

# A Retrieval Mechanism for Semi-Structured Photographic Collections

Joemon M. Jose and David J. Harper

The Robert Gordon University  
Aberdeen, UK, AB25 1HG.  
{j.jose,d.harper}@scms.rgu.ac.uk

**Abstract.** In this paper, a new approach for retrieval from semi-structured photographic collections is described. We have developed a retrieval model based on the Dempster-Shafer theory of evidence combination. Basic concepts of the Dempster-Shafer theory are explained and the suitability of this theory for information retrieval is explored. A retrieval model for a semi-structured photographic collection is presented. Extensibility of this retrieval model for multimedia information retrieval is discussed. Integration of database and information retrieval concepts is a major requirement for semi-structured multimedia information retrieval and is accomplished in this model. A novel indexing scheme for photographic materials is described. We use spatial features, which are objects and their location, as photographic features. We have developed a multi-modal query interface for querying a photographic collection. A prototype system, Epic, has been implemented and is described in this paper.

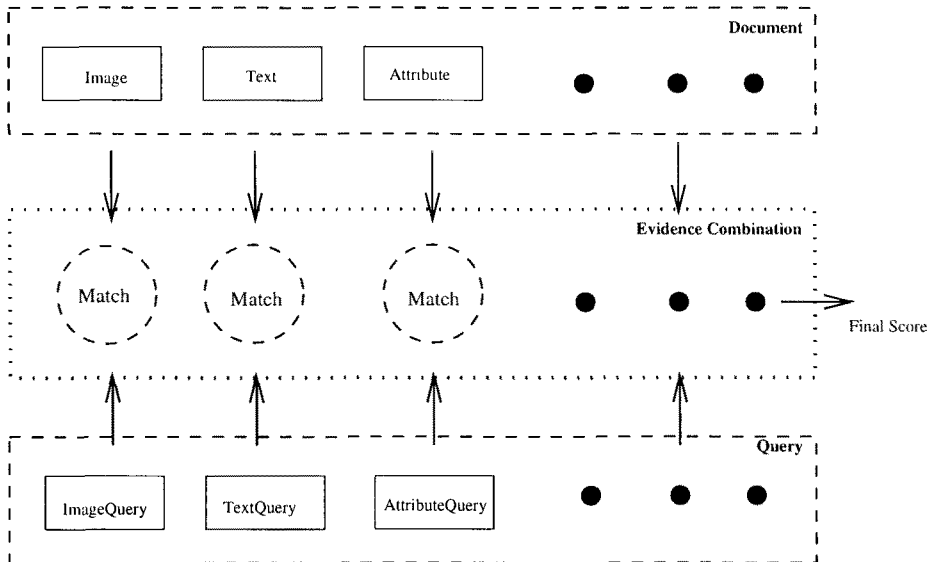
## 1 Introduction

Due to the widespread use of photographs in everyday life, there is a need for storing, managing and retrieving photographs [Orn95]. A large amount of pictorial information is produced in domains like the newspaper industry, the advertising business and the publishing industry. Journalists may look for photographs that describe an event or concern a person. Advertising professionals and publishers may look for photographs that are most useful for conveying their ideas. A number of commercial as well as in-house photographic repositories support these large populations of desirous searchers [Caw93,FMBR95].

Most of these photographic or other similar multimedia repositories are semi-structured. By semi-structured, we mean that the data has no absolute schema fixed in advance, and its structure may be irregular or incomplete. As an example, a newspaper photographic collection may contain photographs, the report or story published along with that photograph, reporter's name, date, photographer's name etc. Searchers may need exact match retrieval based on some attributes of these documents (eg. photographer name) and inexact match retrieval for some other attributes (eg. story) or a combination of both. Moreover, to cater for the wide-ranging access requirements to a multimedia repository,

we need to provide retrieval based on the content of the various media types [GR95]. Unfortunately, with the present state of image processing techniques, we cannot achieve large scale domain independent recognition of the content of images. However, a number of promising approaches indicate some results in fairly restricted domains [NBE<sup>+</sup>93,O'D93]. In the absence of a single ideal feature extraction technique and also considering multiple types of data and various access requirements, it is our contention that retrieval effectiveness can be improved if we combine or integrate different feature extraction and retrieval mechanisms [JH95].

Our approach is depicted in Fig. 1. A multimedia collection can be queried in a variety of ways. The idea is to match a query component with a corresponding component of the document (eg. image query with the image component of the document) and arrive at a similarity value. This activity will result in one or more similarity values for the documents in the collection. What we need next is a flexible and efficient method for combining these similarity values to arrive at a final score. This combination method should also have the power to combine exact (DB) and inexact (IR) match retrieval. We propose to use the powerful Dempster-Shafer method for combining evidence from multiple sources. We have developed a retrieval model based on the Dempster-Shafer approach which satisfies all the above requirements and this model is presented in this paper. The goal of this paper is to describe a new retrieval model that can be



**Fig. 1.** Combining different retrieval mechanisms

used for multimedia information retrieval and also for database and information

retrieval integration. To demonstrate the feasibility of our approach, we have applied this method to a collection of photographs. The outline of this paper is as follows. We will introduce the Dempster-Shafer theory followed by an exploration of the properties of this theory for information retrieval applications. After that, we will describe the application of this theory in photographic retrieval. Then, we will discuss the implementation of a prototype photograph retrieval system. Subsequently, we will discuss related work and conclude.

## 2 Dempster-Shafer Theory

Dempster-Shafer theory provides a framework for accumulation of evidence and its combination. The heart of this theory is Dempster's rule of combination: the combination of degrees of belief or support based on one body of evidence with those based on an entirely distinct body of evidence. This theory is a generalisation of the Bayesian approach, in which the commitment of a portion of belief to a proposition implies that the rest of belief is committed to the negation of the proposition. In the Dempster-Shafer formalism, commitment of a portion of the belief to one proposition does not imply commitment of the rest of its belief to its negation. In the absence of any evidence in support of the negation of a proposition, the remaining belief is assigned to the entire set and thus allows us to represent our uncertainty. In the following, the basic structures and the combination method of Dempster-Shafer theory are described [Sha76].

### 2.1 Frame of Discernment

The *frame of discernment*, denoted by  $\Theta$ , is an exhaustive set of mutually exclusive elements that can be interpreted as hypotheses or propositions. The power set that contains all the subsets of  $\Theta$  is denoted by  $2^\Theta$  and the elements of this power set are also considered as hypotheses.

As an example, consider a collection of documents containing three documents say  $d_1, d_2, d_3$ . Then, the possible propositions of interest would be: with respect to an information need, document  $\{d_i\}, i = 1, 2, 3$  is relevant, documents  $\{d_i, d_j\}, i, j = 1, 2, 3$ , are relevant, and documents  $\{d_1, d_2, d_3\}$  are relevant. Hence, in this case the frame of discernment  $\Theta = \{d_1, d_2, d_3\}$  and the subsets of  $\Theta$  corresponds to our propositions of interest.

When a proposition corresponds to a subset of a frame of discernment, we will say that the frame discerns that proposition.

### 2.2 Basic Probability Assignment

The Dempster-Shafer theory uses a number in the range  $[0, 1]$  to indicate the degree of belief in a hypothesis or proposition given a piece of evidence. This number is the degree to which the evidence supports the hypothesis. And the impact of the distinct piece of evidence on the subsets of  $\Theta$  is represented by a function called a basic probability assignment (*bpa*) or mass function, and is denoted by  $m$ .

**Definition 1.** If  $\Theta$  is a frame of discernment, then a function  $m : 2^\Theta \rightarrow [0, 1]$  is called a *basic probability assignment* whenever

$$m(\emptyset) = 0 \text{ and } \sum_{A \subseteq \Theta} m(A) = 1.$$

The set  $F = \{A | A \subseteq \Theta, m(A) > 0\}$  is called a focal element set and elements of this set are called focal elements of the mass function.

The quantity  $m(A)$  is a measure of that portion of the total belief committed exactly to  $A$ , where  $A \subseteq \Theta$ , and the total belief is one. This portion of belief cannot be further subdivided among the subsets of  $A$  and does not include portions of belief committed to subsets of  $A$ . The quantity  $m(\Theta)$  is a measure of that portion of the total that remains unassigned after commitment of belief to various proper subsets of  $\Theta$ . Thus, the *bpa* assigned to  $\Theta$  represents the ‘uncertainty’ of the evidence.

### 2.3 Belief Function and Belief Interval

In the Dempster-Shafer formalism, the total belief in a proposition  $A$  is represented by an interval, say  $[B(A), P(A)]$ . The lower value,  $B(A)$  represents the support for the proposition  $A$  and sets minimum value for its likelihood. The upper value,  $P(A)$ , denotes the plausibility of that proposition and establishes a maximum likelihood. That is,  $P(A)$  expresses how much we should believe in a proposition if all currently unknown facts were to support that proposition. Thus the true belief in that proposition will be somewhere in the interval  $[B, P]$ . The width of a belief interval can also be regarded as the amount of uncertainty with respect to a hypothesis, given the evidence. The quantity  $B(A)$  includes degree of belief committed exactly to  $A$  and also to the subsets of  $A$ .

An understanding of belief functions and belief intervals is not necessary for the purpose of this paper. For more information see [Sha76].

### 2.4 Combination of Evidence

Mathematically, Dempster’s rule is simply a rule for computing, from two or more belief functions over the same set  $\Theta$ , a new belief function called their orthogonal sum. This rule corresponds to the pooling of evidence: if the mass functions being combined are based on entirely distinct bodies of evidence and the set  $\Theta$  discerns the relevant interaction between those bodies of evidence, then the orthogonal sum gives degrees of belief that are appropriate on the basis of combined evidence.

Let  $m_1$  and  $m_2$  are two mass functions defined in the frame of discernment  $\Theta$ . Then a new mass function,  $m$ , is defined as follows:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) \star m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) \star m_2(C)} \quad A, B, C \subseteq \Theta \quad (1)$$

This is known as Dempster-Shafer evidence combination rule.

### 3 Exploration of Dempster-Shafer Theory

What is the suitability of Dempster-Shafer theory for information retrieval? In this section, we will answer this question by exploring certain aspects of the Dempster-Shafer theory.

In general, the Dempster-Shafer evidence combination is computationally expensive. Considering the need to retrieve from large collections like TREC, the application of the Dempster-Shafer combination method will involve serious computational effort. In the following, we will show how we can simplify the evidence combination method in a particular case where we have positive evidence for singleton hypotheses only. We will also analyse the behaviour of this combination function in some specific cases of uncertainty. Following that, we will show the power of Dempster-Shafer combination method for database and information retrieval integration.

Let us consider a frame of discernment with three hypotheses, i.e.  $\Theta = \{d_1, d_2, d_3\}$ . Suppose,  $m_1$  and  $m_2$  are two mass functions defined on  $\Theta$  based on two distinct bodies of evidence. Then  $m_1(\Theta)$  and  $m_2(\Theta)$  represent the uncertainties in those bodies of evidence. Let us assume that the focal elements of these two mass functions are the singleton hypotheses and the frame  $\Theta$  i.e., we have positive belief for  $\{d_1\}$ ,  $\{d_2\}$ ,  $\{d_3\}$  and  $\Theta$  only.

#### 3.1 Simplification of the Combination Method

In this section, we will show the simplification of the combination method shown in equation (1).

The orthogonal sum  $m_1 \oplus m_2$ , say  $m$ , can be computed using the combination method in equation (1):

$$m(\{d_i\}) = \frac{m_1(\{d_i\}) \star m_2(\{d_i\}) + m_1(\Theta) \star m_2(\{d_i\}) + m_1(\{d_i\}) \star m_2(\Theta)}{1 - \sum_{\{d_k\} \cap \{d_j\} = \emptyset} m_1(\{d_k\}) \star m_2(\{d_j\})}$$

The denominator in the above equation is a normalising factor and is independent of  $\{d_i\}$ . Hence, the above equation can be written as,

$$m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + m_1(\Theta) \star m_2(\{d_i\}) + m_1(\{d_i\}) \star m_2(\Theta) \quad (2)$$

Equation (2) can be used to compute combined degrees of belief and is computationally much less expensive than the Dempster-Shafer combination function given in equation (1). Thus, given distinct bodies of evidence that have positive impact on singleton hypotheses only, we can apply this simplified formula for evidence combination. The resulting mass values  $m(\{d_i\})$  can be normalised by dividing each by  $\sum_i m(\{d_i\})$ . However, this is not necessary as relative belief values are sufficient for ranking retrieved documents.

The rule of combination shown in equations (1) and (2) is a rule for combining a pair of belief functions, but by repeatedly applying it, one can obviously combine any number of belief functions. The Dempster-Shafer combination is commutative ([Sha76], page 61, Theorem 3.3) and hence the order in which evidence is combined is not important.

### 3.2 Boundary Conditions

In this section, we will look into the behaviour of the theory in various cases of uncertainty.

In the case of complete confidence in a piece of evidence we have  $m(\Theta) = 0$ , and then  $m$  reduces to be a probability function.

If we have complete confidence in two bodies of evidence then equation (2) becomes:

$$m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + 0 \star m_2(\{d_i\}) + m_1(\{d_i\}) \star 0 \quad (3)$$

$$\text{i.e., } m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\})$$

That is, in the case of complete evidence, Dempster-Shafer based combination reduces to be a probability based combination method.

In the case where we have complete confidence in one piece of evidence, say in  $m_1$  (i.e.,  $m_1(\Theta) = 0$ .) then equation (2) becomes:

$$m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + 0 \star m_2(\{d_i\}) + m_1(\{d_i\}) \star m_2(\Theta)$$

$$\text{i.e., } m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + m_1(\{d_i\}) \star m_2(\Theta) \quad (4)$$

In this case, the mass function for which we have complete confidence contributes more towards the combined belief.

If  $m_1(\Theta) = m_2(\Theta) = \mu$  then equation (2) becomes:

$$m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + \mu \star m_2(\{d_i\}) + m_1(\{d_i\}) \star \mu$$

$$\text{i.e., } m(\{d_i\}) \propto m_1(\{d_i\}) \star m_2(\{d_i\}) + \mu \star [m_2(\{d_i\}) + m_1(\{d_i\})]$$

In this case, the proposition with high belief will contribute more towards the combined belief.

With the above observations, we could conclude that the behaviour of the theory accords well with the general intuition on combining evidence.

### 3.3 DB and IR Integration

An integrated framework in which we can provide exact match and inexact match retrieval is an essential requirement for retrieval from semi-structured information repositories. Here, we will describe the power of the Dempster-Shafer combination method in handling exact match and inexact match retrieval in one computational framework.

We have seen above that in the case of complete confidence in one body of evidence, say in  $m_1$  (i.e.,  $m_1(\Theta) = 0$ .) the equation (2) reduces to the equation (4).

In equation (4), if  $m_1(\{d_i\}) = 0$  then the combined belief  $m(\{d_i\}) = 0$ .

That is, if we have complete confidence in a body of evidence, then this combination mechanism will impose a filtering condition on the combined belief. We will have final belief only for those hypotheses which have positive belief

based on complete evidence. Thus, for exact match retrieval we model mass functions without any uncertainty. Hence, when we apply equation (2) to combine evidences, the resulting mass function will provide belief for hypotheses which satisfy the exact match criterion.

This property makes the combination function in equation (2) more suitable for combining exact and inexact match retrieval. Moreover, like with handling vague queries in conventional database systems we could incorporate distance metrics for attribute-value queries.

We have seen the behaviour of the theory in situations where mass functions generate positive evidence for singleton hypotheses. In such situations, application of equation (2) for evidence combination reduces computational expenses. Now, we will show the application of this theory in a photographic retrieval situation where we model mass functions with focal elements as singleton hypotheses.

## 4 A Photographic Retrieval Model

In this section, we will explain our retrieval model, based on the simplified Dempster-Shafer theory, applied to the case of photographic information retrieval.

We have a photographic collection in which the photographs are stored along with various textual information. To simplify the discussion, and without loss of generality, we assume that a document in the collection has only two components, a picture component and a text component. Hence, we define a document as follows.

**Definition 2.** A *multimedia document*  $\mathcal{M}$  is a structure  $\mathcal{M} = \langle P, T \rangle$  where  $P$  is the photographic component of  $\mathcal{M}$ , and  $T$  is its textual component.

The various querying mechanisms are explained in the next section and for the sake of this discussion we assume that there are two types of querying: text and picture querying. Searchers can use either of these or both. Thus, a query can also be seen as a document: it also has two components, a picture query component and a text query component. Each query component is considered as a body of evidence and we create mass functions based on these. Then, these mass functions are combined to generate a final belief. Later, we will see how a searcher's uncertainty in the representation of a query component is dealt with.

In the following, we will explain how we obtain basic probability assignments based on different sources of evidence. It is assumed that a collection of  $N$  documents exists. Then, the frame of discernment is  $\Theta = \{d_1, d_2, \dots, d_N\}$ . The corresponding proposition of interest is that *a document  $d_i$  is relevant to the searcher's information need.*

Next we will introduce our picture characterisation scheme and the corresponding querying mechanism. Following that, the individual retrieval models for pictorial matching and text matching are described. At the end of this section, we will show the evidence combination approach.

## 4.1 Photograph Indexing and Querying

One important aspect of designing access methods for picture collections is determining the sort of features that can be used for retrieval. We propose to use objects in the photograph and their locations as features for picture retrieval. We refer to these features as the spatial features from a photograph. In this section the spatial indexing and querying method is discussed along with a discussion of the usefulness of spatial features for retrieval.

**Spatial Indexing.** Commonly, images are characterised by associating with them a set of keywords or some attribute-value pairs [Mos94]. These keywords are assigned to the image manually and used for retrieval. However, this approach has all the drawbacks of the standard text retrieval systems that use manually assigned keywords for retrieval. Moreover, the problems would be exacerbated if one uses keywords to describe spatial relationships among objects in the image. This is because numerous spatial relationships are possible, even between two objects, and any particular relationship may be described in various ways.

This situation can be ameliorated by associating the keywords with the corresponding objects in the photograph instead of associating them with the entire picture. Then, the object and the corresponding location constitute an indexing feature.

Photographs can be seen as consisting of objects. These may be objects like people, buildings, automobiles, gardens, etc. The objects have certain locations in the image. Hence we can describe an image  $P$  as  $P = \{(\lambda_1, \rho_1), (\lambda_2, \rho_2), \dots\}$ , where  $\lambda_1, \lambda_2, \lambda_3, \dots$  are the objects and  $\rho_1, \rho_2, \rho_3, \dots$  are their respective regions. These objects and their region are used as picture features and are known as spatial features.

Computing tools can be provided to extract these features with the minimum intervention from the user. To index an image, the indexer identifies objects and specifies the locations of these objects in the image. This is done by drawing a box (rectangle) around the object using the mouse. Then, the indexer describes this object by naming it. From these descriptions, the system automatically generates spatial features. In comparison with the keyword based retrieval approach, the user is performing minimal additional work by identifying the area in which objects appear in an image.

**Why Spatial Indexing.** We will describe briefly why we believe spatial indexing can be useful for effective picture retrieval. More detailed discussion on spatial features is available in [JH95].

Relatively few studies have been conducted into the kind of queries that users are likely to put to an image collection. Enser [Ens93] has done a study on user queries to the Hulton Deutch collection. He says, ‘Sometimes the need to translate the client’s mental image of his information need into a linguistic expression is facilitated by the identification of that need with a particular picture, or by the incorporation of the sketch, in those cases where the request is received by letter



or fax'. If the user has a mental image of a need, we believe this can be more readily translated into a set of objects and their relative or absolute locations in a picture.

Lansdale et.al. [LSW96] have described the need for a spatially depicted interface for retrieval from visual repositories. Their initial experiments have shown that a spatially depictive interface to visual collections will enhance the retrieval effectiveness. Also, an interface that is expressive enough will give a searcher an opportunity to reflect on her information need and modify it as required.

**Spatial Querying.** If the spatial features are useful for picture retrieval then the question is "how do we use them for querying?". Spatial querying is described in here.

A query canvas of the size of a photograph is provided for spatial querying [JHH96]. Searchers can draw rectangles on this query canvas and label them. From these, the system automatically derives spatial query features.

As an example, consider a user who wants a photograph of a fountain with a tree behind it. Assuming that the user has a mental image of a photograph with a tree in one corner and a fountain in the middle, two boxes can be drawn on the query canvas: one in a corner and the other at the center. The box at the corner will be named as 'tree' and the one at the center will be named as 'fountain'. In this way, a user's mental image of the information need can be captured.

In addition to spatial querying, searchers can perform other types of data access such as retrieval based on attribute-value pair and similarity retrieval based on textual features. This is discussed in section 5.

## 4.2 Basic Probability Assignment based on the Photographic Component

In the spatial retrieval model, the aim is to match the characterisation of the spatial component of the query to the characterisation of the spatial component of the documents. The query is characterised by spatial features. Using these features as evidence, a belief is calculated that indicates support for the proposition of interest. To formalise these notions, the following definitions will be used.

**Definition 3.** A *region*  $\rho$  is a structure  $\rho = \langle x, y, w, h \rangle$  where  $(x, y)$  is the origin of the rectangular area defining the region, and  $w$  and  $h$  are the width and height of the rectangular area.

**Definition 4.** A *spatial feature*  $\phi$  is a structure  $\phi = \langle \lambda, \rho \rangle$  where  $\rho$  is a region and  $\lambda$  is a label identifying the object associated with the region  $\rho$ .

**Definition 5.** A *picture* or spatial component  $P$  is a set  $P = \{\phi\}$  where  $\{\phi\}$  is a set of spatial features forming the characterisation of the picture.

**Definition 6.** Given two regions  $\rho_1$  and  $\rho_2$ . Then the *distance* between the regions  $\rho_1$  and  $\rho_2$  is defined as

$$D(\rho_1, \rho_2) = 1 -$$

$$\frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + ((x_1 + w_1) - (x_2 + w_2))^2 + ((y_1 + h_1) - (y_2 + h_2))^2}}{\Delta}$$

where  $\Delta$  is a normalising factor depending on the dimension of the picture  $P$ .

The value of  $D$  is in the range  $[0,1]$  where zero represents no similarity and 1 represents perfect similarity between the two regions. This distance measure takes into account the position as well as size of each region.

Using this measure we get a similarity value of one for two rectangles with the same origin and extent. This similarity value decreases as rectangles move apart and will be zero for two rectangles (points!) at the opposite corners of the space. This distance measure is used as spatial similarity.

**Definition 7.** Let  $\lambda_i$  and  $\lambda_j$  be the labels of two spatial features. Then, we can define a *picture indicator function* as follows

$$I(\lambda_i, \lambda_j) = \begin{cases} 1 & \text{if } \lambda_i = \lambda_j \\ 0 & \text{otherwise} \end{cases}$$

**Definition 8.** Let  $P_D$  be a picture document and let  $P_Q$  be a picture query and  $\phi_q = \langle \lambda_q, \rho_q \rangle$  be a spatial feature in  $P_Q$ . Then, the *Spatial Filter Function* for  $\phi_q$  is defined as

$$\mathcal{F}(\phi_q, P_D) = I(\lambda_q, \lambda_j) \star \max_j D(\rho_q, \rho_j)$$

where  $\phi_j = \langle \lambda_j, \rho_j \rangle$ , and  $\phi_j \in P_D$

When there are more than two instances of a same object in the photograph, this spatial filter function produces a similarity value for the closest object to the query feature. Hence, it avoids counting the same feature more than once and uses the highest similarity possible for a query feature.

Using the above definitions, we can define the evidence (or similarity score) for a document based on the spatial component as follows:

**Definition 9.** Let  $P_D$  be a picture document, and let  $P_Q$  be a picture query. Then, the similarity between  $P_D$  and  $P_Q$  can be defined as

$$\text{sim}(P_D, P_Q) = \sum_{\phi_i \in P_Q} \mathcal{F}(\phi_i, P_D)$$

By computing the similarity between spatial query component and the picture query component, we reach a positive similarity value for some documents in the collection and zero for the rest. The next step is to convert these similarity values into a mass function (*bpas*). Scores are normalised and converted into *bpas* by dividing each by the sum of all the scores.

A searcher is given an opportunity to specify his confidence,  $\mu$ , in each component of the query. This confidence is interpreted as the searcher's certainty in that query representation. Then the resulting uncertainty in that query representation is  $(1 - \mu)$ . This uncertainty  $(1 - \mu)$ , can be propagated by assigning a belief  $(1 - \mu)$  to the set of all documents (i.e. to the frame of discernment  $\Theta$ ). This means that the belief  $(1 - \mu)$  could not be assigned to any smaller subsets of  $\Theta$  based on the evidence at hand, but must instead be assumed to be distributed in some (unknown) manner among other focal elements of  $\Theta$ . That is  $m(\Theta) = 1 - \mu$ . To make  $\sum_i m(\{d_i\}) = 1$ , we multiply each normalised score by  $\mu$ . These resulting  $m$  values constitute the mass function or basic probability assignments based on the spatial query component. This process is shown in the table below.

Document	Score	Norm. Score	$bpa$
$d_1$	$s_1$	$n_1 = \frac{s_1}{S}$	$n_1 * \mu = m(\{d_1\})$
$d_2$	$s_2$	$n_2 = \frac{s_2}{S}$	$n_2 * \mu = m(\{d_2\})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$d_N$	$s_N$	$n_N = \frac{s_N}{S}$	$n_N * \mu = m(\{d_N\})$
$\Theta$	0	0	$(1 - \mu) = m(\Theta)$
Sum	$\sum_{i=1}^N s_i = S$	1	1

### 4.3 Basic Probability Assignment based on the Textual Component

Here, the objective is to match the textual representation of the document with the textual component of the query. We take the textual query features as evidence, and calculate the belief in these propositions. Before explaining the computation, we need the following definitions.

**Definition 10.** A *text component*  $T$  is a set  $T = \{\tau\}$  where  $\{\tau\}$  is a set of text features (e.g. terms in a natural language document).

**Definition 11.** Let  $\tau$  be a text feature. Then a (*inverse document frequency*) weight  $w(\tau)$  can be associated with  $\tau$  as follows

$$w(\tau) = \log \frac{N}{f(\tau)}$$

where  $N$  is the total number of documents in the document collection.

**Definition 12.** Let  $\tau_i$  and  $\tau_j$  be two text features. Then, we can define a *text indicator function* as follows

$$I(\tau_i, \tau_j) = \begin{cases} 1 & \text{if } \tau_i = \tau_j \\ 0 & \text{otherwise} \end{cases}$$

Using the above definitions, we could define the evidence (or score) for a document based on the text component of the query as follows:

**Definition 13.** Let  $T_D$  be a text document, and let  $T_Q$  be a text query. Then, the similarity between  $T_D$  and  $T_Q$  can be defined as

$$\text{sim}(T_D, T_Q) = \sum_{\tau_i \in T_Q} \sum_{\tau_j \in T_D} I(\tau_i, \tau_j) \star w(\tau_i)$$

This evidence is also normalised and the searcher's uncertainty in the representation of the textual query component is also taken into account by assigning it to the frame of discernment. A belief value is computed using the same procedure as used in the pictorial case explained at the end of the section 4.2.

#### 4.4 Evidence Combination

Now the component matching functions have been designed for both the textual and the picture component, evidence coming from both matching processes will have to be combined in order to arrive at one overall relevance score for the document given a query. Since we have positive evidence for singleton hypotheses only, the simplified Dempster-Shafer evidence combination mechanism described in equation (2) is applied.

Using equation (2), a combined belief is computed and the documents are presented to the searcher in decreasing order of belief.

In our present approach, we compute mass functions for a query component. However, we could compute mass functions for individual query features and then evidence combination can be applied. This may increase computational cost considerably and hence has not been considered for implementation.

We can extend this retrieval framework for more than two query components. We can model mass functions for any type of query components in a similar way and then the evidence combination formula can be applied. We have extended this model for simple database queries and described briefly in the next section.

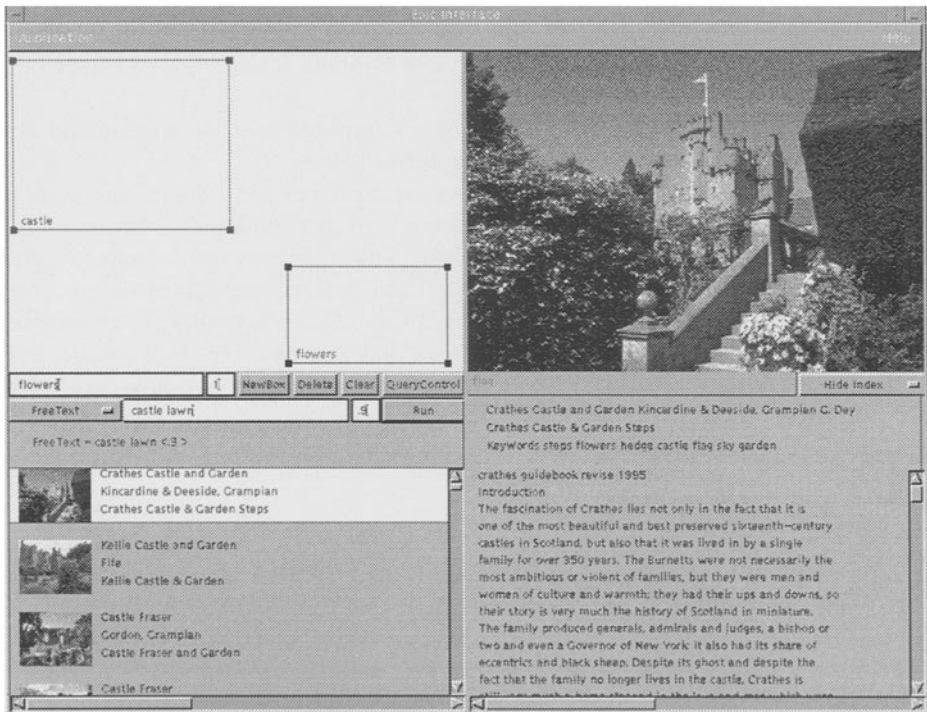
## 5 Implementation

We have demonstrated the practicability of our approach by building a prototype photograph retrieval system, called *Epic*, which is described briefly below. Our implementation is based on a client-server architecture and communication between the client and the server takes place using the HTTP protocol. In order to sustain our retrieval model we need to have an extensible server system. This permits us to add support for different media types and feature extraction and matching techniques. We also need to have an architecture that support the integration of exact and inexact match retrieval.

We have built a server system that satisfies these objectives. The server system is an extension of the *ECLAIR* framework [HW92] and is built on the principles of extensibility and reusability. The idea is to exploit the features of an object-oriented database management system by building a layer of IR framework on top of it. This allows us to utilise the features provided by an

underlying database management system. We have redesigned and extended *ECLAIR* and a light-weight data model that supports the representation of semi-structured collections has been built into this framework. For the representation of indexing features we use an inverted index structure. The data model and this index structure support the integration of exact and inexact match retrieval achieved in our retrieval model.

The client side of the application, namely the user interface, has been built using the JAVA programming language, and a snapshot of this interface is shown in Fig. 2. The upper left part of the figure is a multi-modal query interface and the bottom left is a result viewer. A document viewer is at the right hand side of the interface. The query interface supports various sorts of query mechanisms and the result viewer provides a thumb view of the retrieved documents. A searcher can select documents from the result viewer and view them in the document viewer.



**Fig. 2.** Epic Interface

One component of the query interface is a query canvas where searchers can sketch spatial queries. It also has a text-based query interface where searchers

can issue text queries. Searchers can also specify their confidence in each query component.

In the present implementation of the system we support simple database queries of the form “ name=value”. However, extending this interface for more complex queries is not considered to be a major problem, as we only need to add only the corresponding matching scheme in the server system. At present, a searcher can select a field name and specify their query term. The interface does not separate exact and inexact match queries but by assigning a confidence value of one the searcher is specifying a filtering condition which is equivalent to an exact match retrieval.

## 6 Discussion

Recently, integrated approaches to image retrieval have been discussed by many authors. In [OS95], the problem is approached from a database retrieval standpoint, in which an SQL-type mechanism is used to integrate content-based retrieval with attribute-based retrieval. They use information contained in hand-keyed database fields to supplement image content information. One drawback of this approach is that it cannot rank documents in the order of similarity. Also, the combination or filtering mechanism implemented using SQL is very rigid. WebSeer, a system being developed by Swain et al. [SFA97], uses information derived from analysing the image content to complement the textual information associated with an image and information derived from the image header. They have developed a form-based textual query interface for image searching on the web and they rank images based on their similarity value with the query. However, the final score is reached by an ad-hoc mechanism of combining similarities from various sources and then filtering on certain features. Moreover, searchers are not provided with an opportunity to specify their preference for various query features.

The problems identified in the systems discussed above are tackled in our system. Documents retrieved in our model are ranked in order of similarity. By the application of Dempster-Shafer theory we achieve flexible and efficient integration of various sources of evidence. As shown in Fig. 2, we have implemented the retrieval model for three types of feature extraction and matching. We can perform similarity retrieval based on textual features and/or spatial features. We can also perform attribute-based retrieval (exact match). Since the Dempster-Shafer combination function is commutative, the order of combining belief is not important. We could integrate a new retrieval mechanism into the framework as long as there is a feature extraction, representation and matching technique available. As an extension of this work, we are going to extend this retrieval model for colour and texture feature based retrieval.

What are the strengths of the retrieval framework discussed here in comparison to other approaches? First, we believe, this framework is a general one for multimedia information retrieval and can be extended to incorporate other media types and their corresponding feature extraction and matching techniques. Sec-

ond, this framework allows the integration of exact and inexact match retrieval in one computational model.

Over the years, a lot of work has been done integrating database and information retrieval concepts. Most of this work has tried to exploit an existing database by building an IR system on top of it. A review of such work is available in [HW92]. The *ECLAIR* architecture described in the implementation section is one such approach. The *ECLAIR* architecture integrates the features of an object-oriented database management system with an IR framework. Our work is an extension of this work. However, in the original *ECLAIR* framework the integration of database and information retrieval aspects is not achieved in its retrieval model. In the retrieval framework discussed here, we can achieve the integration of exact and inexact match retrieval in one process. Schauble [Sch93] proposes an architecture for DB and IR integration. The major difference is that in our model we could achieve combination of evidence from multiple sources along with their specified uncertainty. This allows us to extend our system for any number of media and feature extraction techniques. However, we have not dealt with the problems arising from dynamic data. Motro [Mot88] discusses an approach to integrate vague queries into a relational database management system. They ranked documents in order of similarity and users are allowed to provide their own similarity metrics. In our retrieval framework users can also define their own matching mechanism (by extending the server framework). However, the major difference is that our integration mechanism is achieved in one computational framework.

Van Rijsbergen [vR92] discusses the inadequacy of the Bayesian approach for information retrieval applications in the context of the inherent uncertainty involved in various phases of retrieval. He further introduces the Jeffrey's conditioning for information retrieval applications. Jeffrey's conditioning provides a mechanism for revising probability measures in the light of uncertain evidence. Jeffrey's conditioning is a generalisation of the Bayesian approach by incorporating the passage of experience or a searcher's acquired knowledge during the course of retrieval. When the sum of uncertainties involved in the query representations is equal to one (i.e.  $\sum_i \mu_i = 1$ ), our model satisfies Jeffrey's conditioning. However, in our model a searcher can express his uncertainty in a query representation independent of other query components. Moreover, Jeffrey's conditioning does not support exact and inexact match integration.

## 7 Conclusion

This work is grounded in the belief that progress in multimedia information retrieval can be made if different retrieval approaches can be combined or integrated in a flexible and efficient way. The Dempster-Shafer evidence combination mechanism provides a framework for achieving this goal. Since more and more multimedia collections are semi-structured in nature, we believe that it is time to develop a retrieval mechanism for semistructured repositories. We need to identify new features for providing effective retrieval for photographic collections.

Currently there is no single image processing technique that can be applied to images from diverse domains such as photographic collections. Therefore, there is need to develop and use semi-automatic techniques for image feature extraction. It is also very important to have a querying interface that exploits these image features for query representation.

The main contributions of our work are the following: we have proposed a new approach to multimedia information retrieval in which we combine multiple sources of evidence; and we have developed a retrieval model based on the Dempster-Shafer theory. This model provides a framework for retrieving information from semi-structured collections and integrates exact and inexact match retrieval. We have introduced a novel approach for picture retrieval which has the potential to provide more precise retrieval. A prototype system which is extensible and reusable has been built for photographic information retrieval.

## Acknowledgements

We are grateful to Professor Keith van Rijsbergen for introducing us to the Dempster-Shafer theory. Discussions with Jan-Jaap IJdens at various stages of this work was highly beneficial. We would also like to thank David Hendry for his encouragement. Finally, thanks should go to the National Trust of Scotland. The work presented here has been supported by the Principal's Research Fund at the Robert Gordon University.

## References

- [Caw93] A. E. Cawkell. Picture-queries and picture databases. *Journal of Information Science*, 19:409–423, 1993.
- [Ens93] P. G. B. Enser. Query analysis in a visual information retrieval context. *Journal of Document and Text Management*, 1(1):25–52, 1993.
- [FMBR95] S. Flank, P. Martin, A. Balogh, and J. Rothey. Photofile: A digital library for image retrieval. In *Proceedings of International Conference on Multimedia Computing and Systems*, pages 292–295, Washington, D.C., 1995.
- [GR95] V. N. Gudivada and V. V. Raghavan. *IEEE Computer: Special Issue on Content-Based Image Retrieval Systems*, volume 28 of 9. IEEE Press, September 1995.
- [HW92] D. J. Harper and A. D. M. Walker. ECLAIR: an extensible class library for information retrieval. *The Computer Journal*, 35(3):256–267, June 1992.
- [JH95] J. M. Jose and D. J. Harper. An integrated approach to image retrieval. *The New Review of Document and Text Management*, 1:167–181, 1995.
- [JHH96] J. M. Jose, D. J. Harper, and D. G. Hendry. A spatial feature based photograph retrieval system. In H. P. Frei, D. Harman, P. Schauble, and R. Wilkinson, editors, *Proceedings of the Nineteenth Annual International SIGIR Conference on Research and Development in Information Retrieval*, page 341. ACM Press, August 1996.
- [LSW96] M.W. Lansdale, S. A. R. Scrivener, and A. Woodcock. Developing practice with theory in HCI: applying models of spatial cognition for the design of pictorial databases. *International Journal of Human-Computer Studies*, 44:777–799, 1996.



- [Mos94] J. Mostafa. Digital image representation and access. *Annual Review of Information Science and Technology*, 29:91–135, 1994.
- [Mot88] A. Motro. VAGUE: A user interface to relational database that permits vague queries. *ACM Transactions on Office Information Systems*, 6(3):187–214, July 1988.
- [NBE<sup>+</sup>93] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The qbic project: Querying images by content using color, texture, and shape. In W. Niblack, editor, *Proceedings of SPIE-93*, pages 173–187. SPIE, February 1993.
- [O'D93] M. H. O'Docherty. A multimedia information system with automatic content retrieval. Technical Report UMCS-93-2-2, Department of Computer Science, University of Manchester, Manchester M13 9PL, England, 1993.
- [Orn95] S. Ornager. The newspaper image database: Empirical supported analysis of users' typology and word association clusters. In E. Fox, P. Ingwersen, and R. Fidel, editors, *Proceedings of 18th SIGIR Conference*, pages 212–218, Seattle, July 1995. ACM Press.
- [OS95] V. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, September 1995.
- [Sch93] P. Schauble. SPIDER: A multiuser information retrieval system for semistructured and dynamic data. In R. Korfhage, E. Rasmussen, and P. Willett, editors, *Proceedings of the sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 318–327. ACM Press, June-July 1993.
- [SFA97] M. J. Swain, C. Frankel, and V. Athitsos. WebSeer: An image search engine for the world wide web. In *IEEE Computer Vision and Pattern Recognition Conference (Submitted)*. <http://www.cs.uchicago.edu/swain/pubs/CVPR97Sub.ps>, June 1997.
- [Sha76] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [vR92] C. J. van Rijsbergen. Probabilistic retrieval revisited. *The Computer Journal*, 35(3):291–298, 1992.