Can Rule-Based Indexing Support Concept-Based Multimedia Retrieval in Digital Libraries? Some Experimental Results*

Ulrich Thiel André Everts Barbara Lutes
Adelheit Stein
GMD – German National Research Center for Information Technology,
Integrated Publication and Information Systems Institute (IPSI)
D-64293 Darmstadt, Germany

Abstract

The emerging conception of a "digital library" as an information environment in which distributed information sources are made available in an integrated way implies new requirements for indexing and retrieval methods. Therefore, one important research topic in the field of digital libraries is the design of task-specific methods for very precise searching. In addition, we have to anticipate highly dynamic information bases which change continually. This paper outlines an approach which addresses both problems. Rule-based indexing can be used to support highly precise searches in dynamic information bases. The method enables users to perform conceptual searches in image collections, thus going far beyond contemporary content-based retrieval methods for pictorial material. The paper concludes with a discussion of the empirical results obtained so far.

1 Retrieval Support in Digital Libraries: Specific Demands

Rather than reducing the notion of a "Digital Library" (DL) system to the technology needed to handle a large collection of electronic documents, more holistic approaches consider the problems encountered in the development of "information-enriched environments" (cf. Duguid and Atkins 1997). Once this broader perspective is adopted, a variety of issues arises which has to be addressed in order to ensure the usability of this environment for professional and recreational purposes. For instance, the DL will be accessible to different users with varying needs and backgrounds; it will need to provide access to collections containing items of probably very diverse content.

The design of DL access functionality must reflect this diversity in order to consider both the characteristic properties of the collection –which may change over time– and the information needs of a given user community. The perspective we adopt stresses the fact that "information" is not simply static data or a set of documents, but the outcome of a context-dependent communication process involving humans, documents, and the technological as well as semantic environment. This was the main motivation for many research efforts in the area of "intelligent IR". Elsewhere (cf. Belkin et al. 1995, Thiel et al. 1996, Stein et al. 1999) we have proposed to tackle this problem in a two-tiered approach: While a logic-based retrieval method should replace simple matching procedures by inferences which take into account domain-specific rules, thesauri, and other knowledge sources, the query processing should be integrated into a flexible human-machine dialogue, allowing for interactive query construction and modification, inspection of interpretations found for ambiguous queries, and convenient exploration of result lists (cf. Thiel et al. 1996, Stein et al. 1997). Most efforts in the area of logic based IR addressed textual documents, or the textual parts of multimedia documents. In this paper, we will present an approach which aims at exploiting results of image analysis procedures. These techniques usually are the basis of "content-based retrieval" (CBR) approaches which support similarity-based searching of images.

*Some of the work described here was funded by the EU Commission in the ESPRIT Basic Research Project No 9141 "HERMES" (High Performance Multimedia Information Management Systems).

The advantage of taking the usage situation into account and embedding the database accesses into an information-seeking dialogue is the opportunity to support more realistic definitions of "relevance" (Mizzaro 1997). In addition to topicality, a user's relevance criteria (cf., e.g., Froehlich 1994, Green 1995) include usefulness, novelty, originality, origin, type, and style of the document (this is not an exhaustive enumeration) as well as its factual content or subject matter. A user's judgement of relevance is also influenced by a variety of situational parameters, such as previous knowledge, task at hand, and goals (cf., e.g., Wilson 1973).

In the case of non-text media such as images, which are of increasing importance in digital libraries, the needs of many users are only, if at all, specifiable on the conceptual level. Practical experience has shown (Lutes et al. 1996) that the content of non-text document collections can be successfully described via semantically-oriented metadata, for example, by indexing a picture's visible objects, situations, and interpretation context. During search both such abstract concepts and concrete features such as color, contrasts, etc. can be used. For this reason, multimedia retrieval should be possible on both the conceptual and the feature level.

In the remainder of this article we first outline our approach to concept-based image retrieval. Subsequently our rule-based system for the automatic indexing of images is presented along with initial results of experimentation. A quantitative and qualitative discussion thereof as well as future research plans may be found in the final sections.

2 Image Retrieval Based on Feature Analysis and Conceptual Reasoning

In general, a user query can be regarded as a conceptual description of the objects the user is looking for to solve her information problems. As such it is unlikely to match database entries directly. This often leads to repeated modifications of the description. We distinguish two levels on which this may happen: the intensional (or: conceptual) representation of the domain, and the extensional model (i.e. instances retrieved from the database) of an inferred query interpretation. Methods such as "relevance feedback" take an extensional view; concept analysis, on the other hand, is a modification method resulting from the intensional view: if more than one interpretation of a query is possible in a given context, these can be discovered and presented to the user for selection and/or modification. One logic-based method of interpreting user queries in light of a given theory is abduction. Elsewhere, we devise a method for concept retrieval in multimedia databases based on abductive reasoning (see Müller and Thiel 1994, Thiel et al. 1996, Müller and Everts 1997).

Abductive reasoning based on probabilistic propositional logic is formally equivalent to Bayesian networks which can be used to implement probabilistic inference engines for the purpose of IR (as applied, for example, in the INQUERY system, cf. Turtle and Croft 1991, Callan et al. 1992). In accordance with van Rijsbergen's formulation of the retrieval problem (van Rijsbergen 1989), abductive reasoning attempts to prove that a database entry D entails (a part of) the query Q: $D \rightarrow Q$. The abductive reasoning procedure yields a set of additional assumptions (called hypotheses), which are necessary for deriving evidence that D entails Q.

In the case of conceptual image retrieval, we assume the queries to be expressions in propositional logic, e.g., "street scene $\wedge$ daylight". Some intermediate steps may be required which employ heuristic retrieval rules, like "dimension: 3D $\wedge$ objects: artificial $\wedge$ source of light: natural $\rightarrow$ street scene". These rules can be provided by experienced searchers who know the collection and can anticipate common requests of end users. Contemporary picture archives rely on manual annotations or indexing, cf. Lutes et al. 1996. Hence, the compilation of a rule base, although a considerable investment, may pay off for large picture collections, which may be growing or changing quickly. The rules reduce high-level user concepts to combinations of simpler concepts, called "visual content descriptors", which represent abstract or class features of images, e.g., dimension (2D, 3D), type of light source, type of objects shown (e.g., natural vs. man-made) in the case of photos, but also differentiations between photos, paintings, cartoons, etc. For example, a landscape could be described by the descriptors "light source: natural light", "dimensions: 3D", "objects: natural", as well as by having a large blue area in the upper part of the image (see Section 3.1.1 below for an overview of the classification scheme and descriptors used in our experiments).

The content descriptors must satisfy two conditions: First, they must support the needs of the user community, i.e. it should be possible to express common user concepts in terms of content descriptors. Second, they must correspond to patterns of pictorial features which may be extracted from images in an automatic process. The translation of content descriptors into feature patterns is facilitated by a set of "indexing rules". In section 3.1, we will discuss the

process of generating these indexing rules which map content descriptors onto feature patterns in more detail.

Being able to derive these rules (semi-) automatically whenever major changes in the collection occur –here we assume that a DL is usually supervised– enables us to cope with large and quickly growing collections, which then need only to be processed by feature analysis methods rather than be manually indexed. To this purpose, procedures using bitmap technology and statistics on elementary image features are well-known, and adopted here for our analyses.

In the PiClasso system (Everts 1996) we combined 15 algorithms based on the PBM format (portable bit map) to calculate such characteristics as the color distribution within an image, the surface texture of objects in an image, or values which express the degree of similarity in color distribution between two pictures (see also Brady 1982).

Image analysis values of this type are usually stored in meta-databases, where they can be used as the basis for content-based retrieval. However, in most applications no abstraction or interpretation of these data is performed. User queries must either be formulated in terms of existing value ranges, or users must provide a sample image which is then analyzed in the same way to provide query values (cf. Gevers and Smeulders 1992, Hirata and Kato 1992). Since exact matching is not very useful, often intervals are used. Compression techniques such as wavelets are sometimes used to abstract from overly detailed pixel patterns and to accelerate processing.

To enable the use of image analysis results for *more* than similarity search, they must be semantically interpreted. This can be accomplished through model-based object recognition (e.g., Hermes et al. 1995), which allows the identification of specific objects. The modeling effort is high and only justifiable in special cases, e.g., when the information needs of users can be precisely described a priori (e.g., police tasks, cf. Narasimhalu 1993). In general, however, users of a digital library will search for various concepts, e.g., for objects, persons, events, and various motifs. To a certain degree this can be supported by abductive retrieval. In our approach, we first need to define retrieval rules –as described above–, which can be composed of appropriate content descriptors. Second, we need to derive the indexing rules which associate these content descriptors with feature patterns.

A first experiment to generate indexing rules was based on quantile analysis, a standard statistical procedure (see Müller and Everts 1997 for a detailed description): 300 heterogeneous images were randomly chosen from several larger image collections. They were first manually classified by various persons by assigning descriptors which mainly concern visual image characteristics.

For each of the 15 feature extraction methods provided by PiClasso, the following analysis was performed. Let n be the number of images indexed with a certain descriptor and $X_1 <= \ldots <= X_n$ be the sorted result values of a given image analysis algorithm obtained from these images. The α-Quantile X_α is the value for which $\alpha\%$ of all values are less than or equal to X_α and $(1-\alpha)\%$ of all values are greater than or equal to X_α. (If the calculation of $n * \alpha\%$ does not produce an integer it is rounded accordingly).

The interquantile range (IQR) is defined as those values which lie between $X_{75\%}$ and $X_{25\%}$. We define the fence of a set of values as the range from $X_{25\%} - 1.5 * (X_{75\%} - X_{25\%})$ to $X_{75\%} + 1.5 * (X_{75\%} - X_{25\%})$ (see Müller and Everts 1997 and Thiel et al. 1998a). In order to generate rules from the values, individual values as well as linear combinations of values were subjected to a statistical analysis. If the fence of the values in a learning sample showed acceptable behavior, a rule was generated (e.g., $image(I, contrast : 0..200) \rightarrow contour(I, blur)$ tells us that the contrast values in a certain range are associated with the qualitative judgment "soft contours" by human indexers).

These rules were used in the MIRACLE retrieval system to index images (cf. Müller and Everts 1997). To search, users formulate queries as a combination of descriptors. The rules are interpreted abductively, producing range values for the query in a meta-database (see also Mehtre et al. 1997).

Quantile calculation allows conclusions to be drawn regarding the relationship between scalar values and descriptors. But it is often a combination of values which is really characteristic of an image. Thus, the conditional probability of assigning a descriptor based on a given vector of values must be estimated. In the following, we present a method employing these probabilities in the retrieval process.

3 An Advanced Method for Rule-Based Retrieval of Images

In this section, we outline a method for generating indexing rules that take into account whole feature vectors, instead of scalar feature values. Our approach to enabling conceptual queries on images is divided into the following steps: The image indexing (analysis) system employs a number of feature detection algorithms. The results of these algorithms

- called the feature extraction values - are used to find rules which map the values to conceptual terms. For rule generation we employ an empirical approach in which manually indexed images are used as a training set. Generated rules and extracted feature values are stored in the metadatabase. Note that the latter feature values are not restricted to the original sample set used for the rule generation process. Instead, feature analysis results obtained from the many times larger image collection are now used as the basis for semantic access. If the user poses a conceptual query, the retrieval engine analyses the query and maps it to a set of rules which are requested from the metadatabase. The rules are interpreted by an appropriate rule interpreter yielding specification of features extraction values to be searched for. If feature values matching the constraints can be found, the associated images are retrieved. The result is given back to the user as a ranked list. In the following, these steps are described in more detail.

3.1 Generation of Indexing Rules

As a part of our experiments, a number of feature-extraction and comparison algorithms were implemented (Everts 1996) using the PBM (portable bit map) collection of image processing software. Since texture-based classifications are very effective (up to 100% correct classifications, if they are applied carefully, cf. Picard et al. 1993), the PBM texture module (Haralick et al. 1973) was selected as a promising tool.

For the rule generation, we used the following empirical approach: the starting point was a collection of 300 images manually classified by various persons using a simple web-based interface. The interface was sufficiently self-explanatory to be used without any specific training. Controlled index terms were assigned in a number of domain-independent categories including content of an image, contours of objects, source of light and image source (photo, drawing, graphic, computer-generated images). The images were divided into two disjunct sets, a training set and a test set. We used the training set for rule generation and the test set to validate the rules (see Section 3.2).

3.1.1 The Classification Scheme

In the following, the classification scheme used for the manual indexing is described in more detail.

- **Light and Contour:** The attribute "*light*" describes the source of light in the picture; possible values are "natural" (for sunlight) and "artificial" (for flashlight, lamps and the like). If indexers cannot decide which value is correct or if none of them applies, they may choose "unspecified" (which is a generally available option in all the categories/attributes described below). An object with an unspecified (or nonexistent or unknown) source of light can be, for example, 2D graphics and computer icons.

  The attribute "*shadow*" concerns the shadowing in the picture, where "soft" is, for example, indicated by smeared edges (i.e. the shadow is perceptible only as light grey hue in the picture). "Hard" is a very dark, sharp shadow break, "middle" was introduced as intermediate stage between soft and hard shadow, and "none" means not (or very hardly) perceptible.

  "*Contour*" of main object indicates the perceived sharpness of the contour of the prominent object in a picture. The values range from "blurry" for soft outlines up to "sharp" for clear object contours.

- **Dimension:** The attribute "*dimension*" describes the dimensionality of the *representation* of the main object(s). The value "2D" is for a two-dimensional representation (such as in maps, graphics and line drawings), "3D" for three-dimensional representations (such as in photos of persons, houses and the like).

- **Foreground and Background:** The attribute "*front*" indicates whether the foreground in the picture is a natural object (e.g., a lawn or tree) or an artificial/man made object (e.g., a paved street or part of a building). "*Background*" covers the complementary view, distinguishing between the same values as "front" for the foreground, i.e. "natural" background (such as sky or landscape) and "artificial" (such as a wall in a room).

  "*Object-in-front*" defines differences in the brightness between main object, foreground and background. "Light" is to be assigned when the front/main object is brighter than the background (e.g., a white house in front of a group of trees), whereas "dark" indicates that the object in front is darker than the background (e.g., trees or buildings in front of a light blue sky).

- **Content, Kind and Source:** The attribute called "*content*" refers to the main object(s) and roughly distinguishes between four descriptors: "living" for living objects (persons, animals, etc.), "inanimate" for artificial everyday objects (e.g., houses, furniture), "natural" for natural, but inanimate objects and elements (e.g., sand dunes, water, rocks), and "artifact" for works of art (e.g., sculptures, statues).

  "*Kind*" is another content-related attribute, but defines more fine-grained descriptors for the main object(s): "portrait" (half-length portraits of persons), "human" (full view of individuals), "group" (of humans), "landscape", "building", "physical object" (e.g., cars, furniture, toasters), "detail" (detailed view of a part of an object, e.g., its texture), "icon" (computer icons), and "map".

  "*Source*" describes the source of an image, and we distinguish here between "photo", "computer generated" (images generated by a computer program), "painting", "drawing", and "cartoon" (comic drawings).

In the next section we concentrate on the computation of associations between feature values and index terms specified in our classification scheme.

3.1.2 Association Between Feature Values and Descriptors

The association between feature values and manually assigned index terms, i.e., the descriptors chosen from the classification scheme above, can now be accomplished using algorithmic statistical methods, ranging from exhaustive exploration to complex stochastic computation. Which method is applicable depends on the degree of aggregation applied to the original feature values. We start with p-dimensional vectors $\bar{x}$, containing the results from p different feature extraction methods, applied to the image. In the next step, the feature-extraction values may be aggregated to dynamically built constraints, e.g., ranges, or linear combinations of the feature values. If we regard ranges of scalar aggregated feature values, we can derive plausible rules to describe the content of pictures on a general level, by analyzing the feature-extraction values of the manually classified images by a α-Quantile analysis, as described in Müller and Everts (1997) and Thiel et al. (1998b). In a number of cases, however, no useful association pattern could be found due to the fact that the aggregation process was too coarse-grained. In this situation, it is useful to examine the original data, i.e. the feature vectors. In a first exploration study (Everts 1996) we employed a brute force exhaustive search. The results being promising, we changed to a more efficient method.

In our next experiments, we used the Quadratic Classification Function (QCF) (cf. Bortz 1989, p. 572 ff) to calculate the probability that an image matches a classification item. The QCF gives a measure for the distance of the feature extraction values of an image and the mean values of a set of manually classified images.

Let C be a classification attribute (e.g., source of light), $c_j \in C$ be a descriptor (e.g., source of light: natural), $x_{im}, i = 1 \ldots p$, be p feature-extraction values of an image m, and $\bar{x}_{ij}, i = 1 \ldots p$ be the median of all x_i from the images in a training set, manually classified with the same descriptor c_j. Then, a distance vector d_{jm} is defined as

$$d_{jm} = \begin{pmatrix} d_{1jm} \\ d_{2jm} \\ d_{3jm} \\ \vdots \\ d_{ijm} \\ \vdots \\ d_{pjm} \end{pmatrix} := \begin{pmatrix} \bar{x}_{1j} \\ \bar{x}_{2j} \\ \bar{x}_{3j} \\ \vdots \\ \bar{x}_{ij} \\ \vdots \\ \bar{x}_{pj} \end{pmatrix} - \begin{pmatrix} x_{1m} \\ x_{2m} \\ x_{3m} \\ \vdots \\ x_{im} \\ \vdots \\ x_{pm} \end{pmatrix}.$$

This means that an element d_{ijm} of the vector d_{jm} shows the difference between the i-th feature-extraction value of a given image and the mean of the corresponding feature-extraction values of the images of the training set, which are classified as c_j. If we regard $\bar{x}$ as centroid representing descriptor c_j in a vector space, d_{jm} is now the distance vector of an image m to the centroid.

The vectors d_{jm}, however, tell us nothing about the probability that an image with value vector x_m belongs to a descriptor c_j, since those distance vectors are not directly comparable.

To calculate this probability, we need the variance-covariance matrix COV_j for the values of the c_j-classified images. Therefore, we define the matrix D_j as

$$D_j = X'_j \cdot X_j - \overline{X}'_j \cdot \overline{X}_j$$

where X_j is the matrix of the feature-extraction values from images manually classified with the same descriptor c_j (x_{iq} is the i-th feature-extraction value of the c_j-classified image q) from the training set. $\overline{X}_j$ is the matrix of means of all feature-extraction values of all images from the training set, manually classified as c_j (for a given i, $\overline{x}_{iq}$ is the mean feature-extraction value of all c_j-classified images for extraction method i). COV_j is defined as

$$COV_j = D_j \cdot \frac{1}{n_j}$$

where n_j is the number of images, classified as c_j. Using COV_j and the distance vector $d_{j,m}$, we now define

$$\chi^2_{jm} = d'_{jm} \cdot COV_j^{-1} \cdot d_{jm} + \ln|COV_j|.$$

χ^2_{jm} is a distance measure between the feature-extraction values of the image and the c_j classified images of the training set. With this measure we can determine which descriptor c_j belongs to an image m.

However, the χ^2_{jm} are not comparable over the attributes C, so we now calculate the probability $p(c_j|x_m)$ that an image with the feature-extraction values x_m belongs to a descriptor c_j as

$$p(c_j|x_m) = \frac{e^{-(\chi^2_{jm}/2)}}{\sum_k e^{-(\chi^2_{km}/2)}}.$$

where $\sum_k e^{-(\chi^2_{km}/2)}$ is the sum over all descriptors of the attribute C. This probability value can be used to order the images in descending order. In the retrieval process, we can derive the relevance values for an item by combining probabilities according to the composition of descriptors in the query interpretation resulting from the abductive processing of the user query. Interpretations involving more than one descriptor are equivalent to Bayesian networks. Hence, the relevance values are obtained using the Bayesian rule.

Conceptual queries consisting of single descriptors can be processed as follows: The vector x_m tends to be in a certain region of the feature value space, when picture m is assigned a descriptor c_j (the probability $p(c_j|x_m)$ serves as distance measure). Hence, it seems to be useful to encode the corresponding sections and index terms as part of indexing rules. We then can employ abductive reasoning to accomplish the mapping of conceptual query terms onto constraints oo feature representations.

Thus, the conceptual query is translated into search criteria on the feature level which can be executed in the metadata DBMS. The probabilities $p(c_j|x_m)$ are used to compute relevance weights (ranging between 0 and 1) of the retrieved images (an example is shown in Figure 1).

The ranked result list contains entries providing a thumbnail-link for each image. The user can now reformulate her request, if necessary, or view an image.

3.2 Rule Verification

In our verification, we compared the results of our retrieval engine with the outcome of manual indexing. In the first step, a set of 300 images was manually indexed and feature vectors for these images were computed. Then, the set was divided into two disjunct sets with the same number of elements. The first set was used as a training set to calculate the rules, and the second set as a testing set to estimate the quality of the discovered rules.

Then our retrieval engine calculated the relevance values for each descriptor and for all images in the testing set, using the discovered rules, ranked them and stored the relevance value for each (descriptor,image)-pair in a log-file. This log-file was compared with a file containing the manual indexing for the images, using the TREC evaluation software. The software calculates the precision of our rules for several recall stages.

Tables 1 – 4 show the result of the TREC evaluation software for the two alternative rule generation methods, brute force (bf) and QCF (qcf). The percent value in parentheses shows the quota of the descriptor in the testing set.

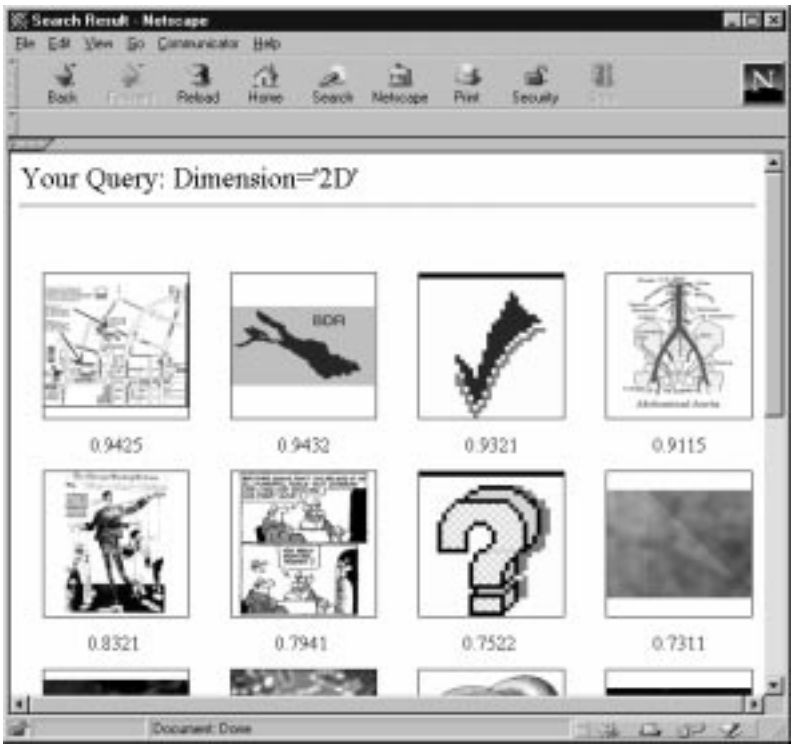

Figure 1: Example Search Result Screen

The brute force algorithm shows slightly better results than the QCF formula. The main advantage of the QCF is that the calculation of the rules is faster than by the brute force algorithm (90 times) with a tolerable loss in accuracy. Similar accuracy results were achieved by the neural networks used by Rowe and Frew (1997), with a more homogeneous picture collection and a simpler classification scheme.

Recall		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	Mean
Natural	bf	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.91	0.82	0.76	0.71	0.93
(68,63%)	qcf	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.87	0.80	0.71	0.95
Artif.	bf	1.00	0.82	0.82	0.79	0.61	0.61	0.61	0.61	0.61	0.60	0.55	0.64
(31,37%)	qcf	1.00	0.90	0.90	0.87	0.87	0.76	0.68	0.68	0.62	0.60	0.60	0.75

Table 1: Light

The experiments showed that retrieval quality is increased if indexers assign descriptors uniformly. For example, the indexing for “dimension” (see Table 2) was more precise than for “object in front” (Table 4). The retrieval quality also decreased when indexers were not consistent in the use of the classification items (e.g., “shadow” was used by different indexers sometimes in different ways). The indexers also had problems with “kind” and “content”, but this is understandable, since they are not visual characteristics and were perhaps not intuitively named.

The experiments also showed that retrieval quality increases with the size of the learning set. In the future we therefore plan to revise the classification scheme, to enlarge the set of classified images, and to train the indexers. We expect of it an improvement of the retrieval quality. For the future we also plan to include a relevance feedback mechanism. Thus, the retrieval quality for both approaches can be improved. Furthermore we plan further statistical testing to improve the quality of the QCF.

Recall		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	Mean
2D	bf	1.00	1.00	1.00	1.00	1.00	0.95	0.95	0.95	0.87	0.70	0.20	0.90
(17,89%)	qcf	1.00	1.00	1.00	1.00	1.00	0.94	0.94	0.72	0.40	0.20	0.20	0.78
3D	bf	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.97	0.95	0.94	0.82	0.96
(82,11%)	qcf	1.00	1.00	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.90	0.93

Table 2: Dimension

Recall		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	Mean
Blurry	bf	0.26	0.26	0.26	0.26	0.26	0.23	0.18	0.17	0.17	0.17	0.14	0.19
(13,49%)	qcf	1.00	0.33	0.16	0.15	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.19
Middle	bf	0.50	0.42	0.32	0.32	0.22	0.20	0.20	0.20	0.20	0.18	0.15	0.25
(12,70%)	qcf	0.18	0.18	0.18	0.18	0.18	0.18	0.17	0.17	0.17	0.17	0.16	0.14
Sharp	bf	1.00	1.00	1.00	0.93	0.93	0.90	0.89	0.86	0.86	0.85	0.73	0.86
(73,81%)	qcf	1.00	1.00	0.93	0.92	0.92	0.82	0.75	0.75	0.74	0.74	0.74	0.84

Table 3: Contour of main object

4 Lessons Learned and Future Work

In the last section we described in some detail the procedure (steps) and particular methods employed to generate rules which map extracted image feature values to conceptual index terms. The aim of the various analyses and experiments carried out was both to confirm our general hypothesis that semi-automatic rule generation can support semantic access to pictorial data, and to explore the potential of different algorithms for calculating the probability that an image matches a manual classification item (efficiency and precision of the rule generation).

In order to be able to perform detailed experiments in a realistic application context (the image retrieval prototype) the *first phase* of our work was pragmatically motivated and mainly explorative. As there exists no generally approved, elaborate theoretical framework or method for concept-oriented automatic indexing and retrieval of multimedia data, we had to rely on our own empirical work and experimentation. Our classification scheme, for example, was not directly derived from some other existing scheme; instead we defined a (relatively small) number of categories that seemed most appropriate as sample categories relevant to different kinds of topical and non-topical relevance criteria which might apply to image retrieval tasks. The categories mainly concerned visual or graphical image characteristics (such as light and dimension) and a few categories for subject indexing ("content" and "kind" of objects depicted). Naturally, this list was far from complete, and we expected to get more evidence from our following experiments on the validity and reliability of the categories and on how the scheme might be improved. In fact, some of the test results discussed above indicated where indexers faced problems and tended to assign index terms inconsistently; the analyses also showed that there were clear differences in the accuracy of the generated indexing rules across categories (e.g.,

Recall		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	Mean
Light	bf	1.00	0.72	0.72	0.65	0.50	0.50	0.50	0.50	0.50	0.50	0.45	0.54
(44,25%)	qcf	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.40
Dark	bf	1.00	0.57	0.57	0.57	0.57	0.57	0.57	0.54	0.52	0.51	0.49	0.54
(55,75%)	qcf	0.85	0.85	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.60	0.60	0.59

Table 4: Object in Front

"content" matched rather poorly, whereas "light" and "dimension" did significantly better—which sounds plausible).

Given the manually indexed test collection (provisionally accepting the observed small inconsistencies in these data), our experiments so far concentrated on the evaluation and comparison of the algorithms for rule generation and verification. They yielded a number of useful results, which could already be used to successively improve the rule generation method. However, more detailed experiments and systematic comparisons are needed in order to confirm and further qualify the present findings. Since these are heavily dependent on the quality and scope of the manually indexed data, we have started to develop a new, more differentiated classification scheme—both in light of our experiences with the first collection and more theoretically motivated considerations. In the *next phase*, this revised scheme will be used to re-index an extended image test collection. This will provide us with a larger, more consistent and richer data basis for improving the system-generated indexing rules, at the same time allowing more detailed statistical testing for the rule verification.

Since the new classification scheme is intended to hold for various image source types (e.g., photos, paintings, drawings, graphics, and computer-generated images), the test collection will be extended to cover large enough samples of each type. Further, we will include a number of new and more detailed categories for *formal* indexing and specification of the image *source* and *type* (e.g., representational vs. non-representational, depiction of natural objects vs. abstract ideas or decorative visual patterns). This will, at least, allow better subsampling of the test collection and systematic statistical comparisons by different image sources and types.

Concerning the specification of (semantic) image content, the distinction between *subject indexing* (subject matter/theme or kind of object depicted) and *visual or graphical characteristics* (color, dimension, foreground/background, etc.) will be made more explicit in the definition of the respective categories and the instructions to the indexers. We will include a small number of categories not considered so far, and both the attributes and the possible value ranges will be defined more precisely and described by examples. This is done by drawing on our experiences with the previous test collection (missing values, inconsistent assignment of index terms) and by analyzing additional sample sets of images (partly from printed material) which include more varied image types and potentially "problematic" cases. First such analyses have already led to re-definitions of some categories and possible values. They also indicated that a "facetted" approach might be necessary, since some attributes apply only to particular image types and are irrelevant for others (e.g., light/shadow or foreground/background in pictures not depicting (real-world) "objects" but rather some texture or color pattern). By running several indexing tests with these samples, the classification scheme will successively be modified and refined. Explicit definitions and more detailed descriptions and instructions than so far will be made available to the indexers, and they will be *trained* to use the classification scheme appropriately until a satisfactory degree of agreement between several independent indexers can be achieved.

As mentioned before, the newly indexed test collection will be used in a similar way as the first one, i.e. part of it as a training set for the automatic generation of indexing rules and the other part as a test set for the rule verification. Detailed statistical testing based on the new data will show whether and in which respects (e.g., for which attributes) the automatic indexing results could be improved. A second strand of future research, on the other hand, will be concerned with improvements of the retrieval quality of the prototype and the user-system interaction, e.g., by incorporating some relevance feedback mechanism and studying end-users in their use of the system and their subjective assessments of the retrieval effectiveness.

5 Conclusions

In this paper we have argued that traditional IR and similarity search approaches alone will not be sufficient for searching in the context of digital libraries. Experience at operating archives and agencies shows (Lutes et al. 1996) that customers often measure retrieval success against very specific evaluation criteria, necessitating detailed annotations or suitable metadata on origin, possible uses, etc. of documents. Our experiments with multimedia indexing and retrieval imply that, above all, the intelligent combination of domain-independent indexing rules and task-, user-, and domain-dependent retrieval heuristics (rules) in a logic-based retrieval system can lead to an increase in precision and can support a differentiated concept of relevance in the context of searching in digital libraries.

In the case of image retrieval, the inadequacy of retrieval methods based solely on physical document content, i.e. pixel data, are particularly evident. Statistical similarity measures between pixel values may very well say little about

the relevance of a given image for a given user. In order to achieve the goal of high-precision search a well thought out set of rules is needed. First experiments with rules formulated on the basis of quantile estimates led to recall and precision rates which are, in the best cases, comparable to those of probabilistic text retrieval systems. Further improvements were achieved by taking into account feature vectors. The Quadratic Classification Function (QCF) approach yielded more reliable indexing rules than the quantile analyses.

Although this approach makes it possible—to a certain degree—to conceptually index large image collections, provided a relatively small manually indexed test set is available, it reaches its limits when searching is aimed at individual concepts such as certain persons, specific situations, etc. This disadvantage can be avoided by storing images embedded in text in a hypertext system. If, for example, a newspaper article on the President of the United States contains an image which is automatically identifiable as a portrait, it can be assumed that the image shows Bill Clinton. Similarly, an analysis of the textual information in a multimedia document often leads to good results (see, e.g., Harmandas et al. 1997). Combining such information with the derived image information can further improve results.

References

Belkin, N. J., Cool, C., Stein, A., and Thiel, U. (1995). Cases, scripts, and information seeking strategies: On the design of interactive information retrieval systems. *Expert Systems and Applications* 9(3):379–395.

Bortz, J. (1989). *Statistik für Sozialwissenschaftler*. Berlin and New York: Springer.

Brady, M. (1982). Computational approaches to image understanding. *ACM Computing Surveys* 14(1):3–72.

Callan, J. P., Croft, W. B., and Harding, S. M. (1992). The INQUERY retrieval system. In *Proceedings of the 3rd International Conference on Database and Expert Systems Application*. Berlin and New York: Springer. 78–83.

Duguid, P., and Atkins, D. E. (1997). Report of the Santa Fe planning workshop on distributed knowledge work environments: Digital libraries. URL: http://www.si.umich.edu/SantaFe/.

Everts, A. (1996). PiClasso – picture classification operators. Master's thesis, TU Darmstadt, Department of Computer Science. In German.

Froehlich, T. (1994). Relevance reconsidered — towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science* 45(3):124–132.

Gevers, T., and Smeulders, A. (1992). Indexing of images by pictorial information. In Knuth, E., and Wegner, L. M., eds., *Visual Database Systems. Proceedings of the IFIP TC2/WG2.6 Second Working Conference on Visual Database Systems*. North Holland: Elsevier. 93–100.

Green, R. (1995). Topical relevance relationships: Why topic matching fails. *JASIS* 46(9):646–653.

Haralick, R. M., Shanmugan, K., and Dinstein, I. (1973). Textual features for image classification. *IIIE Transactions on Systems, Man, and Cybernetics* SMC-3(6):610–621.

Harmandas, V., Sanderson, M., and Dunlop, M. (1997). Image retrieval by hypertext links. In Belkin, N. J., ed., *Proceedings of the 20th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval (SIGIR '97)*, 296–303.

Hermes, T., Klauck, C., Kreyß, J., and Zhang, J. (1995). Image retrieval for information systems. Technical report, Universität Bremen, Fachbereich 3, AG KI.

Hirata, K., and Kato, T. (1992). Query by visual example. In Pirotte, A., Delobel, C., and Gottlob, G., eds., *Advances in Database Technologie - EDBT'92*, 56–71. Berlin and New York: Springer.

Lutes, B., Kutschekmanesch, S., Thiel, U., Berrut, C., Chiaramella, Y., Fourel, F., Haddad, H., and Mulhem, P. (1996). Study on non-textbased information retrieval – state of the art, eu study elpub 106. Technical report, European Commission, DG XIII Information Engineering.

Mehtre, B. M., Kankanhalli, M. S., and Lee, W. F. (1997). Shape measures for content based image retrieval: A comparison. *Information Processing & Managment* 33(3):319–337.

Mizzaro, S. (1997). Relevance: The whole history. *JASIS* 48(9):810–832.

Müller, A., and Everts, A. (1997). Interactive image retrieval by means of abductive inference. In *Proceedings of the RIAO'97 Conference – Computer-Assisted Information Searching on Internet*, 450–466.

Müller, A., and Thiel, U. (1994). Query expansion in an abductive information retrieval system. In *Proceedings of the Conference on Intelligent Multimedia Information Retrieval Systems and Management (RIAO '94), Vol. 1*, 461–480.

Narasimhalu, A. (1993). CAFIIR: An image based CBR/IR application. In *Proceedings of the AAAI Spring Symposium Series*, 70–77.

Picard, R. W., Kabir, T., and Liu, F. (1993). Real-time Recognition with the entire Brodatz Texture Database. In *IIIE Conf. on Computer Vision and Pattern Recognition*.

Rowe, N. C., and Frew, B. (1997). Automatic classification of objects in captioned depictive photographs for retrieval. In Maybury, M. T., ed., *Intelligent Multimedia Retrieval*. AAAI Press/The MIT Press. 65–79.

Stein, A., Gulla, J. A., Müller, A., and Thiel, U. (1997). Conversational interaction for semantic access to multimedia information. In Maybury, M. T., ed., *Intelligent Multimedia Information Retrieval*. Menlo Park, CA: AAAI/The MIT Press. 399–421.

Stein, A., Gulla, J. A., and Thiel, U. (1999). User-tailored planning of mixed initiative information-seeking dialogues. *User Modeling and User-Adapted Interaction* 9(1-2):133–166.

Thiel, U., Gulla, J. A., Müller, A., and Stein, A. (1996). Dialogue strategies for multimedia retrieval: Intertwining abductive reasoning and dialogue planning. In Ruthven, I., ed., *MIRO 95. Proceedings of the Final Workshop on Multimedia Information Retrieval*. Berlin and New York: Springer (eWiC, electronic Workshops in Computing series).

Thiel, U., Everts, A., Lutes, B., Nicolaides, M., and Tzeras, K. (1998a). Convergent software technologies: The challenge of digital libraries. In *Proceedings of the 1st Conference on Digital Libraries: The Present and Future in Digital Libraries, November 1998, Seoul, Korea*, 13–30.

Thiel, U., Hollfelder, S., and Everts, A. (1998b). Multimedia management and query processing issues in distributed digital libraries: A HERMES perspective. In Wagner, R. R., ed., *Proc. of the 9th Int. Workshop on Database and Expert Systems DEXA'98*, 84–89.

Turtle, H., and Croft, B. (1991). Efficient probabilistic inference for text retrieval. In *Proceedings of the RIAO'91, Barcelona, Spain*, 644–661.

van Rijsbergen, K. (1989). Towards an information logic. In Belkin, N., and van Rijsbergen, K., eds., *Proceedings of the SIGIR '89*. New York: ACM Press. 77–86.

Wilson, P. (1973). Situational relevance. *Information Storage & Retrieval* 9:457–471.