Content Based Search System for Ethiopian Art Content on the Web

Hewan Shilema
HiLCoE, Software Engineering Programme, Ethiopia
hewanshil@gmail.ocm

Mesfin Belachew
HiLCoE, Ethiopia
Ministry of Communication and Information Technology
mesfinbelachew@mcit.gov.et

Abstract

In this paper a content based search system is designed and implemented to allow users to access different art works of Ethiopian artists and makes searching information from the Web very easy and quick. Search engines were originally described as automated programs that compile and update databases without human intervention to serve as a bridge between a user and the artworks. Our search engine is domain-specific. Domain-specific Web search engines are effective tools for reducing the difficulty experienced when looking for information from the Web. Existing methods for building domain-specific Web search engines require human expertise or specific facilities. However, we can build a domain-specific search engine simply by adding domain-specific keywords to the user's input query.

In this paper, the necessity of designing an efficient search system and the significance of this work is addressed in considering the given drawbacks in the existing management systems for the available art works. Then an overview of current image database search methods is presented, followed by explanation of the method that was selected for the construction of our content based search system and the reasons for choosing the method. After this, the structure of the proposed contented based search system and its implementation are described in detail.

Keywords: Search Engine; Content- Based Retrieval; Image Database

1. Introduction

In many areas of commerce, government, academia, and hospitals, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of analog photographs, diagrams, drawings, paintings, and prints. Usually, the only way of searching these collections was by keyword indexing, or simply by browsing. Digital image databases, however, open the way for content-based searching [3].

Content-Based Image Retrieval (CBIR), as we see it today, is a technology that in principle helps to organize digital picture archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. This characterization of CBIR as a field of study places it at a unique point in time within the scientific community. While we witness continued effort in solving the fundamental open problem of robust image understanding, a lateral bridging of gaps between some of these research communities is being gradually brought about as a by-product of such contributions, the impact of which can potentially go beyond CBIR. Again, what we see today as a few cross-field publications may very well spring into new fields of study in the foreseeable future [5].

Nowadays there are a number of art galleries in Ethiopia and these galleries have their own websites. Art gallery websites display different art works that they show on the exhibit time and famous painters' works are found on different art galleries. From this anyone who loves to see or buy art works cannot find

art works in a collected way. So having a search system minimizes the difficulty of getting art works from different art gallery sites and it combines all artistic works together.

We developed a system called content based search system for Ethiopian art content on the web. There are many ways available to gather information about art works, such as from newspapers, by asking information from friends, and online websites. Yet, the information available in newspapers also may be mostly from large companies/galleries advertising their newly shows. Information obtained from friends may be limited or similar as they may obtain them from the same source.

In order to improve the relevance and usefulness of the search for art work information for a particular location and therefore to reduce the time the user would need to spend in analyzing the search result, content based search system for Ethiopian art content on the web is developed. It enables and works as a bridge between a user and the artworks. Search engines were originally described as automated programs that compiled and updated databases without human intervention.

Using a search engine, the problems incurred in the previous artwork systems can be resolved. Using efficient and advanced compression techniques, we can digitize a large image effectively.

Therefore, a large piece of artwork can be easily presented to users. By providing a reasonable digitized copy of the original artworks, we avoid directly handling some delicate artworks.

2. Related Work

2.1 Overview of Content-Based Image Retrieval

One of the tools that will be essential for future electronic publishing is a powerful image retrieval system. Image retrieval systems associate keywords or text with each image and require the user to enter a keyword or textual description of the desired image [6]. Content-based image retrieval is a technique which uses visual contents to search images from large scale image databases according to users' interests. Images were first annotated with text and then searched using a text-based approach from traditional database management systems.

As the diversity and size of digital image collections have grown exponentially, efficient image retrieval is becoming increasingly important. Large image databases are difficult to browse with traditional text searches because the task of user based annotation become very time consuming, as the text often fails to convey the rich structure of images. A content-based retrieval system solves this problem where retrieval is based on the automating matching of features of the query image with that of image database through some image-image similarity evaluation. Therefore, images will be indexed according to their visual content such as color, texture, shape or any other feature or a combination of visual features [8].

Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. In typical content-based image retrieval systems, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors [7].

2.2 Content-Based Retrieval Systems

Content-based retrieval uses the contents of images to represent and access the images. A typical content-based retrieval system is divided into off-line feature extraction and online image retrieval [1]. A conceptual framework for content-based image retrieval is illustrated in Figure 1 [1]. In offline feature extraction, the contents of the images in the database are extracted and described with a multi-dimensional feature vector, also called descriptor. The feature vectors of the image constitute a feature dataset stored in a database. In online image retrieval, the user can submit a query example to the retrieval system in search of desired images. Then the system represents this example with a feature vector [1].

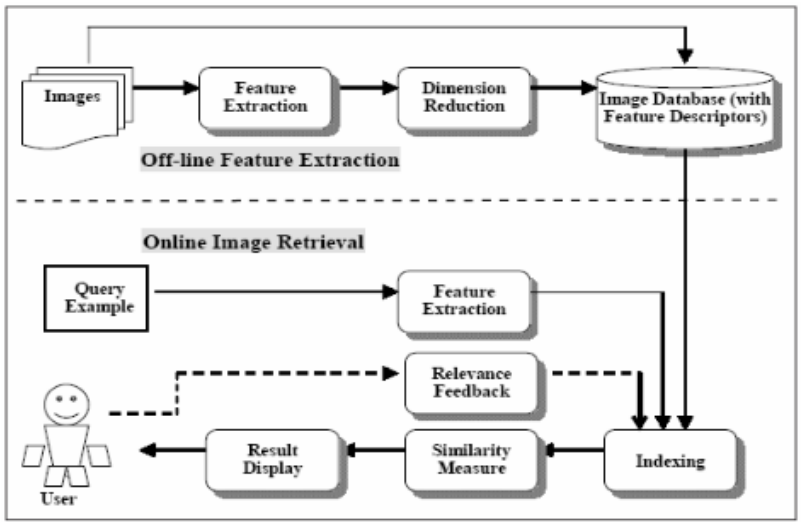

Figure 1: *Conceptual Framework for CBIR*

At the end, the system ranks the search results and then returns the results that are most similar to the query examples. If the user is not satisfied with the search results, the user can provide feedback to the retrieval system, which contains a mechanism to learn the user's information needs. The following sections introduce each component in the system [1].

Most CBIR systems have a single-interaction modality where the user presents an image as a query and obtains matching images as results. This one-shot modality for querying information repositories has a history of success ranging from decades old classical and textual database queries and information retrieval techniques to more recent web searching facilities. Though such one-shot query modality has been effective in dealing with large form of text-based information repositories, we believe that the same modality is undesirable, ineffective, and inefficient for a significant class of image retrieval domains. Hence, interactive CBIR systems are gaining popularity in the past few years [2].

A number of keyword-based general Web search engines allow image media types. A number of other general search engines are more specifically designed for images. There are many special image collections on the Web that can be searched with alphanumerical keywords. [3]

2.3 Content Based Image Retrieval Frameworks

The basic idea of CBIR is that a set of features is used that allows finding images that are similar to the query image. For different properties of images, different features may be used. The goal of feature extraction is to find informative variables based on image data which can be seen as a kind of data reduction [4].

The user interface normally consists of a query formulation part and a result presentation part. Specification of which images to retrieve from the database can be done in many ways. One way is to browse through the database one by one; another way is to specify the image in terms of keywords or in terms of image features that are extracted from the image, such as a color histogram. Still another way is to provide an image or sketch from which features of the same type must be extracted as for the database images in order to match these features [3].

Relevance feedback is about providing positive or negative feedback about the retrieval result so that the system can refine the search. Other features that are used to specify queries are color, texture, shape, spatial layout, and face. Color features are often obtained directly from the pixel intensities. Although a precise definition of texture has been allusive, the notion of texture generally refers to the presence of a spatial pattern that has some properties of homogeneity. In particular, the homogeneity cannot

result from the presence of only a single color in the regions, but requires interaction of various colors [3].

Higher level features are increasingly more specific, and thus less widely used. However, faces are frequently present in pictures and relatively often used as a feature [3].

3. The Proposed Solution

Automating the existing system through centralized database with web based search system is chosen to address current problems that most artists face. Some problems that appear on most of the art work system are the following.

- *Time and place constraints*: In the current art work management systems, the artworks are stored in art gallery rooms, which means that users need to go to the gallery to view these artworks. As far as the user is concerned, this is inconvenient.
- *Difficult or impossible physical access*: It will be difficult or impossible to access some artworks that may be spread across a large physical area or composed of material that may require delicate handling.
- *Safety*: Some artworks are very valuable and precious. Exposing these artworks directly to users probably will result in wear, contamination and even damage. Therefore, it is not safe to expose some real artworks to the public.
- *Lack of Information*: users have no way to view the gallery collection as a whole and have little or no idea about what is contained in the collection.

Automating the existing system outsmarts the other alternative. So we decided to develop the system as web based application and it has the following benefits.

- We want to display multiple images on the screen. Therefore we decided to use thumbnail sized images. One big advantage of using thumbnail sized images is that the user gets a better overview with more than just one image. Also no information is lost because the user is still capable of understanding the image's contents.
- Another advantage is digital storage and retrieval which is essential to World Wide Web programs.
- Online image view is done online which saves time of the buyers of the art and decreases the workload of the galleries.

4. System Architecture

The architecture of the system shows the general structure of the search system for retrieving information from the database server. Figure 2 illustrates the architecture of the system. As shown in the figure, the user search query is sent to the web server and then the web server sends the query to the index server and then the index server refers the database server to search the user query and finally the database server returns the search result for the end user.

Figure 3 shows the user interface for content based search system for Ethiopian art content on the Web. Anyone who wants to use this system can get many art works and detailed information about the artist, art news, the address of art museums in Ethiopia, announcements that help artists and art work lovers about art work exhibitions and upcoming events, and the system has also links that help a user who wants to buy an art work to get detailed information about the art work. There is also a forum link that helps the user to discuss issues by creating user account.

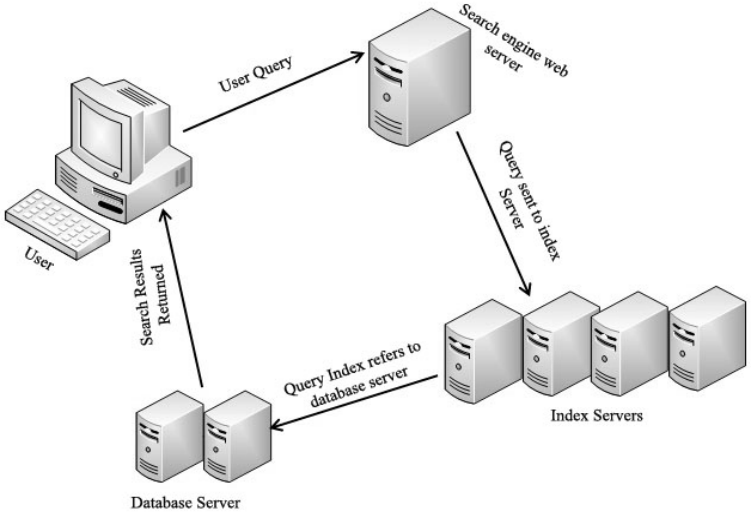

Figure 2: System Architecture for Content Based Search System

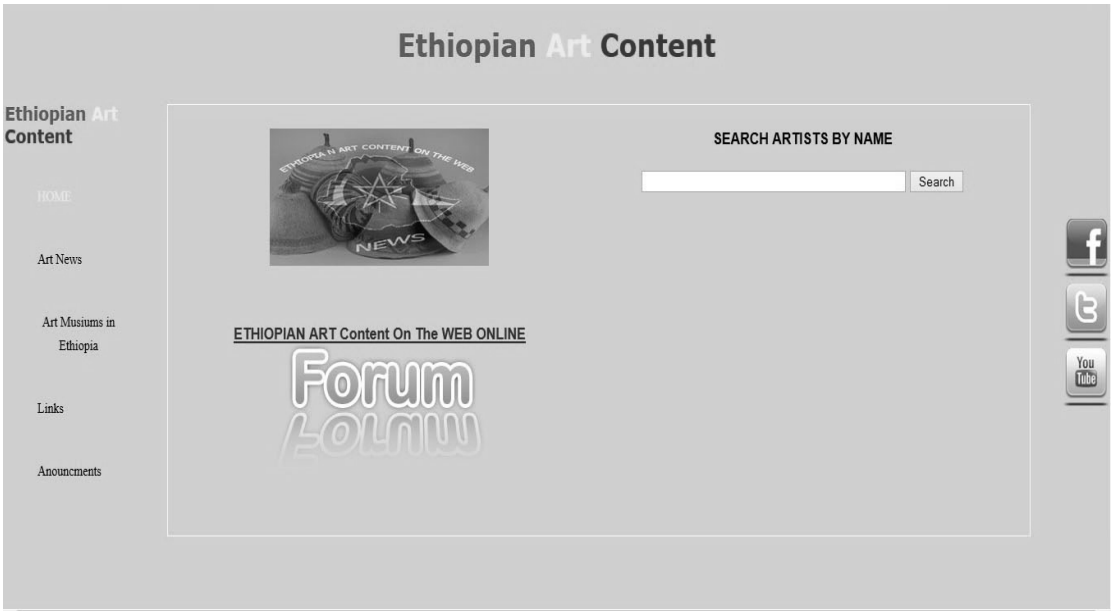

Figure 3: Front Main Menu

Figure 4 displays the interface for the auction site that help the user to get detailed information about artworks, the auction site lists the artworks by category, and anyone who wants to buy an artworks can select from the listed category and can see detailed information about each art works, the price, status of the art work if it is sold or not, the location where the art work is found and which bank is preferable for the payment process and so other information are displayed on this auction site.

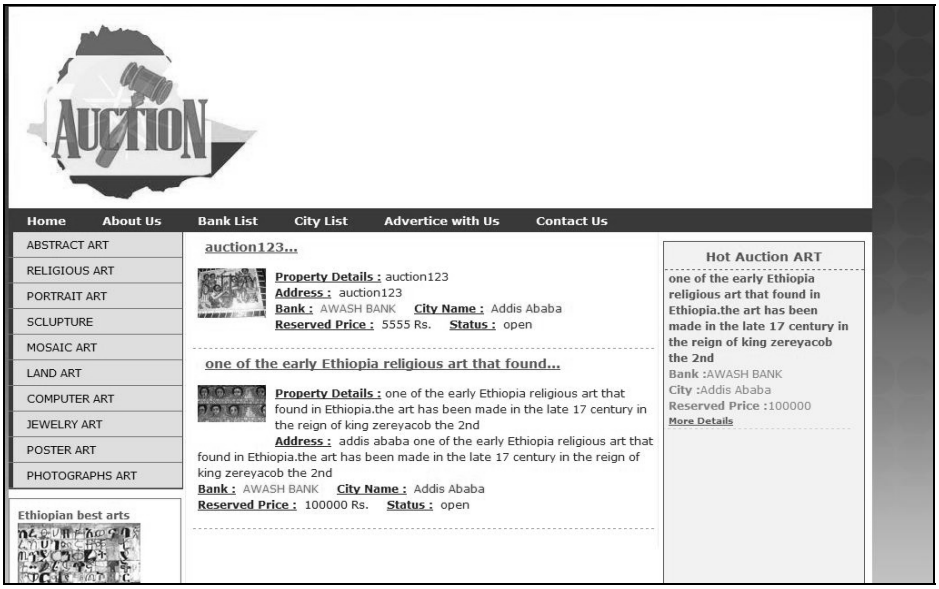

Figure 4: Main Menu for the Auction Site

5. Conclusion and Future Work

Content based search system for Ethiopian art content on the web is developed to provide more relevant search results for users by allowing the user to input information by specifying keywords on the search system.

Currently due to globalization the world is highly becoming competitive, so organizations have to direct their attention on using recent technology to be competitive. This can be real if they are able to use information communication technology (ICT) to successfully achieve their objectives.

Since the system was a proof-of-concept system, it is not fully functional and requires more development in the future.

References

[1] K. Vijay and R. Anitha, “A Content-Based Approach to Image Database Retrieval”, MCA, K. S. R. College of Technology, Tiruchengode.

[2] Sabyasachi Saha and Sandi Sen, “Agent Based Framework for Content Based Image Retrieval”, Department of Mathematics and Computer Science, The University of Tulsa.

[3] Remco C. Veltkamp, MirelaTanase, and Denielle Sent, “Features in Content-Based Image Retrieval Systems: A Survey”, Utrecht University, Department of Computer Science, Utrecht, the Netherlands.

[4] Zainab InrahimAbood, Israa Jemeel Muhsin, and Nabeel Jameel Tawfiq, “Content-Based Image Retrieval (CBIR) Using Hybrid Technique”, International Journal of Computer Applications (0975-8887), Vol. 83, No 12, December 2013.

[5] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age”, The Pennsylvania State University.

[6] Yuri Lausberg, “A Content-Based Image Search Engine in Cyberspace,” Application-Oriented Computer Science Leiden Imaging and Multimedia Group, Department of Computer Science, Leiden University, August 19, 1996.

[7] Fuhui Long, Hongjiang Zhang, and David Dagan Feng, “Fundamentals of Content-Based Image Retrieval”.

[8] Gaurav Jaswal, Amit Kaul, and Rajan Parmar, “Content Based Image Retrieval Using Color Space Approaches”, International Journal of Engineering and Advanced Technology (IJEAT) Vol. 2, Issue-1, October 2012.