# User modelling and adaptivity in visual information retrieval systems

*J. M. Torres and A. P. Parkes*
Distributed Multimedia Research Group
Computing Dept.
Lancaster University
{j.torres@lancaster.ac.uk, app@comp.lancs.ac.uk}

## ABSTRACT

This paper proposes suitable characteristics of an ideal Visual Information Retrieval (VIR) system. The central role of user modelling in such a system is discussed. Approaches to the automated analysis of images in image retrieval systems are considered. A preliminary sketch of the model is then provided. The model is based on Bayesian user modelling techniques and dynamic Bayesian Networks. A brief worked example of the model is provided, showing how the model might acquire knowledge about the user and exploit automatically derived image features to help satisfy a user's specific image retrieval requirements.

## 1 Introduction

Large quantities of information are now becoming available in the form of audio, video or image repositories. This paper is concerned with the problems of visual image retrieval (VIR), and in particular the role of user modelling is VIR systems. Research in VIR has hitherto tended to focus more on VIR system-level development, in particular image analysis tools and low-level image descriptors (Bimbo, 1999). However, there is a growing recognition of the need to consider the representation of the user in VIR systems.

The rest of the paper is as follows. Section 2 discusses a hypothetical ideal VIR system, and presents an intuitive discussion of the requirements of such a system. The central role of user modelling in such a system is discussed, and an outline of current approaches to image processing in VIR systems is provided. Section 3 provides a preliminary sketch of a Bayesian-oriented model to support the user's interaction with the VIR system. Following this, a worked example of the model is given, which suggests how the model may build up information about a user in two overall respects. Firstly, the acquisition of knowledge about the user is described. Secondly, it is suggested how an approach that combines user specification of objects of interest in a retrieved set of images with a feature extraction system might be used to retrieve images that are similar, *in the user's view*, to previously retrieved images.

## 2 The ideal VIR system

The essentials of a VIR systems are three main operations: *Query formulation*, *Search*, and *Information Provision*. In this paper, we consider the first and third of these, for which we have identified the following requirements:

***Query formulation***

- the integration of textual and visual media
- the acquisition of knowledge about the user's specific search requirements
- the provision of interactive assistance in formulating an appropriate query

- the use of other contextual information to increase the specificity of the query

***Information Provision***

- the presentation of information in the most appropriate form to represent the required information content
- the adaptation of the actual presentation of information to suit the particular user

The above requirements suggest a central role for techniques to acquire knowledge about the user and to use that knowledge to guide the query formulation and information presentation components of the system. This is based on *user modelling.* User modelling implies *adaptivity*, in that the VIR system adapts itself to suit the specific user according to knowledge it has acquired about that user.

The second major feature of a VIR system is that it must be able to gain access to the *content* of the information. As the information processed by a VIR is predominantly pictorial, this implies that the system must be able to extract relevant features from the images in order both to determine which images best match the user's query, and then select the most appropriate combination of images to present the user.

The remainder of this section is as follows. User modelling and adaptivity for VIR systems are discussed in more detail. Following this we briefly discuss some of the available techniques for extracting content-based information from digital images.

### 2.1 User modelling and adaptivity in a VIR system

*User modelling* can be defined as the process of acquiring knowledge about a user in order to provide services or information adapted to their specific requirements (McTear, 1993; Kay, 1995). Belkin (1997) and Ellis (1990) describe approaches to user modelling in information retrieval systems.

An *adaptive system* is a system that changes its functionality or interface in order to accommodate the differing needs of the users over time (Benyon et al, 1987).

In a VIR system, user modelling and adaptivity are needed to support the following:

- *adaptation of the queries*
  The user's query may be adapted by the system in order to meet that user's specific needs as identified by the user model.

- *adaptation of the information presentation*
  Determining the suitability of images is based not only on the content of the images, but also on how that content relates to the user's goal and purpose.

### 2.2 Visual information retrieval and presentation

A major focus in image retrieval systems has been on the development of advanced image analysis tools to extract low-level properties of the images to be used during the query and retrieval process. These properties, or *image descriptors,* are normally computed during the indexing stage of the image repository and are typically associated with perceptually meaningful features such as *colour*, *shape* and *texture*. This approach allows a search by image content and has been used in several VIR systems (Rui, 1999; Niblack, 1993; Eakins, 1999).

A common approach is to use a start image as an example in the retrieval process. Following this first step, the user successively refines the query in an iterative browsing process until finding the desired image(s).

Several systems incorporate *relevance feedback*. These systems automatically create new queries based on sample images identified as relevant by the user in previous queries (Rui, 1998).

A *learning* approach is found in some systems. For example, in the FourEyes system, similarity and attribute extraction models are selected and combined according to information provided by examples gathered from the users (Minka, 1996). Similarity is also used in the PicHunter system (Cox *et al.*, 2000)

# 3 Towards a user model for VIR

This section sketches a formal model that represents an attempt to address the issues of user modelling, adaptivity and multimodality discussed above. The model is based on Bayesian User Models and Bayesian Networks (Horvitz, 1998). First, the model is described. Following this, a simple worked example is presented to illustrate how the proposed model might be applied in a VIR system.

## 3.1 Sketch of a user modelling approach for VIR systems

The problem domain of Visual Information Retrieval consists of the following objects:

- A User (*U*), typically a human who has the goal of retrieving a non empty set of images;
- The set of images presented in the repository $I = \{i_1, ...., i_n\}$;
- A subset, *S*i, of *I*, containing images that totally or partially satisfy the user's goal;

To enable the system to compute the solution to the VIR problem, i.e., derive the set $S_i$, it must have access to all of the required information. We represent this information as the following function:

$$f: (U, I) \rightarrow S_i$$

It can be seen from the above that the set $S_i$ computed by the VIR system depends on both the user and on the repository.

The information about the user can be viewed as:

$$U = (U_{goals}, U_{actions}, U_{profile}, U_{context})$$

When using the system, the user may have certain specific goals ($U_{goals}$) concerning the images to be retrieved. The user attempts to reach these goals through a set of interactions with the system ($U_{actions}$). User characteristics such as general interests, cultural information, and so on, is held in $U_{profile}$. Finally, $U_{context}$ features contextual information that may be relevant to result $S_i$.

*Temporal Dependencies*

The arguments of the function *f* evolve over time, so a mechanism is required to deal with this. The temporal variability of *I* can be relaxed, i.e., during a visual information retrieval session one can consider that the set of images remains constant.

The independent variables $U_{goals}$, $U_{profile}$ and $U_{context}$, can be seen as relevant time-dependent information. The variable $U_{actions}$ can be considered to be an array of timestamped actions.

*Dynamic Graphical Model*

The actions performed by the user, $U_{actions}$, are perceived by the VIR system as a temporal series of events, $E_{ti}$. These events are the way in which the user demonstrates his or her goals to the VIR system. We therefore establish a dependency relation between $U_{goals}$ at time $t_i$ and event $E_{ti}$.

The user goals also depend on $U_{profile}$ and $U_{context}$. In general, a goal, or set of goals, arise on the basis of a user requirement, and are related to problems revealed by the situational context. How a user maps these problems and needs onto goals also depends on that user's psychological profile. Given these considerations, a further dependent relation can be established from the pair $U_{context}$, $U_{profile}$ to $U_{goals}$ at time $t_i$.

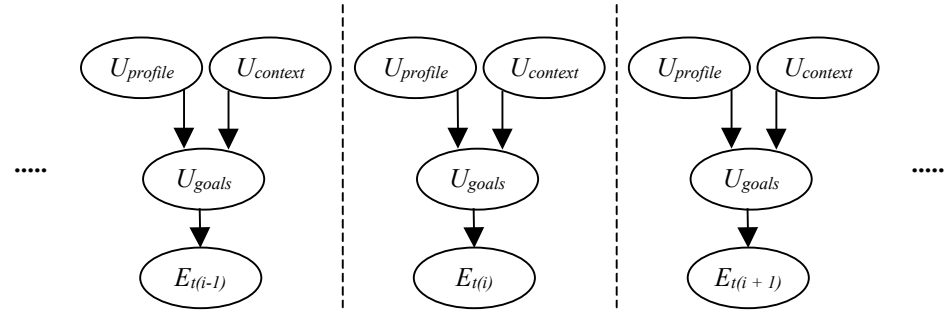


**Figure 1 - Dynamic Belief Network representing the interaction between the User and the VIR system**

Statistical reasoning is a possible approach to modelling the user in the VIR system. In such an approach, each element of *P(I),* i.e. each possible set of images, has an associated value representing the probability that the set of images satisfies the goal of the user. This probability value is updated as time passes on the basis of the interaction between the user and the VIR system. The VIR system uses all possible information gathered from the user and determines which element $S_i$ of set *P(I)* satisfies the user's goal. This can be an iterative process the termination of which can be determined either by the user, declaring that $S_i$ is valid, or the system, which has computed a probability for $S_i$ that is greater than a preset threshold (set in conjunction with the user). A graphical representation of the VIR framework proposed is presented in Figure 1.

Graphical models that can be used to represent the type of problems considered here, namely, Dynamic Bayes Nets (DBNs), can be used to represent random variables that evolve over time. DBNs allow the state of the system to be represented as a set of hidden and observed states in terms of state variables, among which there can be complex interdependencies. Graphical models are graphs in which the nodes represent random variables and the arcs represent direct dependencies between these random variables. When the arcs are directed, such graphical models are called *Bayesian Networks* or *Belief Networks.* In such networks, each node is associated with a conditional probability distribution. The graphical structure provides a convenient means of specifying the conditional independencies, and hence represents a compact parameterisation of the model.

For VIR, given the observed user actions and using a prior graphical model, the objective is to infer the user's goal and ultimately to calculate $S_i$, i.e., the set of target images. The User's goal is a time-dependent hidden variable (an array of hidden variables) of the model that the system must compute in order to compute $S_i$. The other two time-dependent hidden variables $U_{profile}$ and $U_{context}$ are also relevant in the determination of $S_i$, and ultimately will be inferred from the action sequence of the user. In generic terms, the system observes $U_{actions}$ and attempts to compute:

$$P(U_{goals}, U_{profile}, U_{context} | U_{actions})$$

Following the above inference, is is then possible to compute $S_i$:

$$f: ((U_{goals}, U_{actions}, U_{profile}, U_{context}), I) \rightarrow S_i$$

Bayesian models heve been used to diagnose user needs in several applications, enabling those applications to incorporate user modelling capabilities (Horvitz, 1998).

Finally, the system should be designed so that it gathers as much information as possible during interaction with the user. This is a strong reason for the adoption of multimodal interfaces in VIR systems. The presented framework is intuitively adaptive, since it acquires information about the user to aid in the satisfaction of the user's goals. Though the image of the function *f* has been specified above as a set of target images $S_i$, this could be extended to deal with a set of personalised styles of interaction and interaction modalities $P_i$:

$$f: ((U_{goals}, U_{actions}, U_{profile}, U_{context}), I) \rightarrow (S_i, P_i)$$

The above parameters can be used to determine such features as the ways in which the information is presented to the user, the user's preferred interaction modalities, the user's preferred query formulation styles and other adaptive facilities.

### 3.2 The application of the model: a worked example.

A key problem with a model such as the one proposed above is to represent $U_{goals}$, $U_{profile}$, $U_{context}$, and $U_{actions}$ in a stochastic world (to determine the domain of each random variable, empirical studies have to be performed). Further difficulties arises in creating and updating the Conditional Probability Table (CPT) at each node Our initial approach is to assume that for $U_{goals}$, $U_{profile}$, and $U_{context}$, some of the information must be inferred (on the basis of statistical reasoning) and the rest is to be explicitly gathered. Following this approach, $U_{goals}$ can be specified as follows:

$U_{goals}$ = *Inferred User Goals from the Bayesian Networks + Explicit User Goals from the query formulation dialogue*

In order to suggest how the above model may be applied, we now present a simple worked example. The example user is carrying out research into the history of the Kings of England in the 19th century, and requires images to illustrate his work. The following is a description of the relevant features that might be found in the model.

The first, very important, step is to determine the interaction space of the user, i.e., the types of user actions that are possible, such as conversational acts, providing profile information in a form, selecting a subset of a displayed set of images, and so on. Note that in a more restricted interface, the interaction space may be the typical WIMP menu-based scenario. In this case, the system is unable to unobtrusively develop a knowledge-level user model (e.g., to infer that the user is a student in history, and so on). The system is likely to need to request the user to explicitly provide such information.

In systems with a rich dialogue style, the system might infer that our user was a history student from keywords extracted from a discussion with the user (e.g. {history, university, teacher,…}), i.e.

*Kwords{history, university, teacher}-> $U_{profile}$(history_student)*

During the same conversation, the user might use the keywords {19th century, King, England, thesis….}, and the system could infer:

*Kwords{XIX century, King,…}->$U_{context}$(thesis_about_english_monarchy_of_19th_century)*

The user might later refer to the keywords {fight, duel, horse, knight} and the system could infer the goal:

*Kwords{fight, duel, horse, knight}->$U_{goal}$(looking_for_images_featuring_duels_of_mounted_knights)*

The above refers to the *knowledge* level of the proposed user model. Lower level features would also need to be represented in the model. For the purpose of this paper, some of the more interesting of these lower level features will be properties of images used by the system to create content-based associations between images that take account of the user's conceptualisation of these similarities. For example, for our history user, the horses in the images are of interest only because they are ridden by 19th century knights. One approach would be carry out feature extraction on the images to enable the user to indicate the objects of interest (in our example, the horse and the rider, say). The system creates a combined specification of the features related to all of the objects in the image that the user has indicated. This specification is then used by the system as part of its evidence when subsequently retrieving related images in that session. An approach based on the user's notion of similarity can be found in the PicHunter system (Cox *et al.*, 2000)

## 4 Conclusions

In this document we have discussed the need for user modelling and adaptivity in effective VIR systems. We have sketched out a model based on Bayesian networks and Bayesian user modelling, and have demonstrated, using a worked example, how such a model may be applied in a VIR. The example showed how various components of the user model are populated as knowledge of the user is acquired through

dialogue with the user and information about the user's actions. We have suggested how image feature analysis might be combined with user specification of areas of interest within an image in order to determine similarity between images that is specific to the goals and profile of the particular user. The next stage in our work is to prototypically implement the model in a simple VIR system, This will enable us to study the effectiveness of the model over a range of users engaged in the VIR process.

## References


Belkin N. J. (1997): *User Modelling in Information Retrieval*, Tutorial presented in Sixth International Conference on User Modelling, Chia Laguna, Sardinia, June 1997 (url: www.um.org/um_97/contents.html).

Benyon D. et al. (1987): *Modelling users' cognitive abilities in an adaptive system*, 5th Symposium EFISS, Plenum Publishing.

Bimbo A. del (1999): *Visual Information Retrieval*, Morgan Kaufmann Publishers, Inc. San Francisco, California, 1999.

Cox I.J., Miller M.L., Mink, T.P. Papathomas T.V. & Yianilos P.N. (2000) The Bayesian Image Retrieval System PicHunter: Theory , Implementation and Psychophysical Experiments. To Appear in IEE Transactions on Image Processing, VOL. 20, 2000

Eakins J. P.; Graham, M.E. (1999): *Content-based Image Retrieval - A report to the JISC Technology Applications Programme*, Institute for Image Data Research, University of Northumbria at Newcastle, January 1999 (www.unn.ac.uk/iidr/research/cbir/report.html).

Ellis D. (1990): *New Horizons in Information Retrieval*, The Library Association, London, 1990.

Horvitz E. J. Breese D. Heckerman, D. Hovel D. & Rommelse K. (1998): *The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users.* Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, July 1998.

Kay J. (1995): *Vive la difference! Individualised interaction with users*, IJCAI'95, Montreal, August, 1995.

McTear M. (1993): User modelling for adaptive computer systems: a survey of recent developments, Artificial intelligence review 7, 157-184.

Minka T (1996): An Image Database Browser that Learns From User Interaction, MIT Media Laboratory, Cambridge MA, Master of Engineering Thesis, technical report TR #365.

Niblack W. Barber R., Equitz W., Flickner M., Glassman E., Petkovic D. & Yanker P. (1993): *The QBIC Project: Querying images by content using colour, texture and shape*, in SPIE 1908, Storage and Retrieval for Image and Video Databases, February.

Rui Y., Huang T.S. & Chang S.F (1999): *Image Retrieval: Current Techniques, Promising Directions and Open Issues*, Journal of Visual Communication and Image Representation, Vol. 10, March 1999, pp. 39-62.

Rui Y., Huang T.S., Ortega M. & Mehrotra S. (1998): *Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 5, September 1998, pp. 644-655.