Towards Adaptive Ontology-Based Image Retrieval

Johanna Vompras

Institute of Computer Science
Heinrich-Heine-University Düsseldorf
D-40225 Düsseldorf, Germany
vompras@cs.uni-duesseldorf.de

Abstract

Since the use of large image databases gains in importance nowadays, efficient querying and browsing through image repositories becomes increasingly essential. Compared to text retrieval techniques there are even more problems with image retrieval. Particularly, the *semantic gap* between low-level visual features of images and high-level human perception of inferred semantic contents decreases the performance of traditional content-based image retrieval systems. The first important step for the correlation of image data with cognitive processes is the identification of discriminative features in the data. The next decisive step is to extract high-level knowledge from this data in order to provide a confident interpretation of signals into symbols. In this paper we demonstrate our first conceptual notions about the integration of spatial context and semantic concepts into the feature extraction and retrieval process using the relevance feedback procedure.

1 Introduction

Today, the management, storage, and retrieval of large image data repositories is a challenging problem. Due to their properties images are hard to handle with, particularly because of their diversity and the inadequate possibility to estimate and represent perceptional semantics. The basic difficulties emerging in analyzing and retrieval of image data often result from the following facts: Images mostly contain a non-uniform image background or a textual noise in the foreground, such as phone numbers or URLs, which make it impractical to differentiate between the relevant or irrelevant patterns. There are other problems emerging due to the data quality, like sharpness, contrast, and brightness of images. The heterogeneity of image data is also an issue to address to. The contents of the images may range from grey-scale to 24-bit color images and may have different sizes and formats.

Beside all these problems, the most important disruptive factor in image retrieval is the *lacking semantics*. Images are not structured data, which can be browsed on certain patterns which are arranged in a predefined manner. Views and shots can be taken from a variety of camera positions or images may contain many types of objects, like persons or animals. The objects may differ from each other at the visual feature level despite of belonging to the same semantic category.

The known traditional retrieval techniques based on textual image annotations ignore these difficulties. Here, the images are first labeled manually by keywords and are retrieved by their corresponding annotations [1]. Due to the high manual effort and the inconsistency in image interpretation and keyword assignment among indexers, this approach becomes impracticable [2].

Efficient and accurate information retrieval is of great demand in the field of image databases. A first important pre-processing step in the image mining process is the feature selection and extraction which have been the basis for numerous heuristic retrieval methods and machine-learning methods with relevance feedback in content-based image retrieval (CBIR).

Nevertheless, users are highly interested in querying images at a conceptual and semantic level, not only in terms of features like color, texture, or shape [3]. Since it is very difficult to involve the human perception to an automated image retrieval task, methods which combine human cognition ability and the automatic computation are of great interest [4].

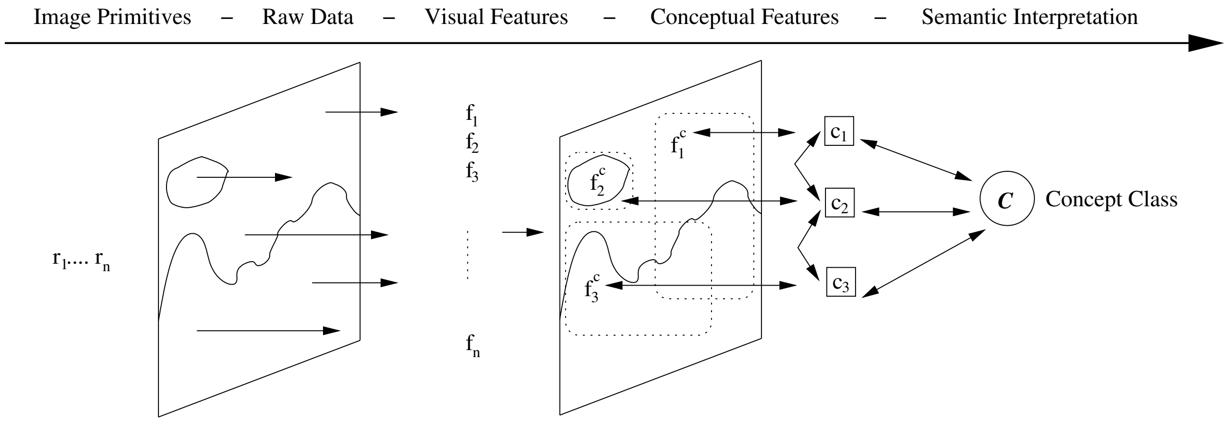

Figure 1: Levels of Image Content Representation

In this paper we demonstrate our first conceptual notions about the integration of spatial context and semantic concepts into the feature extraction and retrieval process using the relevance feedback procedure. Through an interactive system which allows both feeding the system with additional semantic information and capturing the correspondences between the low-level features and high-level *perceptual semantics*, the quality of retrieval results could be improved. Thus, query formulations at the semantic level will enhance current information systems.

2 Data Representation

In order to achieve our above-mentioned aims, an image and semantic data model has to be developed which can be coupled for retrieval purposes. Thus, a modified multimedia object model from [2] is used for representing low-level features and high-level perceptual semantics. Accordingly, an image object O is represented as

$$O = O(D, F, R, C, S) \tag{1}$$

where D represents the raw image data, $F = \{f_i\}$ a set of low-level visual features and $R = \{r_{ij}\}$ a set of representations for a given feature f_i. The model introduced in [2] is extended by the component C which denotes the set of semantic concepts $\{c_k\}$ and their inter-concept relationships $S = \{s_{kl}\}$. A given concept is also described by a set of features f_i^c which are characteristic for it in the low-level space. In addition, a semantic feature space is built with references to image objects or image regions in the database. The semantic network is represented by a set of keywords having links to the image objects or regions in the images. The initial network can be formed during the extraction of images from the web or digital television by using textual information. More concepts can be learned from the user's feedback and thus expand the semantic network.

In order to support a multiple-level description of image contents, weights are used at various levels. The aim of relevance feedback is not only the optimization of the weights to model the users information need but also modeling high-level data and thus forming a semantic space.

2.1 Image Representation

Since there is no semantic information in the data at the beginning of the retrieval session, the contents of image objects are only characterized by their visual features f_i which result from the computation of the image processing primitives $r_1 \ldots r_n$.

2.2 Semantic Representation

An ontology gives an explicit specification of an abstract view of the world, which is represented by inter-relationships among a set of representational terms that we call *concepts*. Let us suppose that our world can be described by the set $\mathcal{K}$ of structured concepts which are hierarchically organized. Our real-world model is organized as a net of concepts linked together and providing a formal description of the relationships of the concepts. Since the description of the whole world is relatively voluminous, it is desirable to use a simplified specification for a particular application field which reveals the correspondences between low-level features and semantic knowledge.

In the semantic feature space, each image $\mathcal{I}$ includes a set of n different concepts which are represented by a set of m characteristics $p_1 \ldots p_m$ and their weights w_{mkl}. Each characteristic is represented by one relationship between different concepts as displayed in Table 1 and the corresponding weights. From this it follows that the spatial and semantic relationships $rel(c_k, c_l)$ between two concepts c_k and c_l are values in a m-dimensional vector in the semantic space, where m is the overall number of existing relations between concepts. Thus, the semantic content of an image object is represented by a vector $\vec{v}$

$$\vec{v} = \begin{pmatrix} c_1 \\ c_2 \\ .. \\ .. \\ c_n \end{pmatrix} \text{ and } rel(c_k, c_l) = \begin{pmatrix} w_{1kl} \\ w_{2kl} \\ .. \\ .. \\ w_{mkl} \end{pmatrix}$$

where its components c_k are represented by:

$$c_k = \begin{cases} 1 & : \text{ if object is classified to concept } k \\ 0 & : \text{ else.} \end{cases}$$

The learning mechanism must be able to learn the classification of image objects into concepts, and thus deliver an appropriate answer to a query. An image region belongs to a particular concept c with its *partial concepts* or *subconcepts* $\overline{c}_1 \ldots \overline{c}_n$ if

$$\sum_{1}^{n} w_i \overline{c}_k > \Theta \tag{2}$$

where Θ is a application specific threshold and the w_k represent the weights assigned to the subconcepts $\overline{c}_k$ within the concept c. The weights of the subconcepts are adapted during the learning process. For example, if we consider the concept *tennis match* which is characterized by description *'a match between two tennis players'* in a ontology: This concept has a `has-part` relationship with the terms *tennis ball*, *tennis players*, *tennis court*, and *green grass*. Through the relevance-feedback by the user the weights of the subconcepts are adjusted to the users perception and thus mirror the meaningfulness of the subconcepts in the whole context. For example, if there are also sand-colored courts in our world, the user will specify a lower weight to the component *green grass*.

The membership of the over-all image contents to a *concept class* Ω is computed by checking if its containing concepts c_k are located in immediate neighborhood N of the class Ω. The whole image is associated with class Ω if

$$N(\Omega, r) := \{c_k \in C \mid \forall k \ \mathrm{dist}(\Omega, c_k) \leq r\} \tag{3}$$

where r denotes the radius of the semantic neighborhood of the concept class Ω. The distance between two concepts in the ontology is measured along the shortest path, which is formed by the smallest number of undirected arcs that connect the entity classes. These undirected arcs represent semantic relationship between two concepts (e.g `is-a-kind-of` and the `has-part`). A concept class Ω denotes a superordinate concept which describes the aggregation of the subconcepts c_k in this special context is assigned to a certain image according to the semantic hierarchy or defined by the user during the relevance feedback procedure.

In order to link semantic knowledge of our world into the retrieval process, an appropriate representation supporting inter-conceptual and intra-conceptual modifications of weights and a suitable similarity function is needed.

3 Relevance Feedback Method in Retrieval

A decisive step in the embedding of semantic information in the retrieval process is the mapping of low-level feature space into a high-level semantic space. Every detected object in the image is an additional information of our world. Through this information the users' *semantic space* is formed out of the recognized concepts and their semantic relationships are updated in each retrieval step, so that this

Spatial and Topological Relations	Lexical and Semantic Relations
a `left-of` b	a `is-a-kind-of` c
a `right-of` b	a `has-part` c
a `above` b	a `is-synonym-for` c
a `below` b	a `is-antonym-for` c
a `near-to` b	a `is-member-of` c
a `far-to` b	
a `inside` b	
a `outside` b	
a `in-front-of` b	

Table 1: Spatial and semantic relations between objects in an image.

created knowledge is used in subsequent query sessions. The relationships between images are also captured in the semantic space. After several steps of relevance feedback the user has determined a pool of images which are relevant to his query [5]. We can make the assumption that these images belong to the same semantic class. Due to this result set of images, the subsequent query object O^q is extended by additional query points $q_1 \dots q_n$ in the feature space that are related with the same semantic class.

3.1 Construction of semantic feature space

Since the semantic description of many fields of application is infinite in size, the semantic space is formed out of a partial mapping of real world objects (ontology space) into a semantic feature space. This space should summarize the most important concepts and relations between them discovered by humans. In analogy to the *cognitive space* introduced in [6] it is in constant change as a result of internal processes and interaction with the user. The relevance feedback procedure is a user-centered loop [2] with the following iteration steps:

1. Initialization of the low-level representation's weights and the semantic representation's weights into a set of no-bias weights.
2. User query in form of a labeled sample image or formal description of low-level features. The query object O^q is described by its features f_i^q.
3. The similarity of the query object O^q to the image objects is computed iterative over all representation levels according to the corresponding similarity measure.
4. The returned k image objects are ordered by their similarity $d(.)$ to O^q.
5. User labels a subset of the retrieved objects according to his perception subjectivity.
6. Weights are updated according to the user's feedback such that the adjusted O^q is a better approximation to the user's information need.
7. The semantic space is extended such that the recognized concepts are inserted into the space hierarchy and the positive labeled image objects results in a incremental movement of concept points in the semantic space.
8. Go to 2 and start a new iteration of retrieval.

3.2 Similarity Measure

The similarity function may deliver a suitable distance measure for both low-level image features and users subjective semantic features. In step 4 of the feedback loop (Section 3.1) we use a distance metric function d to measure the similarity between the query vector O^q and the stored image objects. The query vector O^q is formed either by an image object or rather by its visual features like in the *query-by-example approach* or as a combination of an image object with semantic description. As all the weights in the system, the similarity measure is also a dynamic function depending on the current state of the system i.e. weights and semantic information. Generally speaking, the distance between an query object O^q and an image object O in the database is defined by

$$d(O^q, O) = \frac{\alpha\, d_l(O^q, O) + \beta\, d_s(O^q, O)}{\alpha + \beta} \tag{4}$$

where α and β are from range $[0 \ldots 1]$. They denote the degree of belief in the values computed by d_l and d_s which represent distance functions for both low-level and semantic representation of the query object O^q. The parameters α and β are determined by counting the number of query points in the semantic space which are close to the point of the query object. If the number is small the similarity computation is rather based on low-level features.

4 Future Work

CBIR is an important technology for an effective and efficient retrieval in image databases. Since it is difficult to capture high-level semantics of images when images are only represented and queried by low-level features, methods which embed learning semantic concepts are of great demand. The weighting of low-level features and high-level semantics both in the data representation and similarity computation assists in finding a set of images which should be close to the association of what users are looking for. Of course, there are some questions left:

- How can the accuracy of different similarity functions be refined in order to be optimized the precision of the query result?
- How can semantic feature representation enhance the detection of objects by inference mechanisms?
- How can the model be used to deliver a graphical framework for user interaction and knowledge extension?
- How can the semantic model be compressed and summarized in order to improve the retrieval performance?

References

[1] G. Salton and M. McGill. *Introduction to modern Information Retrieval.* New York.

[2] T. Huang, Y. Rui, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998.

[3] J. Torres, A. Parkes, and L. Corte-Real. Region-based relevance feedback in concept-based image retrieval. *Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisboa, Portugal.*, 2004.

[4] A. B. Benitez and J.R. Smith. New frontiers for intelligent content-based retrieval. *Proc. SPIE Vol. 4315, Storage and Retrieval for Media Databases, 141-152*, 2000.

[5] X. He, O. King, W.-Y. Ma, M. Li, and H.-J. Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, 39-48.*, 2003.

[6] G. B. Newby. Cognitive space and information space. *Journal of the American Society of Information Science and Technology. 52, 1026-1048*, 2001.