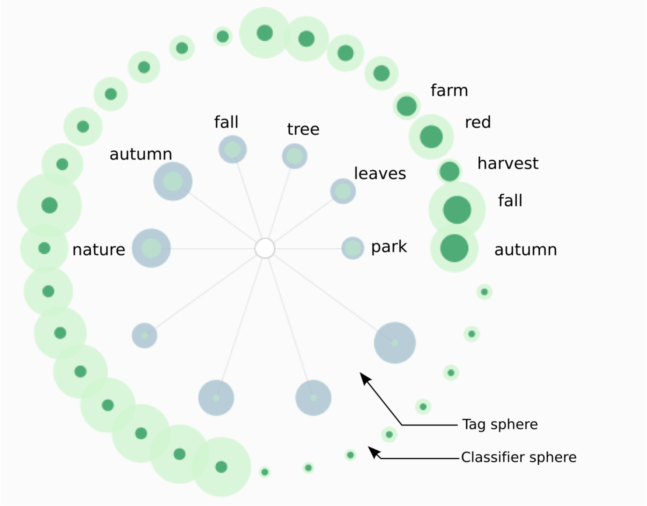


Fig. 2. Visualisation of Tags

Figure 2 shows an example visualisation, where a search has been performed on an image found by the tag “fall” (see first image User Collection, Figure 1). The two spheres show that there is indeed some correspondence between the tags of the search results and the visual features. In the classifier-sphere, there are at least five sets, where the inner circle takes a large part. The tags corresponding to these sets are “autumn”, “fall”, “harvest”, “red”, “farm”. This shows, that there is a high percentage of images among the search results, which are labelled with these tags. A similar behaviour can be observed in the tag-sphere after the user added some images from the results to his collection. Here, the sets with a large inner part are “park”, “leaves”, “tree”, “autumn”, and “nature”. These results are very interesting, since all these tags can be seen as reflecting the “fall-theme” of the user.

3 Technical Details

We tested our system with photographs downloaded from Flickr. Currently, we use about 3000 photographs from 12 randomly chosen users. In the following, we describe the visual features and the implementation of our retrieval process. Most of the techniques we used reflect either the state-of-the-art in image retrieval or are well-established standards in image analysis and pattern recognition. The idea of our system is to advance neither of these fields but to use the available tools in a new, intuitive, and creative way.

3.1 Features

We intentionally employ simple global features in our system. Rather than trying to recognise objects or even explain the meaning of an image, we seek to measure a certain "atmosphere", or a vague visual pattern, which we believe is possible to capture by low-level image features.

The visual features we used are colour and texture, i.e.

$$F = \{f_i\} = \{\text{colour,texture}\}$$

Colour Features Comparison of colour histograms is known to be sensitive to small colour variations caused e.g. by lighting conditions. In order to obtain a more robust and simpler measure of the colour distribution, we calculate the first two moments (mean and standard deviation) in RGB colour space. In addition, we use the standard deviation between the means of the three colour channels. Intuitively, this yields a measure for the "colourfulness" of an image. The feature has a value of zero for grey-scale images and increases for images with stronger colours. We map the values to a logarithmic scale in order to distribute them more equally. In total, the colour feature vector has thus seven dimensions.

Texture Features Texture refers to the properties that represent the surface or structure of an object. In our work, we seek to employ texture features that give a rough measure of the structural properties, such as linearity, periodicity, or directivity of an image. In experiments, we found *oriented gaussian derivatives* (OGD) to be well-suited for our purposes [16]. This feature descriptor uses the steerable property of the OGD to generate rotation invariant feature vectors. It is based on the idea of computing the "energy" of an image as a steerable function.

The features are extracted by a 2nd order dyadic pyramid of OGDs with four levels and a kernel size of 13x13. The generated feature vector has 24 dimensions. The first order OGD can be seen as a measure of "edge energy", and the second order OGD as a measure of the "line energy" of an image.

Feature Integration The distance between a query image and an image in the database is calculated according to the $l2$ norm (Euclidean distance). We use a linear combination of the distances in the colour and texture spaces to combine both features. In order to give the same initial weight to all features, the values are normalised linearly before calculating the distance. The joint distance d between a database image x_l and a query image s_k over all features spaces f_i is thus

$$d(x_l, s_k) = \sum_{i=1}^{N} w_i d_i, \qquad \text{with} \quad \sum_{i=1}^{N} w_i = 1$$

where N is the number of features in the set F and w is a weighting factor. In our implementation, w was set to $\frac{1}{N}$.

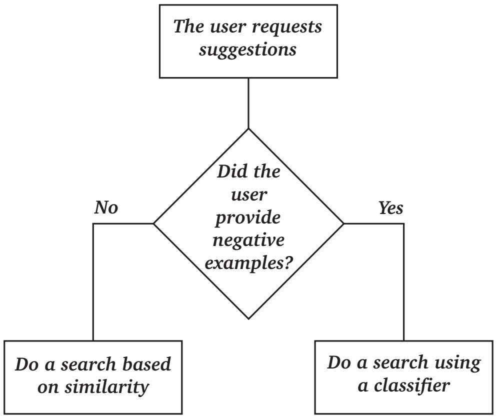

The user requests suggestions
Did the user provide negative examples?
No
Yes
Do a search based on similarity
Do a search using a classifier

Fig. 3. Based on the user selection, KORU either performs a similarity search, or a classification.

3.2 Search Process

The search for visually similar images starts with one or more images selected by the user. These initial images can be found through tags. In our implementation, we focussed on a totally user defined process: Not only is the number of selected images left to the user, he is also free in all further actions to take. When the results of the similarity search are displayed, the user can either (1) exclude images, (2) select images for refinement, (3) combine (1) and (2), or (4) simply not take any action. This distinguishes our approach from methods suggested for *relevance feedback* in image retrieval (see e.g. [17]), where the user is forced to take certain actions, such as giving feedback to every retrieved image, or where he has to follow a strict order of interaction.

Image Selection In case the user selects several images for his query (multi-image query), we think of these images as representing different classes. Thus, we accept images for retrieval that are similar to one of the query images. An alternative approach would be to average over the selected images which is, however, rarely relevant because the user might select visually distinct images. To give a simple example, a user selection of a yellow and a blue image should not yield green images as a result, but images that are either yellow or blue. Selection of the retrieved images is performed according to the following equation. Let X denote the archive and let x_l denote the l-th image in the archive. Let S denote a set of query images selected by the user. The distance D of x_l to S is then defined by

$$D(x_l, S) = \min_k d(x_l, s_k) \tag{1}$$

where d represents the distance of x_l to an image s_k contained in S, and k denotes the number of query images in S.

Refinement of Results If the user is not entirely satisfied with the retrieved images, he has the possibility to refine the results. He can choose (1) one or more images as positive examples, or (2) one or more images as negative examples, or (3) combine (1) and (2). In case only positive examples are chosen, these are added to the initial query images and the query is started anew by evaluating Equation 1 and selecting the n closest images. If the user chooses to provide the system with one or more negative examples, the retrieval process becomes a classification problem. (see Figure 3). The set of all user-selected images can then be seen as prototypes labelled either “positive” or “negative”.

It is important to note that the user might choose very different examples for the same label, i.e. he might choose for example, a red image with a very smooth texture, and a green image showing high contrast leaves both as positive examples. Therefore, a parametric classification method is not suited since it assumes the distribution of the underlying density function to be unimodal. In our case, it is a much better choice to employ a non-parametric approach that can be applied for arbitrary distributions and without the assumption that the forms of the underlying densities are known. Furthermore, it is important to ensure a smooth transition between retrieval and classification in order to avoid a drastic change of the results as soon as negative examples are selected.

A method that fulfils these requirements is a simple nearest neighbour classifier. Equation 1 basically defines the distance of an image in the database to a set of query images to be the distance between the test image and its nearest neighbour in the query set. For this reason, nearest neighbour classification is the natural choice to follow similarity retrieval. Let $P^n = \{x_1, \ldots, x_n\}$ denote a set of n labelled prototypes and let $x' \in P^n$ be the prototype nearest to a test point x. Then the nearest neighbour rule for classifying x is to assign it the label associated with x'.

4 Navigation Map

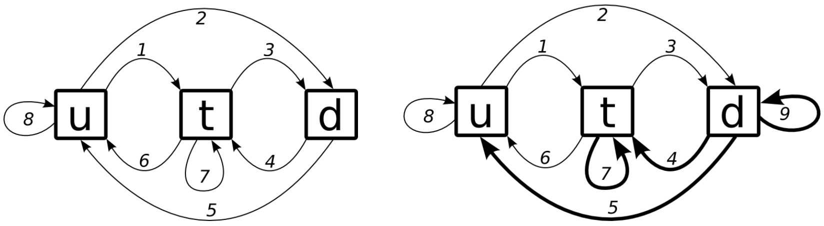
In this section, we introduce our concept of a navigation map to analyse the relationships between users, tags, and data in tagging systems. We compare the navigation possibilities between a simple tagging system, a system exploiting co-occurrence relationships between tags, and our proposed system. We show that we are able to introduce new links into the navigation map by combining tagging and data analysis.

4.1 Notation

We will refer to the set of all photos available on the server as the set P. The set of users of the tagging site will be denoted by U. The relation $\pi \subset U \times P$ defines which photos belong to which users. We define the set T to be the set of all the tags of all users. The relation $\tau \subset U \times T \times P$ represents which tags are used by which user to label a photo.

In addition to the entities presented above, users can create groups, in some form or another, and they can organise their photos in *sets*. For the sake of simplicity, we will focus here only on users, tags, and photos as the main entities.

Given the above mentioned relations, it is possible to define the navigation map that allows users to navigate from one entity (a user u, a photo p, or a tag t) to any other one. In general, any entity can provide an entry point for a navigation, based on its name (user name, tag, file name, photo title).



(a) Tagging only (b) Tagging and Similarity Search

Fig. 4. The navigation map with links between users (u), tags (t), and data (d).

4.2 Navigation Using Tags

In the following, we describe the navigation possibilities on a typical tagging site like Flickr, without taking into account co-occurrence relations. In such a system, a user can view all available tags (link 1, see Figure 4(a)). He can also directly access the photo collection of himself or another user (link 2). Through selection of a tag, the user can view all photos annotated by this tag (link 3). For a given photo, its associated tags can be displayed (link 4). There is a direct navigation possibility from a given photo to its owner (link 5). Most tagging sites allow users to establish *contacts*, which are direct links to other users (link 8).

4.3 Navigation by Co-Occurrence of Tags

One way to improve navigation is to exploit the co-occurrence relations between tags. This feature is already provided by Flickr, where it is referred to as "clusters". Navigation between related tags is represented in the navigation map by link 7 (see Fig. 4(a)). Clustering allows to solve the problem of synonymy to some extent. For instance, the tags “fall” and “autumn” appear in one cluster. However, this approach relies on a large number of people tagging their images explicitly with both “fall” and “autumn”. For example, establishing links between tags in different languages using co-occurrence does not work very well because most users tag in their native language only.

4.4 Augmenting the Navigation Map by Visual Features

The navigation map can be augmented further by introducing links that are not represented directly in the relations $\pi \subset U \times P$ and $\tau \subset U \times T \times P$ but can be found through analysis of the image data (see Figure 4(b)).

As mentioned in the introduction, tagging has some inherent shortcomings. Among them, the most problematic are the use of synonyms or different languages by different users, as well as incomplete tagging or the lack of any tags at all. To overcome these problems, we propose an approach that extends the navigation map by introducing an important link that is not represented directly in the relations π and τ. This link is based on visual similarity, which gives a direct relation between data (link 9, Fig. 4(b))). This new relation augments the possible navigation paths through the photo archive even further. By retrieving visually similar images, their associated tags are accessible as well, which yields additional links of type 4. Since the retrieved images link also to users, we get supplementary links of type 5. Furthermore, the relationship between the tags of the query image(s) and the retrieved images provides additional links of type 7.

In summary, visual features can introduce a link between photos that might be very remote in the existing structure provided by the relations π and τ. This way, it is possible to discover new photos that were previously difficult or even impossible to find.

5 Finding New Relations – An Example

The following example shows a possible navigation path of a user through a data collection. We illustrate, how these new links that connect data elements based on visual features can be found in our implemented system.

5.1 Initial Query

The user might want to start his search by a certain tag, e.g. “rust” in our example. Looking at the set of images labelled by this tag, he might select two interesting photos showing old, rusty windows (see Figure 5).

5.2 Similarity Search

Inspired by his initial selection, the user might want to find more images related to the “window” theme. He has two possibilities to continue his search: either

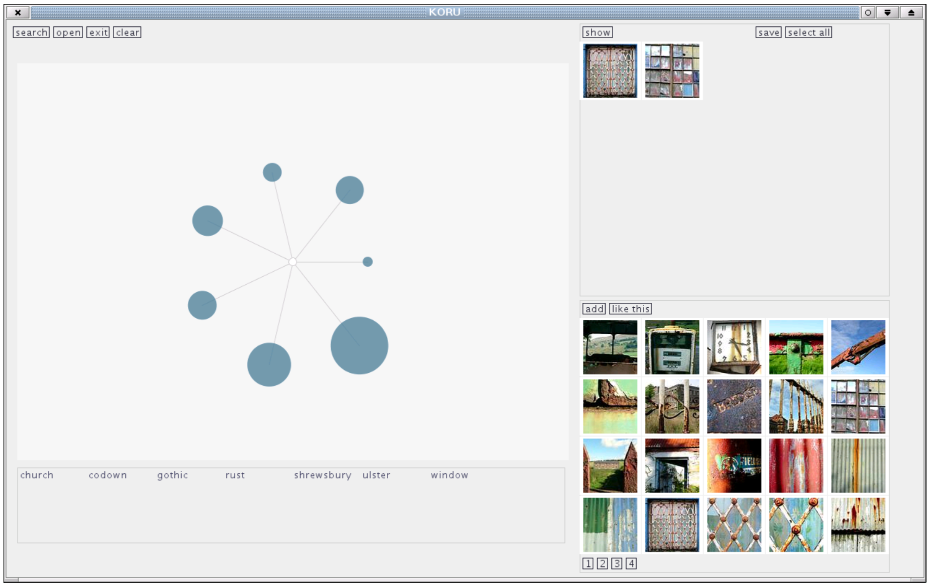


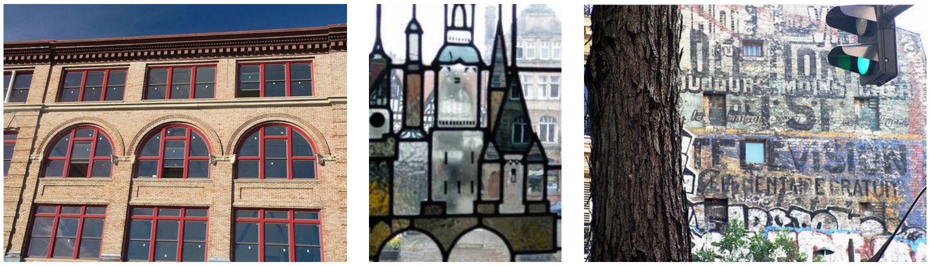
Fig. 5. Screenshot of the Interface. Suggestion Display: shows first 20 images tagged “rust”, User Collection: images selected from this set.

going directly to the tag “window” and view other images labelled with this tag, or he can start a similarity search. Let us assume that the “window” set does not provide any images that suit the user’s taste. Instead, he wants to do a similarity search based on the two photos he selected (see Figure 4, link 9). In Figure 6, three examples of the images retrieved by this search are shown. Among the tags attached to the image shown in Figure 6(c) is “fenêtre” (french for “window”) but not “window” itself. The image shown in Figure 6(a) is tagged only “downtown” and “oakland”, while no tags at all are provided for the photo shown in Figure 6(b). It can clearly be seen that none of these images could have been found with “window” as tag.

The results of the search can be further refined by the user as described in Section 3.

5.3 Further Exploration Possibilities

The new links the user has found by similarity search give starting points for further data exploration. He can, for example, follow the link for “fenêtre”, leading him to a whole new set of images of windows (see Figure 4(b), link 4). Another possibility is to start by the tags related to the set of retrieved images (shown in the *Tag Menu*, see Figure 1 and 5) and continue navigating through these tags and their co-occurrence relations (see Figure 4(b), link 7). For example, the results also include an image showing a photo of a high-rise building tagged with “tokyo”. The user might find this relation interesting and might want to



(a) Tags: downtown, oakland

(b) No Tags

(c) Tags: fenêtre, façade, france, paris, rivegauche, ...

Fig. 6. Examples of photos retrieved by similarity search, and corresponding tags.

look at more images tagged with "tokyo". And indeed, this data set contains several photos of high-rise buildings with many windows. Other, perhaps interesting new links are provided by tags such as "façade", "reflection", "grid", or "downtown". A further possibility is to navigate from an interesting photo to its owner (see Figure 4(b), link 5).

6 Conclusions

We introduced an approach that combines social tags and visual features in order to extend the navigation possibilities in image archives. Our user interface including an intuitive visualisation of tags was presented as well as the implementation details described. Furthermore, we explained our concept of a navigation map and showed how the initial map based on tags can be augmented by using visual features. A simple example illustrated how such additional links can be found in our implemented system.

We showed an example of our tag visualisation for a similarity search on an image tagged "fall". The results indicate that there is some correspondence between the visual features of the search results, and the tags that are attached to the result images. Although we cannot expect a simple one-to-one mapping from a tag to a visual category, there is indication that visual features can support the suggestion of new tags.

The work presented in this paper concentrated on establishing a framework for analysing our concept as well as a first implementation. An important next step will be to develop a quantitative measure of the improvement in navigation using formal methods and through user studies.

7 Future Work

Future work will also concentrate on exploiting the new relationships between data objects in order to propose tags for unannotated images. Moreover, we will investigate possibilities to add new user-to-user links based on profile matching according not only to the users' tags, but also to visual similarity of the users' data sets. Furthermore, we plan to show the generality of our approach by extending it to the music domain as well as to video clips.

Acknowlegements

This research was carried out and funded by the Sony Computer Science Laboratory in Paris with additional funding from the EU IST project TAGora (FP6-2005-IST-5).

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American (2001)
2. Cattuto, C.: Collaborative tagging as a complex system. Talk given at International School on Semiotic Dynamics, Language and Complexity, Erice (2005)
3. Steels, L.: Semiotic dynamics for embodied agents. IEEE Intelligent Systems (2006) 32–38
4. Steels, L., Kaplan, F.: Collective learning and semiotic dynamics. In Floreano, D., Nicoud, J.D., Mondada, F., eds.: Advances in Artificial Life: 5th European Conference (ECAL 99). Lecture Notes in Artificial Intelligence 1674, Springer-Verlag (1999) 679–688
5. Steels, L., Hanappe, P.: Interoperability through emergent semantics: A semiotic dynamics approach. Journal of Data Semantics (2006) To appear.
6. Wiederhold, G.: Mediators in the architecture of future information systems. IEEE Computer (1992) 38–49
7. Rahm, E., Bernstein, A.P.: A survey of approaches to automatic schema matching. VLDB Journal: Very Large Data Bases (10) (2001) 334–350 `http://citeseer.ist.psu.edu/rahm01survey.html`.
8. Tzitzikas, Y., Meghini, C.: Ostensive automatic schema mapping for taxonomy-based peer-to-peer systems. In: Proc. of CIA–2003, 7th International Workshop on Cooperative Information Agents - Intelligent Agents for the Internet and Web. Number 2782 in Lecture Notes in Artificial Intelligence (2003) 78–92
9. Santini, S., Gupta, A., Jain, R.: Emergent semantics through interaction in image databases. IEEE Transactions on Knowledge and Data Engineering **13** (2001) 337–351
10. Aberer, K., Cudré-Mauroux, P., Ouksel, A.M., Catarci, T., Hacid, M.S., Illarramendi, A., Kashyap, V., Mecella, M., Mena, E., Neuhold, E.J., Troyer, O.D., Risse, T., Scannapieco, M., Saltor, F., Santis, L.D., Spaccapietra, S., Staab, S., Studer, R.: Emergent semantics principles and issues. In: DASFAA. (2004) 25–38 `http://www.ipsi.fraunhofer.de/~risse/pub/P2004-01.pdf`.
11. Staab, S.: Emergent semantics. IEEE Intelligent Systems (2002) 78–86 `http://www.cwi.nl/~media/publications/nack-ieee-intsys-2002.pdf`.

12. Steels, L.: Emergent semantics. IEEE Intelligent Systems (2002) 83–85
13. Bumgardner, J.: Experimental colr pickr. `http://www.krazydad.com/colrpickr/` (2006)
14. Langreiter, C.: Retrievr. `http://labs.systemone.at/retrievr/` (2006)
15. Grosky, W.I., Fotouhi, F., Sethi, I.K., Capatina, B.: Using metadata for the intelligent browsing of structured media objects. ACM SIGMOD Record **23**(4) (1994) 49–56
16. Alvarado, P., Doerfler, P., Wickel, J.: Axon2 – a visual object recognition system for non-rigid objects. In: Proceedings International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA). (2001)
17. Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S.: Relevance feedback: A power tool for interactive content–based image retrieval. IEEE Transactions on Circuits and Systems for Video Technology **8**(5) (1998) 644–655