# Using Dempster-Shafer Theory of Evidence in Visual Information Retrieval[⋆]

Joemon M Jose, Jan J IJdens and David J Harper[1]

*School of Computer and Mathematical Sciences, The Robert Gordon University, Aberdeen AB25 1HG, United Kingdom.*

**Abstract**

This paper describes the use of the Dempster-Shafer Theory of Evidence to develop a retrieval model for a visual information retrieval system.

**Key words:** AI Applications, Information Retrieval, Uncertainty Handling, User Interfaces.

## 1 Introduction

Information retrieval is defined as the quest to find those information objects relevant to a given information need. The information retrieval process is not unlike a standard approach taken to many problems in artificial intelligence. In information retrieval, one needs to describe or represent the data (knowledge representation), represent the problem and then search the space of possible solutions to find those that satisfy the problem specification. In information retrieval, these three fundamental steps can be described as follows.

**knowledge representation** The individual information objects (generally called *documents*) in the information space (called *document collection*) are characterized (or *indexed*). For example, in text retrieval, where the documents consist of natural language text, this is generally done by taking all words in the text document, removing those that occur too frequently (as they lack discriminatory power), and reducing the remaining words to their word-stem. The aim of the characterization process is to transform the

[⋆] This work is supported by The Principal's Research Fund at The Robert Gordon University.

[1] Email: {j.jose,j.ydens,d.harper}@scms.rgu.ac.uk

documents into such a form that can be handled by the retrieval system, which will have to match the documents to the problem specification.

**problem specification** Generally, it is assumed that the searcher has a specific need for information. This need is a mental state, sometimes referred to as *Anomalous State of Knowledge* or *ASK* [1]. The searcher's information need is subject to changes and can be easily influenced by external factors. In order for an information retrieval system to be able to satisfy the information need, it has to be externalized or expressed by the searcher. In information retrieval, this expression of the information need is called the *query*. The query will have to be represented or characterized in a way similar to the way in which the documents have been characterized. One of the main problems in this stage of the process is that often searchers are not aware of their precise information need until they are presented with documents ("I don't know what I am looking for but I will know it when I see it"), and therefore the specification of the information need might not correspond perfectly to the actual need itself.

**searching the solution space** Documents or information objects can be seen as (partial) solutions to the problem (query) specified by the searcher. It is the task of the retrieval system to determine whether a given document is relevant to the information need. It will have to do this by considering whether the *characterization* of the information objects has any correspondence to (*matches*) that of the query (which, as described earlier, is the *specification* of the information need). To be able to do this, the information retrieval model uses a so called *matching function*. This function takes a characterized object and a query, and establishes whether the object matches the query (and sometimes also attaches a numerical value to the closeness of the match). Different information retrieval systems differ mainly in the way this matching function is designed.

In the field of information retrieval, various well understood and well researched techniques are available for dealing with the problem of text retrieval, where the documents consist of natural language text. As described before, characterization of text documents is relatively straightforward. Also, various methods of query specification exist. The main difference among the various retrieval models is the way the matching function has been implemented. Some retrieval models make a simple binary judgment of relevance (e.g., the Boolean Model). Most others return a list of documents that have been deemed relevant, ranked in decreasing order of perceived usefulness. In this category, the most important systems are Probabilistic Retrieval [8] in which a probability of relevance is calculated for every document in the document collection, and Logical Retrieval [10,11], in which the retrieval system tries to logically deduce that the query (seen as a set of axioms) is a logical consequence of a document, thereby establishing it is relevant to the (specification of) the information need.

When moving from the text retrieval case into domains in which documents are multimedia in nature, the information retrieval problem is complicated considerably. Multimedia documents are made up of various components that potentially can be represented by different media. In this paper, we will look at documents that consist of a text and a visual (image) component. Whereas text retrieval is well understood, extending the domain to also consider a visual component changes all of the steps of the information retrieval process. In the *knowledge representation* stage, one now has to find a way to characterize the visual component so that the essence of the visual data is captured. Similarly, in the *problem specification* stage, the searcher has to be able to externalize their information need in such a way that it can be used for matching against visual data. The problem in the *matching* process is twofold: first, one needs a matching function that can establish the similarity between the specification of the visual component of the query and that of a document. Furthermore, as the documents (and queries) consist of multiple parts that are matched independently, the retrieval model is presented with evidence from multiple sources. Therefore, a suitable evidence combination strategy has to be employed in order to be able to come up with a final relevance judgment.

In the remainder of this paper, we propose a retrieval model for visual information retrieval. This is done by describing a way of characterizing the visual data (as proposed in Jose&Harper [5]), by describing a matching function for both the visual and the textual components of a document, and by proposing a method for implementing the matching process that is based on the use of Dempster-Shafer theory for combination of evidence from the textual and visual retrieval sources.

## 2 Dempster-Shafer Theory and its use in Information Retrieval

### *2.1 Dempster-Shafer Theory of Evidence*

In Dempster-Shafer theory one represents the beliefs in propositions of interest and combines beliefs from multiple sources to reach a common belief. The set of all propositions of interest is called the frame of discernment $\Theta$. Beliefs are assigned to a subset of propositions in the frame of discernment. The subset of propositions for which we assign positive beliefs are called focal elements.

As an example, let $D$ stand for a conclusion or proposition. Suppose one rule or source of evidence implies $D$ with strength 0.8 and another rule implies it with strength 0.9. In the Dempster-Shafer theory of evidence, these strengths are called basic probability assignments (or BPAs). If the focal element is a singleton then the belief and BPA are the same. If the focal element is a set of possibilities, then the belief is sum of BPAs of all the possibilities. The belief in a proposition is represented by a sub-interval [S,P] of the unit interval [0,1]. The lower value S represents the support for that proposition and sets a minimum value for its likelihood. The upper value, P, denotes the plausibility of that proposition and establishes a maximum likelihood. Support may be interpreted as the total positive effect a body of evidence has on a proposition, while plausibility represents the total extent to which a body of evidence fails to refute a proposition. The degree of uncertainty about the actual probability value for a proposition corresponds to the width of its interval.

In Dempster-Shafer theory, a set of propositions in which we are interested is called the frame of discernment $\Theta$. The individual propositions are called focal elements. A Belief function $Bel : 2^{\Theta} \rightarrow [0,1]$ is defined on a frame of discernment and has the following property:

$$Bel(A) = \sum_{x \subseteq A} m(x)$$

where $m$ is the Basic Probability Assignment (BPA) such that $m(\phi) = 0$ where $\phi$ is the empty set and

$$\sum_{P \subseteq \Theta} m(P) = 1$$

where $m(P)$ represents the degree of belief that is exactly committed to P.

### *2.1.1 Combination of Evidence*

The Dempster combination rule aggregates two bodies of evidence defined within the same frame of discernment into one body of evidence. Let $m_1$ and $m_2$ be two bodies of evidence defined in the frame of discernment $\Theta$. The new body of evidence is defined by a belief function $m$ as follows:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \sum_{B \cap C = \phi} m_1(B) \times m_2(C)}$$

## *2.2 The Use of Dempster-Shafer Theory in Information Retrieval*

There have been some efforts to apply the Dempster-Shafer formalism in the information retrieval field. Two different logic-based models of information retrieval have been based on the Dempster-Shafer framework. Both models are directed towards text based retrieval systems. In these models a document is represented by a frame of discernment, and the propositions in this frame represent information items. In a model developed by Silvia and Milidiu [9], the query is also represented as a body of evidence associated with the frame of discernment. A model by Lalmas [6] is based on the principle of transformation. In their model relevance is based on obtaining a transformed document that contains the information need (a principle similar to minimal axiomatic extension).

## 3 Visual Information Retrieval Model Using Dempster-Shafer Theory

We have developed a picture retrieval system that integrates text and picture features for retrieval [5]. The model uses objects in the picture and their locations as image features (viz. spatial features). These features are derived semi-automatically. The image matching considers the spatial similarity between a query object and an image object by using a distance measure. Standard information retrieval techniques are used for text indexing and matching. We applied Dempster-Shafer theory for combining the evidence received from the text matching and the picture matching by considering certainty factors for each component. A prototype system has been implemented and will be briefly discussed in section 4.

In the following, it is assumed that a collection of documents exists that forms the frame of discernment $\Theta$, (i.e. $\Theta = \{d_1, d_2, \ldots\}$). Also, it is assumed that a searcher has expressed an information need as a query, which will be considered as evidence. The documents for which a belief value is available (i.e. there is a value (belief) associated with the proposition that the document is relevant to the query) are considered as the focal elements, (i.e. $\{d_i | m(d_i) \succ 0\}$ ). Different beliefs eminate from the textual and the pictorial components of the document and query and these beliefs are normalised and combined using the Dempster-Shafer mechanism.

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.

**Definition 1 (Document)** *A* multimedia document $\mathcal{M}$ *is a structure* $\mathcal{M} = \langle P, T \rangle$ *where* $P$ *is the visual component of* $\mathcal{M}$*, and* $T$ *is its textual component.*

A query can also be seen as a document: it also has two components, a visual query component and a text query component. In the remainder of this section, the individual retrieval models for pictorial matching and text matching will be described, and an evidence combination method will be discussed.

### *3.1 Belief based on the Visual Component*

In the visual retrieval model, the aim is to match the characterization of the visual component of the query to the characterization of the visual component of the documents. The proposition of interest is that a certain document is relevant to a query based on the spatial features in the characterization of the document. The query is characterized by spatial features. Using these as evidence, a belief is calculated that indicates support for the proposition of interest. To formalize these notions, the following definitions will be used.

**Definition 2 (Region)** *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 3 (Spatial Feature)** *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 4 (Picture)** *A* picture *or visual component $P$ is a set $P = \{\phi\}$ where $\{\phi\}$ is a set of spatial features forming the characterization of the picture.*

**Definition 5 (Distance Measure)** *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 normalizing 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 distance as well as size of each region.

**Definition 6 (Picture Indicator Function)** *Let $\lambda_i$ and $\lambda_j$ be the label components of two spatial features. Then, we can define a* picture indicator function *as follows*

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

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

**Definition 7 (Spatial Similarity)** *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*

$$\mathrm{sim}(P_D, P_Q) = \sum_{\langle\lambda_i,\rho_i\rangle\in P_Q} \sum_{\langle\lambda_j,\rho_j\rangle\in P_D} [I(\lambda_i, \lambda_j) \times D(\rho_i, \rho_j)]$$

The next step is to convert these scores into Basic Probability Assignments (BPAs). Scores are normalised and converted into BPAs by dividing each by the sum of all scores. As an example, consider 5 documents with scores $\mathrm{sim}(d_1)$, $\mathrm{sim}(d_2)$, $\mathrm{sim}(d_3)$, $\mathrm{sim}(d_4)$ and $\mathrm{sim}(d_5)$. Then these can be normalised to get a BPA, say $m(d_i)$, by dividing each $\mathrm{sim}(d_i)$ by $\sum_{i=1}^{5} \mathrm{sim}(d_i)$. That is,

$$m(d_i) = \frac{\mathrm{sim}(d_i)}{\sum_{i=1}^{5} \mathrm{sim}(d_i)}$$

in order to ensure that $\sum_{i=1}^{5} m(d_i) = 1$.

A searcher's confidence in a component of the query can also be incorporated into the model in the following way. The 'confidence' is interpreted as the 'certainty' in a given piece of evidence (e.g. in the spatial component of the query) and hence the 'uncertainty' is 1 - certainty. Now, uncertainty, say $\mu$, can be propagated by assigning a belief $\mu$ to the set of all documents (i.e. frame of discernment $\Theta$). This means that the belief $\mu$ could not be assigned to any 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 $\Theta$. Hence, $m(\Theta) = \mu$. To make $\sum m(d_i) = 1$, we multiply each $m(d_i)$ by $(1 - \mu)$.

### *3.2 Belief based on the Textual Component*

In the case of text matching, the proposition of interest is that a document is relevant to a query based on the text features in them. We take the query features (text features) as evidence, and calculate the belief in the proposition. Before explaining the computation, we need the following definitions.

**Definition 8 (Text Component)** *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 9 (Weighting Function)** *Let $\tau$ be a text feature. Then a 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 10 (Text Indicator Function)** *Let $\tau_i$ and $\tau_j$ be two text features. Then, we can define a* text indicator function *as follows*

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

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 11 (Textual Similarity)** *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*

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

This evidence is also normalised and from this a belief value is computed using the same procedure as used in the pictorial case explained at the end of section 3.1.

### *3.3 Evidence Combination*

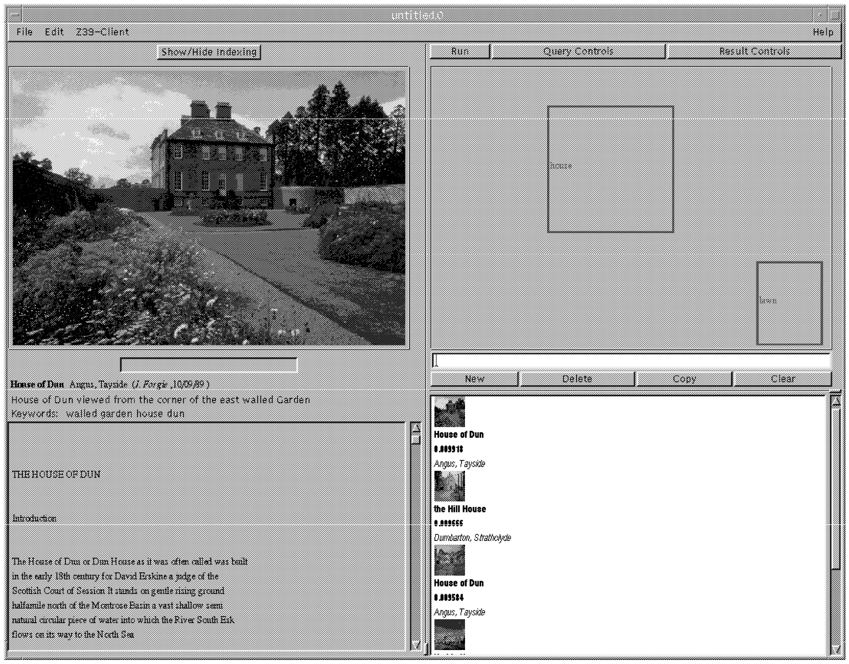
Now component matching functions have been designed for both the textual and the picture component, evidence coming from both matching processes will have to be combined in order to arrive at one overall relevance score for the document containing the various components given a query. For the evidence combination, the Dempster-Shafer evidence combination mechanism is used.

**Definition 12 (Evidence Combination)** *Let $m_s$ be the belief function based on the pictorial evidence and let $m_t$ be the belief function based on the textual evidence. Then the combined evidence is given by the following formula*

$$m(d_i) = \frac{\sum_{d_j \cap d_k = d_i} m_s(d_j) \times m_t(d_k)}{1 - \sum_{d_j \cap d_k = \phi} m_s(d_j) \times m_t(d_k)}$$

Using this mechanism, documents are presented to the searcher in decreasing order of similarity (which, according to the definition of the models will be viewed as decreasing order of relevance).

## 4 Epic: A Prototype Visual Information Retrieval System



**House of Dun** Angus, Tayside (*J. Forgie* ,10/09/89 )
House of Dun viewed from the corner of the east walled Garden
Keywords: walled garden house dun

THE HOUSE OF DUN

Introduction

The House of Dun or Dun House as it was often called was built in the early 18th century for David Erskine a judge of the Scottish Court of Session It stands on gentle rising ground halfamile north of the Montrose Basin a vast shallow semi natural circular piece of water into which the River South Esk flows on its way to the North Sea

**House of Dun**
**0.009918**
*Angus, Tayside*

**the Hill House**
**0.009666**
*Dumbarton, Strathclyde*

**House of Dun**
**0.009584**
*Angus, Tayside*

Fig. 1. The Epic Retrieval System

To be able to experimentally investigate the effectiveness of the visual retrieval model, a prototype implementation of the method has been constructed using *Eclair* [3] , an extensible class library for constructing information retrieval applications. The Eclair library provides abstractions for retrievable objects, indexing, storage, querying etc. Best match and Dempster-Shafer based retrieval models are provided by the Eclair server. The query interface, a snapshot of which is shown in figure 1 has been built using the FireWorks architecture [4]. The prototype is being implemented as a client-server application using the Z39.50 information retrieval protocol.

In *EPIC*, searchers can enter queries consisting of picture components (labeled rectangles) and text components. Searchers can also specify the relative importance they attach to each query component. From this, the *EPIC* system generates spatial query features and text query features as described before, and matches them to the images in the database using the methods discussed above. After this, the component similarity values are combined using the Dempster-Shafer mechanism.

In figure 1, the upper right hand window is the query canvas. Here, searchers can enter the picture component of their query by drawing boxes representing objects they want to appear in the result images, and by labeling them. Furthermore, below the query canvas a field for entering the text query component is provided; after a search has been performed, the lower right hand window displays the list of results of the query. By default, every result is displayed with a thumbnail image. However, the presentation of the result list is fully customizable. The left hand side windows are used for viewing documents from the result list. In the upper pane a full-sized view of the picture part of a retrieved document is displayed. The lower pane contains any textual information associated with the image.

Lansdale et.al. [7] have described the need for a spatially depicted interface for visual information retrieval. Their initial experiments have shown that a spatially depictive interface to visual collections will enhance the retrieval. Enser [2] has shown that when searchers seek visual information they have a mental model of the image they want. The query canvas is designed to enable a searcher to capture this mental model of her information need. If the searcher has a mental model of their information need a tool which allows them to express this will reduce the uncertainty in the query formulation process (this uncertainty is one of the major problems in information retrieval, and in the IR field one of the basic assumptions that has to be made is that the query is a perfect specification of any underlying information need that caused the searcher to issue the query). Also, an interface that is expressive enough will give a searcher an opportunity to reflect on her information need and modify it as required. More information on spatial features and their use in visual information retrieval can be found in Jose&Harper [5].

To enable us to do user experiments, we have created a picture collection of 800 photographs from the National Trust for Scotland (NTS) photographic archive. In this collection, each photograph is stored along with some associated text related to the contents of the photographs (this text was taken from various publications from the NTS).

## 5 Conclusion

In this paper, we have introduced a new strategy for picture retrieval. The approach presented integrates evidence obtained from the textual component(s) and the picture component(s) of a document for document retrieval. Well-known text retrieval techniques are applied to the text component of the document. Picture features are used for matching the image components of a document with those of the query. The picture features represent objects and their spatial location. A Dempster-Shafer based retrieval model, which combines the evidence from the text component and the picture component is given. The similarity measure discussed considers the relative importance of the various components of the query and the document.

A prototype visual retrieval system called *EPIC* has been implemented and will be used in user evaluations of the system.

## Acknowledgement

We are much indebted to Professor Keith van Rijsbergen for inspiring the development of this model and for valuable discussions. We would also like to thank David Hendry for being a constant source of ideas and criticism. Finally, thanks should go to the National Trust for Scotland. The work presented here has been supported by the Principal's Research Fund at the Robert Gordon University.

## References

[1] N. J. Belkin, R. N. Oddy, and H. M. Brooks. ASK for information retrieval. *Journal of Documentation*, 38:61–71,145–164, 1982.

[2] P. G. B. Enser. Query analysis in a visual information retrieval context. *Journal of Document and Text Management*, 1(1), 1993.

[3] 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.

[4] D. G. Hendry and D. J. Harper. An architecture for implementing extensible information-seeking environments. 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*, pages 94–100, August 1996.

[5] 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.

[6] M. Lalmas. Modelling information retrieval with dempster-shafer's theory of evidence: A study. Technical report, Department of Computer Science, University of Glasgow, 1996.

[7] M. 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.

[8] S. E. Robertson and K. Sparck-Jones. Relevence weighting of search terms. *Journal of American Society for Information Science*, 27(3):129–146, 1976.

[9] W. Teixera de Silva and R. L. Milidiu. Belief function model for information retrieval. *Journal of the American Society of Information Science*, 4(1):10–18, 1993.

[10] C. J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6):481–485, 1986.

[11] C. J. van Rijsbergen. Towards an information logic. In *Proceedings of the Twelfth ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 77–86, 1989.