

Rerum Novarum: Interactive Exploration of Illuminated Manuscripts

Daniele Borghesani, Costantino Grana, Rita Cucchiara
Dipartimento di Ingegneria dell'Informazione, University of Modena and Reggio Emilia
Via Vignolese 905/b, Modena, Italy
name.surname@unimore.it

ABSTRACT

This paper describes an interactive application for the exploration and annotation of illuminated manuscripts, which typically contain thousands of pictures, used to comment or embellish the manuscript Gothic text. The system is composed by a modern user interface for browsing, surfing and querying, an automatic segmentation module, to ease the initial picture extraction task, and a similarity based retrieval engine, used to provide visually assisted tagging capabilities. A relevance feedback procedure is included to further refine the results.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: Multimedia Information Systems

General Terms

Design

Keywords

Illuminated manuscripts, image retrieval, tagging, relevance feedback, covariance matrices, user interaction, visual similarity

1. INTRODUCTION

Among all the fields in which modern multimedia research can bring a leap forward in the user experience on Digital Libraries (DLs), cultural heritage is undoubtedly one of the most promising. In our case, we deal with Renaissance illuminated manuscripts, whose digitalization allows a tremendous freedom of data elaboration. Experts in the fields of art, religion and literature will have the possibility to deeply study all the details of the works, making them interoperable and ubiquitous through platforms and places, they could personalize their research, extend their research to a wider set of works with similar characteristics, even mix research

results. On the other hand, normal people (museums visitors or people keen on art) can easily take a closer look to the work, increasing their experience and their interest.

In this technical demo, we present the initial results of the “Rerum Novarum” project (“new things” in latin), a human-centered multimedia system for enjoying illuminated manuscripts DLs, which substantially differs from the classic archive management tools usually employed for the consultation of digital works. In our system, we employed image analysis techniques for image segmentation, search and retrieval based on visual features and relevance feedback. In particular, we exploit covariance matrices as visual features [4], and we adopt a relevance feedback methodology based on a query reformulation and a feature space warping to improve the retrieval results based on user feedback [1]. Aside these techniques, the main novelty of the work is the joint use of visual similarity and relevance feedback for a *visually assisted tagging* of the manuscript, which eases the annotation process adding a remarkable amount of very useful textual information.

2. SURFING ARTISTIC DOCUMENTS

Multimedia indexing and representation are tasks that are highly desirable to be automated with limited role of user interaction. In fact, the goal of most systems is to remove the user from the indexing loop and to achieve full automation. This is very important in light of the huge volumes of multimedia data. However, it is unlikely that fully automated multimedia archiving systems can be achieved in the near future. In order to achieve a usable multimedia system today, we need to involve the user in the retrieval loop. This is not just because of the lack of today's technologies to achieve a fully automated system, but mainly because different users have different interests in their multimedia data and, therefore, efficient, usable, multimedia representations need to be personalized [3]. The correct understanding of the user intent in the process of interaction with a multimedia system is fundamental for a successful design of the system itself.

In our system, we tried to include all the typical interactions usually available in a multimedia system [2], i.e. *browsing* (when the end-goal of the user is not clear), *surfing* (when the end-goal is moderately clear) and *searching* (when the end-goal is very clear). To enclose all these modalities in the same user interface, we leverage the use of a visually assisted tagging procedure, which moves the user at the center of the multimedia experience. Starting from a clean system, with no prior information about the work (which

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10 ...\$10.00.

can be trained with a small amount of ground truth - just some pages - to provide a first automatic segmentation of the more meaningful pictures), with no prior tags (except for the ones automatically extracted from experts-made commentaries and proposed by the system itself), the user begins his analysis by *browsing* the pages of the work, and correcting the automatic segmentation if necessary. Once the user finds a particularly interesting detail, he proposes a tag to the selected picture and then continues his analysis by *surfing* by visual similarity. The system automatically provides back a set of similar pictures, for which the user can further provide relevance feedback. The results marked by the user as similar at the end may be given the same tags, so with minimal effort the user will accomplished the otherwise demanding effort to tag all pictures in the dataset sharing the same visual content. Finally, the user can keep on analyzing the work by *searching* using specific and reliable (combinations of) tags, which will cause the system to filter out the visualized dataset allowing the user to focus his attention to the sections of the work he is mainly interested on. Basically, it is a virtuous loop in which the similarity search by visual content will allow the extraction of similar pictures (pictures which will likely share the same tags), and tags will help the user in the process of search of content inside the manuscripts, and in the process of filtering results by topic.

3. UI DESIGN

The user interface and the user interaction paradigm is a fundamental aspect for a multimedia system, because is the only part of the system which will link directly to the user's emotion. For this reason, if the proposed user interaction is good, the user will be pleased to come back using the application.

For our system, we were inspired to the most successful user interfaces proposed in the market and in literature, and we came out with a very easy-to-use, minimal but yet effective interface which enclosed all the necessary functionalities. The entire interface is multitouch-friendly, since the finger is the most easy-to-use pointing device we have. The most complex functionalities are described by (a small amount of) clearly understandable icons, while —when it is possible— all the other minor functionalities (for example selection, navigating, zooming and so on) are generally triggered by convenient gestures (like swipe, pinch, etc...). Screenshots of the UI are provided in Fig. 1.

The entire surfing capabilities of the system is built around an overlay component called *tag navigator*. This component provides the user the access to the tags available in the system (Fig. 1(b)), allowing to move to higher or lower hierarchy levels. When selecting one or more tags, the system filters results in background, in order to provide the exact subset of data on which the user wants to focus on. We can exploit tags to conveniently limit the amount of information visualized to the user.

The selection of tags can be combined as an AND logic for a more fine grained filtering capabilities. For example, selecting a document tag focuses the attention of the system inside that document; selecting a document and a page tag focuses the attention of the system inside that particular page, visualizing all the pictures extracted from that page; finally, selecting a document and a picture tag will visualize all the picture of the document with that tag. The defined

tag navigator module is provided in the user interface as discussed in the next section. Document and page tags are assigned a priori, in particular a tag cloud-like approach is used and accordingly customized to extract the most relevant keywords from the commentary. Picture tags instead are built incrementally by users.

To each page of the document, the system can provide an automatic annotation, in the form of a list of selections (Fig. 1(c)). Each selection corresponds to a region of interest of the page, to which we can assign one or more tags (Fig. 1(d)). Once a particular page is selected, an *annotation editor* can be invoked in order to correct the automatic annotation provided by the system, or providing a custom annotation if necessary (Fig. 1(e)). A zooming functionality is included to perform a fine grain selection of pictures (Fig. 1(f)).

The entire process of managing tags for picture, assign a picture to folder of favorites pictures, or add comments to pictures is performed within a component called *social dashboard*. This component is currently highly experimental: it will be completed with the required functionalities as we start collecting significant amounts of user opinions. The set of pictures subdivided by tags, as well as the favorites set of pictures customized by users, will be visualized as a stack of pictures in an appropriate view called *pile viewer*. Comments are saved in this dashboard in the form as a sticky note (to mimic the traditional way librarian use to take notes about art works).

The entire process of visually assisted tagging is performed within the *similarity dashboard*. It is a fundamental component for the user experience in this application, since it cares about the visualization of the similarity search given a selected image (query) and gathers feedback information exploiting the pictures marked by users (through touching) as relevant (Fig. 1(g) and 1(h)). The similarity dashboard can subsequently trigger a refinement of the search exploiting the relevance feedback collected at each step, and finally can trigger the process of extending tags from query to relevant results, in order to improve the textual information available to the system and used to populate accordingly the tag navigator.

4. ACKNOWLEDGMENTS

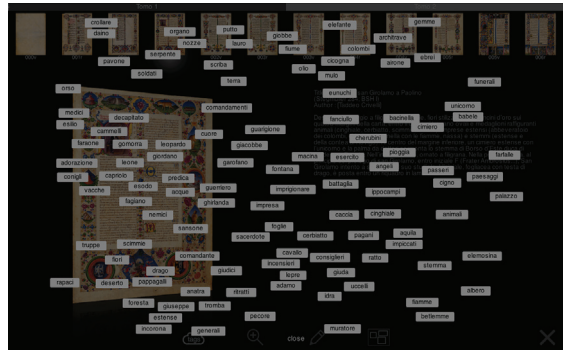
This work is supported by Franco Cosimo Panini Spa. We would like to thank Biblioteca Estense Universitaria for the availability of their digital library of illuminated manuscripts.

5. REFERENCES

- [1] Y. Chang, K. Kamataki, and T. Chen. Mean shift feature space warping for relevance feedback. In *IEEE Image Proc*, pages 1849–1852, 2009.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. In *ACM Comput Surv*, 40(2):1–60, 2008.
- [3] A. Elgammal. Human-centered multimedia: representations and challenges. In *Intl. Workshop on Human-centered multimedia*, pages 11–18, 2006.
- [4] O. Tuzel, F. Porikli, and P. Meer. Pedestrian Detection via Classification on Riemannian Manifolds. In *IEEE T Pattern Anal*, 30(10):1713–1727, 2008.



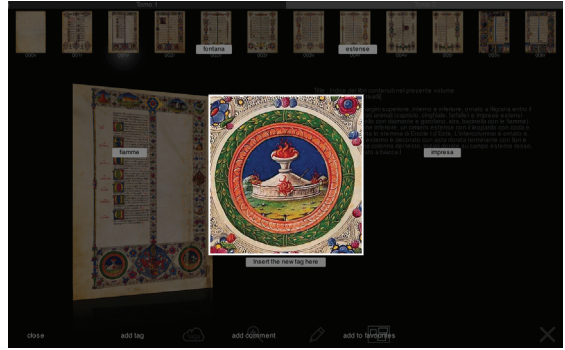
(a)



(b)



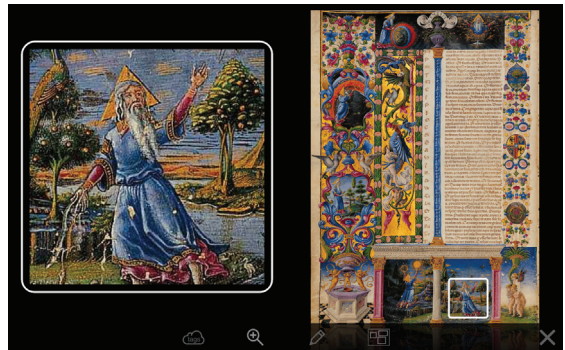
(c)



(d)



(e)



(f)



(g)



(h)

Figure 1: Screenshots of the “Rerum Novarum” project User Interface