

SELF CONFIGURABLE QUERIES FOR DIGITAL IMAGE LIBRARIES

Panagiotis Karagiannis^{*}, *Nikolaos Doulamis*⁺, *George Varkas*^{*} and *Konstantinos Paparrizos*^{*}
^{*}Department of Applied Informatics, University of Macedonia, Egnatia 156, 54006, Thessaloniki,
⁺National Technical University of Athens, Athens, Greece

ABSTRACT

Since multimedia information is characterized by motley types of media with different properties, multimedia content retrieval in digital libraries requires dynamic reconfigurable architectures. In this paper, mobile agents are being used as the basis of such a dynamic reconfigurable architecture. The proposed architecture is further enhanced with innovative relevance feedback algorithms for interactive multimedia retrieval. In particular, an optimal recursive relevance feedback mechanism is examined by adaptively estimating a parametric correlation-based similarity measure.

1. INTRODUCTION

The recent explosion in using media-rich applications has resulted in an appreciation for the value of multimedia content, and a realization of the challenges in managing that content [1, 2]. New multimedia applications such as content based retrieval, require new tools and algorithms for efficient and reliable organization and management.

Humans perceive and characterize content using high-level concepts not related to image pixels [3]. Furthermore, they usually interpret the same semantic content in a different way, at different times. This “time varying” multimedia content interpretation and classification can be addressed by introducing adaptive multimedia content management schemes.

In order to overcome the aforementioned difficulties, on line learning strategies have been proposed, called relevance feedback, originally developed in the traditional text-based information retrieval systems [4]. In the relevance feedback approach, the human is considered a part of multimedia management process in an interactive framework.

Recently, relevance feedback algorithms have been extended from text-based information retrieval to Content-Based Image Retrieval (CBIR) systems [4-10]. These approaches however, are either ad hoc methods exhibiting no optimal learning strategy, or yield unstable performance.

In this paper, an efficient dynamic learning scheme is proposed for updating not only the similarity measure used for ranking multimedia data, as in all the above feedback schemes, but also configuring the system while it is functioning. Mobile agents are being used as the basis of the proposed a dynamic reconfigurable scheme yielding to a highly dynamic multimedia retrieval system which allows adding and removing components at run-time. The presented architecture allows adding or removing system components at run-time based on interaction by the user, who provides information about the relevance (irrelevance) of the content.

A code on demand approach is proposed to support the dynamic reconfiguration of the agent properties. Upon instantiating an agent, its code, an initial configuration and its

programming interfaces are transferred to the target site. The results return to the end user and an evaluation process commences based on the user’s interaction. User’s evaluation activates an on-line learning scheme which is responsible for dynamically reconfiguring the system’s architecture. Reconfiguration is conducted by analyzing the important visual descriptors (as obtained after the user has classified the results as relevant / irrelevant) in more details.

This paper is organized as follows: Section 2 presents an overview of the system’s architecture adopted for dynamic reconfiguration based on user’s interaction. In section 3, we describe the reconfiguration algorithm. In particular, we present the optimal relevance feedback approach, followed by the method for code selection and user’s profile estimation. Finally, section 4 concludes the paper.

2. SYSTEM ARCHITECTURE

The proposed systems’ architecture consists of three main modules; the acquisition, the adaptation and the retrieval module. The first module is responsible for receiving and processing a visual query submitted by a user, while the adaptation module for reconfiguring the system using a relevance feedback mechanism based on user’s interaction. Finally, the retrieval module is used for activating and monitoring the mobile agents used for media retrieval. The basic components of these three modules are described in the following sub-sections.

2.1. The Acquisition Module

The internal architecture of the acquisition module is depicted in Figure 1(a). As can be seen, the module consists of two main components; the Portal and the Processing component.

The portal component is responsible for receiving user’s queries. It has an authentication unit for authorizing a user. The portal server component follows the authentication component. It includes a web-based portal that functions as entry point for users and administrators.

The goal of the processing component is to apply processing algorithms on the submitted visual data (query) before the retrieval module initiates the retrieval process. Since raw image data are not appropriate for describing rich visual content, content modeling is required. This is done by extracting several descriptors (features) so that a better representation of the visual information is accomplished. In our implementation two different types of visual descriptors are used; the global and the object based descriptors [10].

2.2. The Retrieval Module

The retrieval module is responsible for creating and dispatching the mobile agents used for mining and retrieving multimedia information. The retrieval module consists of two platforms; The Home and the Destination platform. The home platform is used for activating, polling and monitoring the mobile

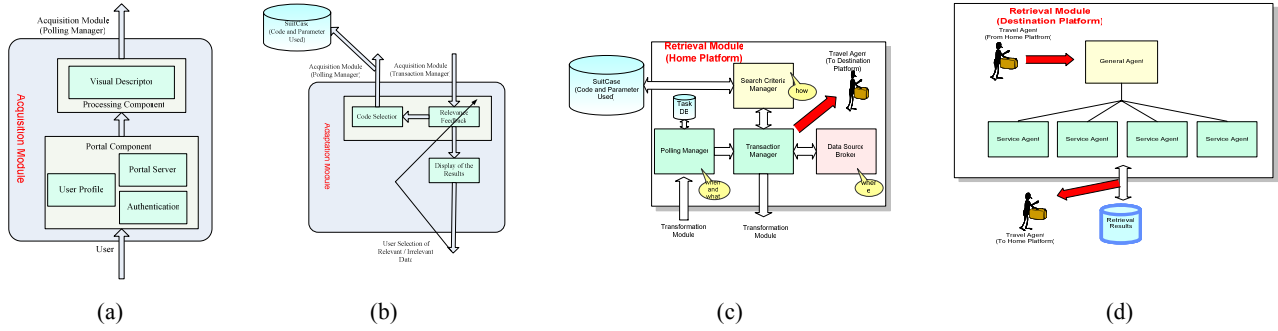


Figure 1: The block diagram of the (a) acquisition and (b) adaptation module, c) retrieval module (home platform), and d) retrieval module (destination platform).

agents, while the destination platform refers to the remote web site in which the mobile agent is sent.

A set of mobile and stationary agents are embedded in the home platform unit capable of interacting and exchanging relevant information. Figure 1(c) provides a graphical representation of the components included in the home platform. The Polling Manager (PM) is a stationary agent under the task of deciding when a new instance of a Travel Agent (TA) has to be created. This is achieved by consulting the appropriate records in the task database.

The Transaction Manager (TM) is a stationary agent responsible for instantiating and launching travel agents when triggered by the Polling Manager. It is also responsible for creating the configuration parameters for the mobile agents. During a travel agent's journey, the transaction manager monitors its activity.

Travel Agents are mobile agents, capable of fetching information. The travel agent can carry all necessary parameters for conducting the acquisition process. These parameters are stored in the suitcase database created by the Transaction Manager and the Adaptation Module.

The Data Source Broker is a stationary agent that maps the set of submitted data into the corresponding destinations where the desired information might be found. The data source broker keeps a rated list of available data sources that are indexed.

The Search Criteria Manager is a stationary agent that encapsulates the knowledge for selecting the appropriate algorithms for retrieving and processing multimedia data. During the system's reconfiguration, the Search Criteria Manager communicates with the suitcase database which contains the processing algorithms as they have been selected by the Adaptation Module.

A block diagram of the components included to each destination platform is illustrated in Figure 1(d). The destination platform comprises the general and the service agents, described in the following.

The General Agent is a stationary agent constituting the mediator between the arriving travel agent and the Service Agents.

The Service Agents are stationary agents that perform the information extraction from the web data sources.

2.3. The Adaptation Module

The purpose of the adaptation module is to reconfigure the system according to the user's information needs and preferences. The core component of this module is the relevance feedback mechanism. There are also two additional components, the

Display of the Results and the Code Selection component respectively (see Figure 1(b)).

Relevance feedback is an on-line learning scheme for updating the system response according to the user's information needs and preferences. In this framework, a feedback is established from the user to the system. In our approach, relevance feedback is used for estimating the importance of different types of descriptors to the similarity measure. Then, the importance energy of the descriptors is used for estimating an appropriate software component (code) used by the mobile agent, in order to retrieve multimedia information that better characterize the user's information needs.

The Code Selection component estimates the appropriate code used by the mobile agents. The algorithm is based on the results of the relevance feedback component. More specifically, the code and software components which correspond to the descriptors of highest energy (importance) are selected. The selected codes are stored in the suitcase database, in which mobile agents mine.

The Display of the Results component is responsible for distributing the results to the end user. This component is required for the implementation of the relevance feedback algorithm to allow user evaluation about the relevant retrieved multimedia data.

3. THE INTERACTIVE LEARNING PROCEDURE

In this section, we describe the relevance feedback algorithm and the method used for code estimation and selection.

3.1. Relevance Feedback

Let us denote as \mathbf{f}_i the feature vector of the i^{th} image in the database and as \mathbf{f}_q the feature vector of the query image as it has been obtained by the visual descriptor component of the acquisition module.

As similarity measure for multimedia data ranking the correlation of the vectors \mathbf{f}_q and \mathbf{f}_i is used.

$$\rho_w(\mathbf{f}_q, \mathbf{f}_i) = \frac{\sum_k w_k \cdot f_{q,k} \cdot f_{i,k}}{\sqrt{\sum_k w_k^2 \cdot f_{q,k}^2} \cdot \sqrt{\sum_k f_{i,k}^2}} \quad (1)$$

where w_k are the descriptor weights and $f_{q,k}$, $f_{i,k}$ the k^{th} feature element of the vectors \mathbf{f}_i and \mathbf{f}_q respectively. It should be mentioned that the similarity measure is not symmetric (the weights are only used for the query image). This means that only

the query feature vector is scaled (using the weights w_k) in order to fit the user's information needs as provided by a set of relevant / irrelevant images as selected by the user.

The correlation measure similarity measure is better than a distance- based similarity measure (such as the method of [7]), since it is scale and translation invariant and it also gives a better physical relation to the similarity of two signals.

3.2. Optimal Weight Updating and Code Selection

Let us first assume, without loss of generality, that m out of the M best retrieved data are selected as relevant by the user and let us denote as $\mathbf{y}_i, i=1, \dots, m$ ($1 \leq m \leq M$) the feature vectors of the m selected data. Then, the weights w_k are estimated so that the following criterion is maximized,

$$C(\mathbf{w}) = \sum_{i=1}^m \rho_{\mathbf{w}}(\mathbf{f}_q, \mathbf{y}_i) = \sum_{i=1}^m \frac{\sum_k w_k \cdot f_{q,k} \cdot y_{i,k}}{\sqrt{\sum_k w_k^2 \cdot f_{q,k}^2} \cdot \sqrt{\sum_k y_{i,k}^2}} \quad (2)$$

The energy $E_{\mathbf{y}_i}$ of \mathbf{y}_i is given as $E_{\mathbf{y}_i} = \sqrt{\sum_k y_{i,k}^2}$.

Then, the optimal weights are estimated by differentiating $C(\mathbf{w})$ and setting the $\partial C(\mathbf{w}) / \partial w_n = 0$, for all weights w_n

$$\frac{\partial C(\mathbf{w})}{\partial w_n} = 0 \Rightarrow \sum_{i=1}^m \frac{y_{i,n}}{E_{\mathbf{y}_i}} \cdot \left(\sum_k w_k^2 \cdot f_{q,k}^2 \right) = \left(\sum_{i=1}^m \sum_k w_k \cdot f_{q,k} \cdot \frac{y_{i,k}}{E_{\mathbf{y}_i}} \right) \cdot w_n \cdot f_{q,n} \quad \forall n \quad (3)$$

As a result, (3) corresponds to a non-linear system of equations. By dividing two equations of the form shown in (3) one over the other, for example the ones corresponding to $\partial C(\mathbf{w}) / \partial w_n = 0$ and $\partial C(\mathbf{w}) / \partial w_l = 0$, $n \neq l$, the following relation of weights w_n and w_l , is obtained

$$w_n = w_l \cdot \frac{f_{q,l}}{f_{q,n}} \cdot \frac{\sum_{i=1}^m \frac{y_{i,n}}{E_{\mathbf{y}_i}}}{\sum_{i=1}^m \frac{y_{i,l}}{E_{\mathbf{y}_i}}} \quad \forall n \neq l \quad (4)$$

Equation (4) requires an additional equation for estimating the unknown weights. This constraint is the norm of the weights, which is assumed to be constant, $\|\mathbf{w}\|_2 = 1$

An explanation of this is due to the properties of $C(\mathbf{w})$. Indeed scaling the feature vector has no impact on the correlation. Let us assume, without loss of generality, that the free variable of (4) is the first weight w_0 , i.e., $l=0$. Then, expressing all weights $w_n, n \neq 0$ with respect to w_0 and using equations (4) and the constraint, we can conclude that

$$w_0 = \sqrt{\frac{1}{1 + \sum_k A_k^2}} \quad \text{with} \quad A_k = \frac{f_{q,0}}{f_{q,k}} \cdot \frac{\sum_{i=1}^m \frac{y_{i,k}}{E_{\mathbf{y}_i}}}{\sum_{i=1}^m \frac{y_{i,0}}{E_{\mathbf{y}_i}}} \quad (5)$$

3.3. Reconfiguration

Let us consider that the visual descriptors are classified into L different categories C_1, C_2, \dots, C_L . For each feature element (visual descriptor), we assign an importance degree of relevance,

say w_i . All weights which are assigned to a particular category C_j , are included in a weight vector denoted as \mathbf{w}_{c_j} .

In order to perform system reconfiguration, we initially estimate the user's preferences. This is accomplished by calculating the energy of the weights of a particular category. More specifically, the energy of the weight vector \mathbf{w}_{c_j} is estimated as follows

$$E_{\mathbf{w}_{c_j}} = \frac{\sum_{i \in C_j} w_i^2}{\|C_j\|} \quad (6)$$

where $\|C_j\|$ denotes the cardinality of the set C_j , and w_i all the weight belonging to the category C_j . Then, the algorithm estimates all the categories C_j whose respective energy contributes more (defined by a factor α) to the total energy, i.e., $\hat{j} : \sum_j E_{\mathbf{w}_{c_j}} = \alpha \cdot E$ (7)

where E refers to the total energy, i.e., $E = \sum_j E_{\mathbf{w}_{c_j}}$, while factor α usually equals to 80%.

Using equation (7), we estimate the categories that almost approximate the total descriptor energy, or in other words we estimate those categories which correspond to the most important descriptors with respect to the current user's information needs and preferences as expressed by the relevance feedback algorithm. This means that the descriptors of the selected categories are the most important to the user's profile and therefore, these descriptors should be analyzed in more details in the following retrieval iteration.

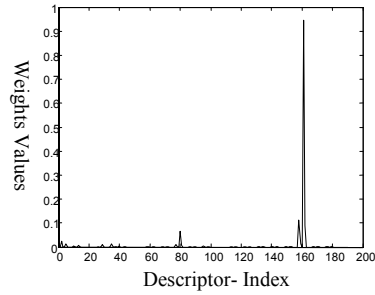


Figure 2: The weight values after a feedback iteration. As is observed only some weights are important (high weight values) in this particular example.

4. A TYPICAL SCENARIO

In this section, we describe a typical scenario of the proposed reconfigurable architecture. Initially, the user submits an image query (query by example), which is received by the web server (portal component of the acquisition module). The image query is analyzed by the visual descriptor component to extract the default descriptors. In the following, the retrieval module is activated and a mobile agent is sent to remote web sites for mining and retrieving visual content that is perceived close to the submitted query. In particular, the polling manager of the home platform unit is triggered followed by the mobile agent and the transaction manager. The latter monitors the mobile agent's itinerary.

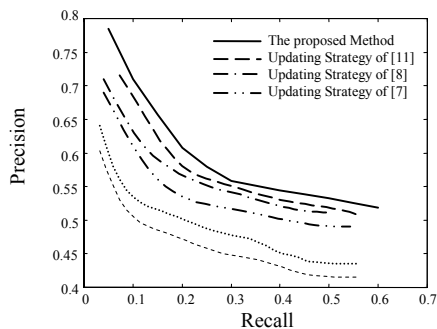


Figure 3: The precision recall curve of the proposed user centric multimedia description along with other dynamic updating mechanism.

The retrieved results are displayed to the user, who interacts to the process by selecting a set of relevant and/or irrelevant data. Then, the on-line learning algorithm, described in section 4.1, is activated to identify the user's profile (most important descriptors). In the following the descriptor categories, which are closer to the user's information needs are selected and then, the system is reconfigured to satisfy the current user's needs. The descriptors of the important categories are analyzed in more details by supporting additional processing algorithms. Figure 2 presents the degree of importance (weights) of the descriptors as they have been estimated by the algorithm of section of 3.2, using a set of relevant / irrelevant images. As is observed, the weights of some descriptors present much higher values. The descriptor categories associated with the greatest weight values are analyzed in more details using additional descriptors of the same category. Therefore, a *multi resolution* reconfigurable scheme is presented. The polling agent of the acquisition module is triggered to commence a new retrieval iteration phase.

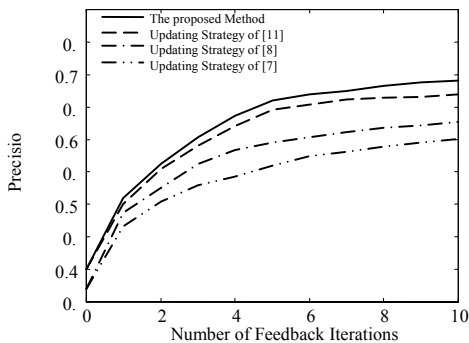


Figure 4: Precision versus the number of feedback iterations of the proposed scheme along with other dynamic updating mechanism.

Figure 3 presents the precision-recall curve of the proposed method along with other dynamic updating strategies. As is observed, the proposed reconfigurable scheme outperforms the examined ones (higher precision at all recall values). This is due to the fact that in the proposed approach adaptation is achieved using a user centric approach instead of just updating the descriptors weights (degree of importance) as the conventional approaches do. Figure 4 depicts the precision versus the number of feedback iterations, where similar conclusions are drawn.

Different Algorithms	ANMRR
The proposed Algorithm	0.09
The Method of [11]	0.11
The Method of [8]	0.12
The Method of [7]	0.14

Table I: The performance of the proposed reconfigurable architecture compared to several relevance feedback schemes in terms of the Average Normalized Modified Retrieval Rank (ANMRR) criterion.

Table I shows a comparison of the proposed method with several updating strategies. The evaluation has been performed using the Average Normalized Modified Retrieval Rank (ANMRR) as proposed by the MPEG-7 standard (low ANMRR means good performance). As is observed from Table I, the proposed methods presents the best results among the compared algorithms, since in the presented system the code used for data retrieval is modified instead of the similarity measure weights as happens in the other approaches.

5. REFERENCES

- [1] O. Etzioni, "The World-Wide Web: quagmire or gold mine?," *Communications of the ACM*, Vol. 39, No. 11, pp. 65-68, 1996.
- [2] Special Issue on Content-Based Image Retrieval Systems, *IEEE Computer Magazine*, Vol. 28, No. 9, 1995. Guest Editors: Venkat N. Gudivada and Jijay V. Raghavan.
- [3] N. Vasconcelos, and A. Lippman, "Statistical Models of Video Structure for Content Analysis and Characterization," *IEEE Trans. on Image Processing*, Vol. 9, No. 1, pp. 3-19, January 2000.
- [4] Rocchio, *Relevance Feedback in Information Retrieval: The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice Hall, 1971.
- [5] I. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos, "Pichunter: Bayesian Relevance Feedback for Image Retrieval," in *Proc. Inter. Conf. Pattern Recognition*, Vol. 3, pp. 362-369, 1996.
- [6] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," *IEEE Trans. CSVT*, Vol. 8, No. 5, pp. 644-655, Sept. 1998.
- [7] Y. Rui, and T.S. Huang, "Optimizing Learning in Image Retrieval," *Proceeding of IEEE int. Conf. on Computer Vision and Pattern Recognition*, Jun. 2000.
- [8] Xiang Sean Zhou, and T. S. Huang, "Small Sample Learning during Multimedia Retrieval using BiasMap", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Hawaii, Dec. 2001.
- [9] A. D. Doulamis, Y. S. Avrithis, N. D. Doulamis and S. D. Kollias, "Interactive Content-Based Retrieval in Video Databases Using Fuzzy Classification and Relevance Feedback," *Proc. of the IEEE International Conference on Multimedia Computing and Systems (ICMCS)*, Vol. 2, pp. 954-958, Florence, Italy, June 1999.
- [10] A. Doulamis, and N. Doulamis, "Generalized Non-Linear Relevance Feedback for Interactive Content-Based And Organization," *IEEE Trans. on Circuits and Systems for Video Technology*, special issue in audiovisual analysis for interactive multimedia services, Vol. 14, No. 5, pp. 656-671, May 2004.
- [11] N. Doulamis, and A. Doulamis "Fuzzy Histograms and Optimal Relevance Feedback for Interactive Content-based Image Retrieval," *IEEE Trans on Image Processing* (under second review for final acceptance).
- [12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGrawHill, 1984.