

18 Searching Distributed and Heterogeneous Digital Media: The VisualHarness Approach¹

Amit Sheth, Kshitij Shah, Krishnan Parasuraman,
Srilekha Mudumbai

Large Scale Distributed Information Systems (LSDIS) Lab
Computer Science Department, University of Georgia,
415 GSRC, Athens GA 30602-7404 USA
amit@cs.uga.edu, <http://lsdis.cs.uga.edu>

18.1 CHALLENGES IN RESOURCE DESCRIPTION AND DISCOVERY FOR DIGITAL MEDIA

Current Web-based search engines do a reasonable job when dealing with primarily textual and in some cases semi-structured data. Web-based interfaces to traditional databases also allow us to exploit structured databases. However, there is little help when we wish to find relevant information in visual and so-called new digital media (with significant temporal and spatial components), which are being put on the Internet at an exponential rate. Today, usually a user needs to know the specific repository and use the specific access mechanisms and interfaces that have been provided by the repositories of such information. There is also little help when we wish to deal with a broad variety of heterogeneous media. What if we don't know which type of digital media we need to target the query to? What if the response to the query is best satisfied by a collection of artifacts of different media? Thus one challenge in the current Internet environment is to have the ability to search and access information in various media, including visual media.

The second challenge is to have different ways of describing information of interest. The current Web-based search engines do reasonably well to exploit keyword-based search techniques, primarily derived from the field of

¹ Research reported here is funded in part by "InfoHarness: A System for Scalable Search of Heterogeneous Information" in the Massive Digital Data initiative managed by the Office of Research and Development under contract No. 95 - F138400 - 000.

information retrieval. But, any user is well aware of the lack of precision, selectivity or quality of results, and the resulting information load when too many hits are returned. The users' own lethargy in picking good keywords and a good set of keywords connected using advanced query options is partly to blame. The concept-based searches that exploit statistical techniques, the human-developed concept hierarchies, and information categorization, provide good improvements in certain situations. However, additional techniques involving ontology supported attributed search and content-based search techniques need to be exploited, especially when dealing with new media types.

The third challenge is to be able to have semantic-level modeling of related and relevant information. How do we describe semantically related distributed and heterogeneous information (e.g., a "person portfolio" consisting of his structured database records, his/her articles or articles about him/her in semi-structured form, software that s/he has created, his/her photograph, and a video of his/her talk)? How do we describe information at a logical or semantic level that may be independent of media or may involve multiple media? How do we describe the context of his/her information search such that the system can distinguish between medical instruments when the term instrument appears in his/her portfolio?

This paper deals with some of the aspects of the first two challenges. It presents the VisualHarness system developed at the LSDIS lab that supports a customizable search involving keywords, attributes and (visual) content of heterogeneous data (currently text, structured databases and image repositories). The third challenge is being addressed in our InfoQuilt project. Section 18.2 presents the background, while section 18.3 discusses metadata that play key roles in the VisualHarness system. Section 18.4 presents a brief overview of the VisualHarness including its extensible architecture. Section 18.5 presents the novel black box approach for supporting content based access of images using third party visual information retrieval engines. Section 18.6 provides a summary and outlines some of the ongoing research. The Appendix at the end presents early results showing efficacy of the novel black box approach when using a naïve strategy.

18.2.1 REQUIREMENTS FOR RESOURCE DESCRIPTION AND DISCOVERY

The challenges discussed earlier translate into the following set of emerging requirements for Web-based information searches (as well as associated information management functions involving access, filtering and integration):

- support for heterogeneous digital media,

- support for complimentary access strategies involving keyword-, attribute- and content-based access, preferably with an ability to combine these components with different weights, or to use them iteratively,
- specification of media independent queries,
- specification of information correlation at a logical level, and
- support for semantics, including context of a query and available information, possibly supported by use of ontologies and profiles

A key to our approach is to exploit the metadata to a much fuller extent than was perhaps done before. The approach involves

- identification of a broad variety of metadata,
- extraction of these types of metadata,
- logical correlations involving any type of metadata (for different types of media) [Sheth and Kashyap 96, Shah and Sheth 98], and
- specification of customizable (with different weights or sequences) information requests, involving keyword-, attribute- and content-based engines, and associated metadata-driven information request processing

The run-time system exploits and adapts a number of well-known and emerging technologies, including

- modified Web-server technology to utilize the broad variety of metadata in processing information requests,
- use of multiple (third-party) indexing techniques for textual data as well as other digital media,
- database management systems to manage metadata, and
- object-oriented modeling and software management, distributed object management

Our system that supports the above is called the VisualHarness system. Although it shares a number of similar features and capabilities with other contemporary systems that support integration of heterogeneous information, such as GARLIC [GARLIC], HERMES [HERMES], InfoHarness [I-HARNESS], Information Manifold [I-MANIFOLD], InfoSleuth [I-SLEUTH], SIMS [SIMS], TSIMMIS [TSIMMIS], and several others, it also has a number of differences. The features we focus on in this paper that are by and large different or unique are (a) use of a broader variety of metadata, (b) support for multiple third party indexing including a black-box approach to adapting visual information retrieval engines to capture metadata from visual information, and (c) integrated and customizable support for keyword, attributed *and* content-based access to distributed and heterogeneous information. While individual techniques used are not novel, except for the

black-box approach to using visual information retrieval engines, their integration and the issues related to integration are what we believe is novel.

A part of this paper focuses on the content-based retrieval of images. Here the related work includes (a) systems for content-based retrieval called Visual Information Retrieval (VIR) systems such as Virage's VIR [Gupta 95], MIT Media lab's Photobook [Pentland et al 96], IBM's QBIC [Ashley et al 95], among several others, and (b) content-based image retrieval using metadata and relaxation techniques [Chu et al 98]. VisualHarness system complements and significantly extends the current VIRs to support a broader search strategy over a broader variety of distributed and heterogeneous data. In fact, for its black-box approach it uses Virage's VIR technology [Virage] for computing the image properties. Our approach can be seen as a metadata based approach that is extensible such that any metadata type for multiple domains can be obtained using extractors, the corresponding access method for that metadata can be supported, and different access strategies can be combined to achieve better quality results. More detailed comparison can be found in [Mudumbai 97].

18.3 METADATA FOR HETEROGENEOUS DIGITAL MEDIA

Metadata represent information about the data. Metadata can be regarded as an extension (albeit a significant one) of, the concept of a schema in structured databases. They may describe, or be a summary of the information content of the individual databases in an intentional manner. They typically represent constraints between the individual media objects that are implicit and not necessarily represented in the databases themselves. Some metadata may also capture content-independent information like location and time of creation. Examples of what we consider media types are structured data (data in relational or object-oriented databases), textual data (of different formats, such as Word files, source code, etc.), images (of possibly different modalities such as X-Ray, MRI scan), audio (of possibly different modalities such as monaural, stereophonic) and video.

Although there are a number of ways to classify metadata (e.g., see [Boll et al 98] for a more detailed discussion), the criteria we use to classify the metadata [Kashyap et al 95] is the extent to which they are successful in capturing the (data and information) content of the artifacts or documents represented in various media types. The level of abstraction at which the content of the documents is captured is very important. We believe that to capture the semantic content (i.e., at a level of abstraction closer to that of humans), it is important for the metadata to model application domain-specific information.

Thus, we classify the various kinds of metadata depending on whether they are based on the (data or information) content of the artifacts/ documents or not. The basic kinds of metadata we identify are:

- content-dependent metadata,
- content-descriptive metadata, and
- content-independent metadata.

Content-dependent metadata, as the name suggests, depends only on the content of the original data. A text index, like the document vectors in the LSI [Deerwester et al 90] index and the complete inverted WAIS [Kahle and Medlar 91] index (among many others) are examples of metadata that is determined by the content, i.e. the frequency and position of text units in the document. This kind of metadata is referred to as content-dependent metadata for textual data. When we associate metadata with the original data, which describes the contents in some way, but cannot be extracted automatically from the contents themselves, we call it content-based metadata. This kind of metadata could be determined exclusively by looking at the content, or is derived intellectually by automatic or semi-automatic means. However, it could not have been derived on the basis of content alone.

Table 18.1: Metadata for different digital media²

<i>Metadata</i>	<i>Data Type</i>	<i>Metadata Type</i>
Q-Features [Jain and Hampapur]	Image, Video	Domain Specific
R-Features [Jain and Hampapur]	Image, Video	Domain Independent
Meta-Features [Jain and Hampapur]	Image, Video	Content Independent
Impression Vector [Kiyoki et al.]	Image	Content Descriptive
NDVI, Spatial Registration [Anderson and Stonebraker]	Image	Domain Specific
Speech Feature Index [Glavitsch et al.]	Audio	Direct Content Based
Topic Change Indices [Chen et al.]	Audio	Direct Content Based
Document Vectors [Deerwester et al.]	Text	Direct Content Based
Inverted Indices [Kahle and Medlar]	Text	Direct Content Based
Content Classification Metadata [Bohm and Rakow]	MultiMedia	Domain Specific
Document Composition Metadata [Bohm and Rakow]	MultiMedia	Domain Independent
Metadata Templates [Ordille and Miller]	Media Independent	Domain Specific
Land Cover, Relief [Sheth and Kashyap]	Media Independent	Domain Specific
Parent Child Relationships [Shklar et al.]	Text	Domain Independent
Contexts [Sciore et al., Kashyap and Sheth]	Structured	Domain Specific
Concepts from Cyc [Collet et al.]	Structured	Domain Specific
User's Data Attributes [Shoens et al.]	Text, Structured	Domain Specific
Domain Specific Ontologies [Mena et al.]	Media Independent	Domain Specific

Content-descriptive metadata can be both domain-dependent and domain-independent. Domain-dependent metadata uses domain-specific concepts as a basis to determine the actual metadata created. Domain-independent metadata, on the other hand, relies on no such domain-specific concepts. A typical

² The citations in the table appear in [Klas and Sheth 94, Sheth and Klas 98].

example of a domain-independent metadata would be the one that describes the structure of a multimedia document.

Content-independent metadata, on the other hand, does not depend on the content. This kind of metadata can be derived independently from the content of the data. This is like attaching a tag to the data irrespective of the data contents. Examples of content-independent metadata of a document are its date of creation and location. Table 18.1 lists a few examples of different kinds of metadata.

18.4 VISUALHARNESS

The VisualHarness system is aimed at providing rapid access to huge amounts of heterogeneous data available over the World Wide Web. Our work so far has explicitly dealt with repositories of a variety of textual data, relational databases, and images, although this paper primarily focuses on the support for image data by this system. The VisualHarness system uses the components of the InfoHarness (IH) [Shklar et al 95, I-HARNESS] platform that supports textual data, with extensions to deal with visual data (using the ZEBRA system for supporting image data). Its metabase, that is the database of metadata, can consist of indices (e.g., full text index for textual data, feature based index for image data), and attribute value pairs used to support attribute based access as well as content-based access of visual data using a novel black-box approach that converts feature vectors into structured metadata. The VisualHarness system maintains metadata about the information space without restructuring, reformatting or relocating the original information enabling access to information by logical units of interest. An object-oriented layer (using InfoHarness Objects - IHOs) supports logical structuring of the metadata objects and thus allows arbitrary relationships amongst the represented information artifacts. VisualHarness is built using the IH integration platform. The VisualHarness system architecture is open and extensible. It provides hooks using which different third party indexing engines for textual data, and third party VIR engines, such as the Virage's VIR for image content based access, can be integrated.

Figure 18.1 shows a high level view of the VisualHarness architecture. VisualHarness supports comprehensive querying that is performed as follows. The IH server accepts a user query as a client request from a browser. The Query Engine module of the Query Processing Unit (QPU) creates subrequests for the relevant search components. The search components use metadata—both precomputed and stored in metabase, as well as that computed at run-time, to determine references to the relevant data and provide them to the result composition module of the QPU. QPU performs normalization, rescaling and formatting of the result. The IH server then

displays the result to the user. When the user selects one or more data objects to be displayed, the IH server accesses the appropriate repositories directly to retrieve data. Table 18.2 lists some sample image metadata used by the VisualHarness system. Similar metadata classification can be given for other digital media accessible by the system.

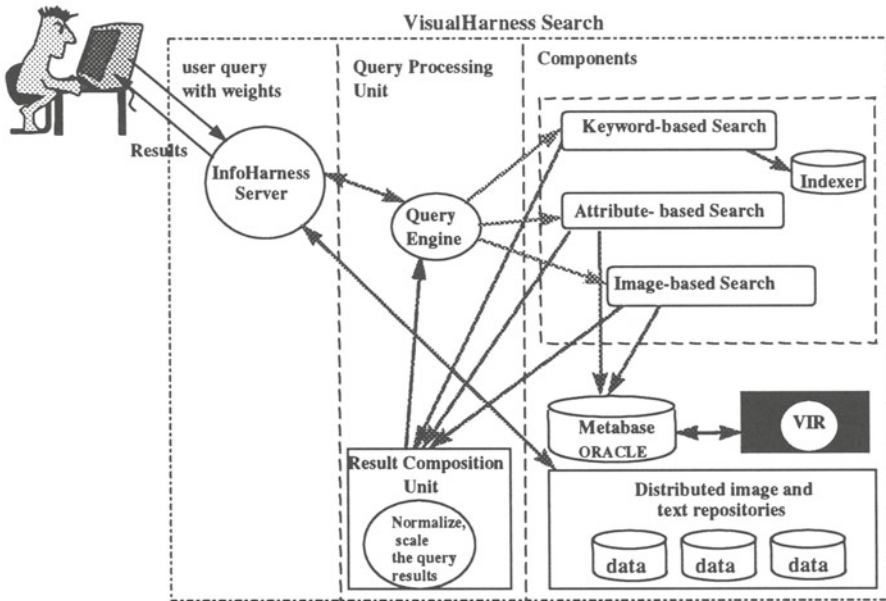


Figure 18.1: The VisualHarness architecture

The query processing subsystem uses weighting strategies to provide a scalable approach. By scalable approach, we mean that a user can assign different weights to different properties on which the similarity is based. Information retrieval from the database is restricted according to the user assigned weights. For example, if we have three properties say, P_1 , P_2 , and P_3 , supported by the VIR engine, then the user can assign different weights i_1 , i_2 and i_3 to each of these properties so that the retrieval by the VIR is based on

$$i_1P_1 + i_2P_2 + i_3P_3 \text{ where } 0.0 \leq i_1, i_2, i_3 \leq 1.0$$

The resulting values are normalized and scaled in order to give a ranking to each of the objects retrieved from the database.

Property weights in the VisualHarness system refer to the user weights assigned for different properties of the image via the user interface. The property weights vary between 0 and 1.0. If O_1, O_2, \dots, O_n are the objects in the image database, P_1, P_2, \dots, P_n are the different properties supported for an object O_i and Q is the input query object, the score S , obtained for each retrieved object O_i for the user assigned property weights i_1, i_2, i_3 and i_4 would be

$$S = i_1z_1 + i_2z_2 + i_3z_3 + i_4z_4$$

where $z_1 = \text{abs}(P_1 \text{ value of } O_i - P_1 \text{ value of } Q)$ and similarly for z_2, z_3 and z_4 . Scaling of property weights is done by multiplying the property weights into the appropriate difference in the property values z_1, z_2, z_3 and z_4 . The normalization is performed by giving the highest ranking with value 1.0 to the object that has the highest score. For all other objects retrieved, we normalize them by dividing the score of that object by the score of the highest ranked object (prior to giving it the value of 1.0). This gives the overall ranking of the objects that are retrieved from the image database.

Table 18.2: Some of the image feature metadata in the VisualHarness system

Metadata classification	Requires processing raw data	Uses semantic (domain-specific) knowledge	Typical Format	Examples from image data
Content-independent	No	No	attribute values	height, width, size, date of creation etc.
Content-dependent	Yes	No	feature vectors or attribute values	color, composition, texture, structure etc.
Content-based	No	Yes	attribute values	color, height, fragrance, hybridizer of a flower; model, category etc. of an aircraft.
Content descriptive	Maybe	Either	text and/or keywords	Descriptions of a flower, an x-ray, an air-craft (domain-specific); General descriptions related to the image (domain-independent)

Figure 18.2 shows an example of the comprehensive search screen. A user could focus on any one of the three search strategies, keyword-based, attribute-based and content-based, or combine the three using relative weights. Within the content-based search, it is possible to use additional features, such as color, structure, texture and composition in the case of images, as shown in Figure 18.2. Iterative refinements are also possible.

The access method mentioned above will apply to any of the VIR engines if one has knowledge about, and access to, the feature vectors of the image objects in the database. For systems such as the VisualHarness system that do not know about the internals of the VIR engine, this access method might not be applicable. An alternative, in cases where we either do not have access to the actual feature vectors or have no way of interpreting them, is the Black Box Approach (BBA) originally introduced in [Shah et al 97]. In this we try to compare the objects based on their differences with a reference image rather than a direct comparison between the objects themselves. We now describe

the original BBA that involved use of what are terms as null images as reference images [Shah et al 97], followed by our recent improvements.

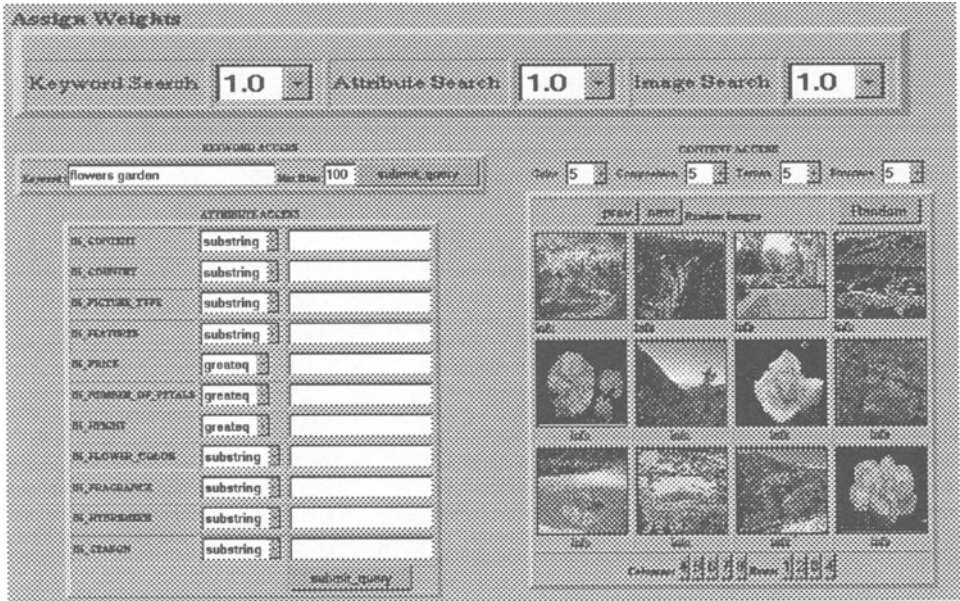


Figure 18.2: Comprehensive search in VisualHarness

18.5 THE BLACK BOX APPROACH FOR CONTENT-DEPENDENT METADATA

Feature vectors from an image refer to the features extracted from different topological spaces. Distances between the objects and the input query object are required in order to obtain a ranking of the objects similar to the given query object. As discussed above, since the VisualHarness system does not have access to the actual feature vectors of an image object or has no way of interpreting them, it uses the VIR engine as a black box. The BBA tries to compare objects based on their differences with a reference image rather than a direct comparison between the objects themselves.

If R is a reference image and the objects in the database are O_1, O_2, \dots, O_n then the feature distance

$$D(O_1, O_2) = \text{abs}(D(O_1, R) - D(O_2, R))$$

i.e., the distance between any two objects O_1 and O_2 in the feature space would be equal to the absolute value (or the Euclidean distance) of the difference between each object compared with the reference image for a particular property. Feature vectors of the object sequence in the database

based on different properties of an image are mapped to a point in the feature space; a query with tolerance e becomes a sphere of radius e . This process is shown in Figures 18.3 and 18.4.

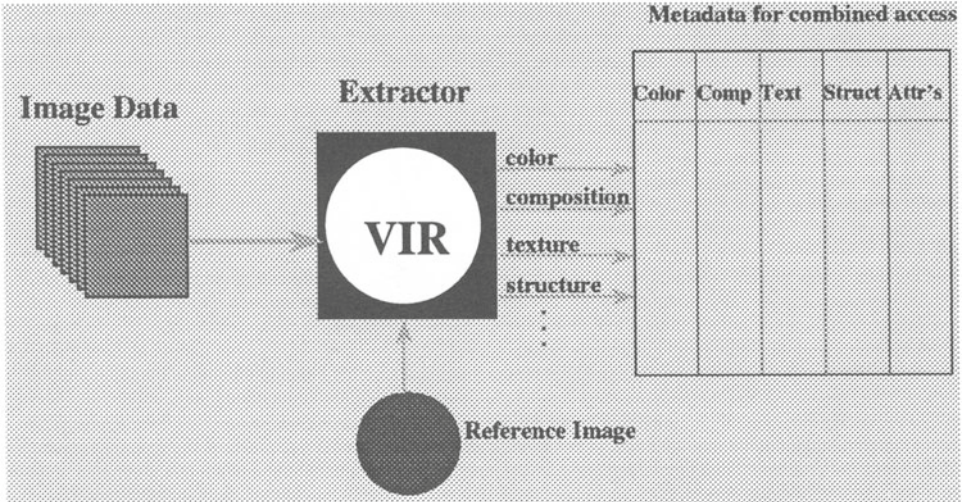


Figure 18.3: Image feature extraction in VisualHarness' Blackbox approach using a reference image

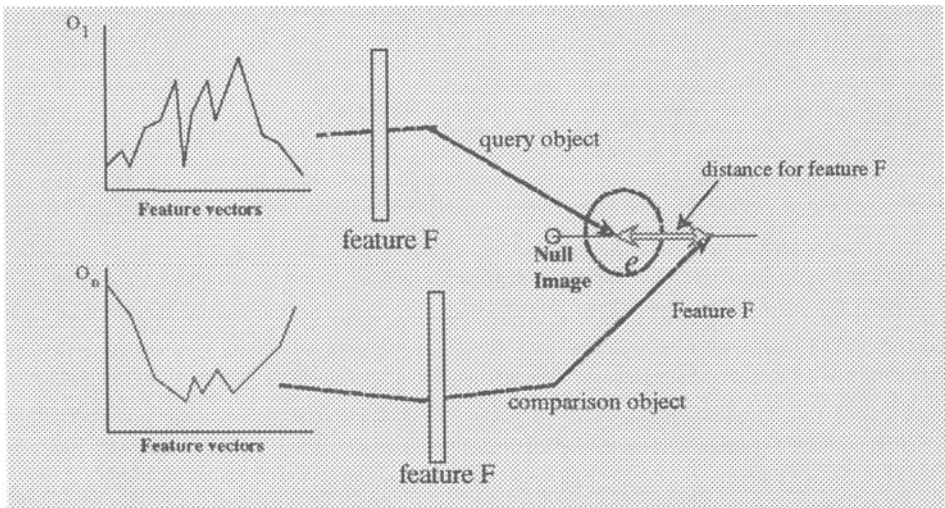


Figure 18.4: Translating n dimensional image feature vector spaces to Euclidean distances

The BBA allows the VisualHarness system to be very scalable since the information retrieval is not limited to a particular VIR engine and its corresponding image database. Any engine can be hooked up into our search system and multiple databases at different locations can be accessed. Runtime

computation is not expensive as we pre-compute the distance between each object and the reference image for each of its properties and store it in a database. Run time computation basically involves retrieving the appropriate results from the database by converting the user query image, Q , into a database query $D(Q,R)$. Without this approach, we would have to compute the distance between the query image and each image object in the databases during runtime in a sequential manner. This would be computationally very expensive. With our BBA, we can also employ different weighting strategies to combine the distances obtained in comparing each object with the reference image in that topological space. We can also try and combine features computed using different engines since we are using normalized distances. Our ongoing work studies the effectiveness of the black-box approach. The early results using a naïve strategy of using null images as reference images are reasonably good as compared to the VIR's results as the target set in most situations as discussed in the next subsection and as demonstrated in the Appendix.

18.5.1 Selecting a Reference Image

Our initial BBA strategy was to use a *null image* as a reference image. We chose an entirely black or an entirely white image as a null image, hypothesizing that such an image does not have any specific feature and hence no properties of its own. Using this naïve choice for a reference image, we received quite decent results as shown in the appendix. However, we continued to look for a better strategy for choosing a reference image so as to obtain more accurate results compared with the ones obtained by directly using the VIR engine.

18.5.2 Problems with Null image and the strategy based on centroid of the feature space

Consider three image objects O_1 , O_2 and O_3 in a feature space. Let O_{null} be the null object and d_1 , d_2 and d_3 be the respective distances between the objects and the null image. Figure 18.5 represents relative positions of the objects in the feature vector space. The feature distance between O_1 and O_2 is computed as:

$$D(O_1, O_2) = \text{abs}(D(O_1, O_{null}) - D(O_2, O_{null})), \text{ i.e., } D(O_1, O_2) = \text{abs}(d_1 - d_2)$$

The spatial distribution of the objects in the feature space suggests that O_1 is closer to O_2 than O_3 . The VIR would have given us $D(O_1, O_2) < D(O_1, O_3)$. The distance between O_1 and O_2 is less than the distance between O_1 and O_3 and this would have led one to conclude that O_1 is more similar to O_2 than O_3 .

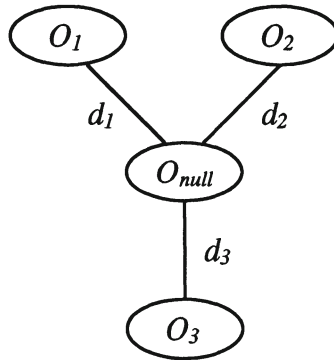


Figure 18.5: Objects in Feature Space.

Suppose the VIR had given distances $D(O_2, O_{null}) < D(O_3, O_{null})$ i.e., $d_2 < d_3$. If $d_1 > d_3$, then using the BBA, we would have concluded that the $D(O_1, O_2) > D(O_1, O_3)$. This is contradictory to VIR results! This shows that the null image may not be the ideal reference image. Theoretically a null image should be devoid of any features, but in practice even a plain black or plain white image has certain inherent features, which adds a certain amount of bias to the feature space.

In this case the null object was more or less a random object. The null object was not a part of the initial object collection and it was added such that it was at a random location in the feature space. As it was shown earlier, this might lead to contradictory results. In our second strategy, a better reference image, rather than being a random object, is the "centroid" in the feature space. Ideally it should be equidistant from all the other objects. But such an object would be difficult to construct, so we choose an existing object that is close to the "centroid" (see Figure 18.6).

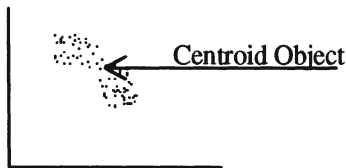


Figure 18.6: Centroid Object in Feature Space.

Finding an existing object that has the minimum distance from every other object is computed as follows:

Suppose there are n objects in the database, $O_1, O_2, O_3, \dots, O_n$ then $D(O_i, O_j)$ is the distance metric returned by the VIR for objects O_i and O_j . For

two objects to be similar $D(O_i, O_j)$ has to be minimum and to find the reference object our objective is to minimize the difference metric.

We calculate $D(O_i, O_j)$ for all objects O_i and O_j in the database and then SD_i , the standard deviation of an object O_i with respect to every other object in the database. $\min(SD)$ returns the object with the minimum standard deviation, which would be our reference image. A potential weakness of this strategy is the cost of recomputing the reference image every time the image database changes.

We have compiled quantitative data using the centroid based reference object that shows noticeable improvement over the null image strategy. These results, while not included here, due to time and space limitation, will be presented at the next available opportunity and at our Web site. Next we discuss a more speculative strategy that is currently being investigated, and for which quantitative results have not been obtained.

18.5.3 Reference image based on semantic correlation of objects

Our objective in this third strategy is to improve the quality of results by semantically correlating various objects into semantic groups. Members of such a group would have some binding feature and objects could belong to multiple semantic groups (i.e., we could "thread" objects based on some predefined semantics). By semantically correlating the objects we make an effort towards trying to better understand the intent of the user submitting the query.

18.5.3.1 Correlation Strategies

We outline two methods that can be use to correlate the objects. These are content semantics and context semantics. Grouping based on content semantics is purely based on statistical principles and can be mathematically formulated whereas context based grouping might be automated, manual or knowledge driven.

18.5.3.2 Content semantics

In content semantic grouping, the objects are correlated or grouped together based on their contents. We extract n features from an image object and map it on to n -dimensional feature spaces as points. If we then run any standard clustering algorithms on these points, we would be able to group the points into *clusters* (see Figure 18.7). Each member in a cluster would be correlated with one another based on content semantics.

Objects could also be correlated based on context. For instance, the objects that represent similar context could form one semantic group. As the simplest

case we could use some qualifying attribute of the object and group all objects having a common value for that object. We could also use an ontology and group all objects with terminological differences resolved for a particular attribute value or user domain knowledge, and group object within the same domain.



Figure 18.7: Clustering of objects in the Feature space

The query processing that involves semantic correlation involves the following strategy. Information retrieval from the database is restricted according to user assigned weights. If the VIR supports three properties, say P_1 , P_2 and P_3 , then the user can assign different weights, w_1 , w_2 and w_3 , to each of these properties and the overall weight of the retrieval w would be

$$w' = w_1P_1 + w_2P_2 + w_3P_3 \text{ where } 0.0 \leq w_1, w_2, w_3 \leq 1.0$$

Apart from assigning individual weights to properties, the user can also assign a value f , which is the scaling factor. The scaling factor imparts relevance to the semantic groups.

All the objects retrieved from the query object's semantic group have their overall weight w' multiplied by the scaling factor. If the scaling factor f had a value of 1.0, all the objects in the collection, whether they belong to semantic groups or not, will have the same relevance. If the scaling factor is greater than 1.0, then the objects in the querying object's semantic group have a higher preference.

As indicated earlier, quantitative evaluations for a semantic correlation based approach are yet to be performed.

18.6 SUMMARY

In this paper we discussed a metadata-based approach to search heterogeneous digital media accessible on the Internet and Intranets. Through the extensive use of metadata, we can support keyword-based, attribute-based and content-based searches, as well as their combinations. The content based search is particularly facilitated by the black-box approach that can use third party engines to deal with specific types of digital media, coercing them to extract

metadata. Once we are able to extract the metadata, the Web-based VisualHarness server can provide access to any Web-accessible data using any combination of the three search alternatives and allow us to combine access to data of heterogeneous media. Extensible metabase and open architecture allow for adding or extending metadata extractors at any time for existing and new media types, allowing the system to grow with new data, new types of media and new ways of processing data of different media. The second version of VisualHarness is implemented in Java.

While this paper focused on image data, we have used a similar approach and architecture to support access to textual, structured and video data. A related project at the LSDIS lab, VideoAnywhere, focuses on a variety of video data and can be seen as a Web search engine for video data.

Another relevant on-going work at the LSDIS lab is on supporting the concept of information correlation involving heterogeneous digital media (called MREF) [Sheth and Kashyap 96] and corresponding information request processing techniques. RDF and XML have been exploited for MREF representation and implementation [Shah and Sheth 98].

References

- Ashley, J., Flickner, M., Hafner, J., Lee, D., Niblack, W., and Petkovic, D. (1995). The Query By Image Content (QBIC) System. *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, San Jose, CA.
- Boll, S., Klas, W., and Sheth, A. (1998). "Overview on Using Metadata to Manage Multimedia Data," in [Sheth and Klas 98].
- Chu, W. W., Hsu, C., Jeong, I. T., Taira, R. (1998). "Content-Based Image Retrieval Using Metadata and Relaxation Techniques," in [Sheth and Klas 98].
- Deerwester, S., Dumais, S., Fumas, G., Landauer, T., and Hashman, K. (1990). "Indexing by Latent Semantic Indexing," *Journal of the American Society for Information Science*, 41(6).
- The GARLIC system, <http://www.almaden.ibm.com/cs/showtell/garlic>
- Gupta, A. (1995). *Visual information retrieval: A Virage perspective*. Technical Report, Virage, San Mateo, CA.
- The Hermes system, <http://www.cs.umd.edu/projects/hermes>
- The InfoHarness and VisualHarness systems, <http://lsdis.cs.uga.edu/proj/proj.html>
- The Manifold system, <http://www.research.att.com/~levy/imhome.html>
- The InfoSleuth system, <http://mcc.com:80/projects/infosleuth>
- Kahle, B. and Medlar, A. (1991). "An Information System for Corporate Users: Wide Area Information Servers," *Connexions - The Interoperability Report*, 5(11), November.
- Kashyap, V., Shah, K., and Sheth, A. (1995). "Metadata for building the MultiMedia Patch Quilt," *Multimedia Database Systems: Issues and Research Directions*, S. Jajodia and V.S.Subrahmanian, Eds., Springer-Verlag.

- Kashyap, V. (1997). *"Information Brokering over heterogeneous digital data : A metadata based approach,"* Doctoral thesis, Dept. of Computer Science, Rutgers University.
- Klas, W. and Sheth, A. Eds., (1994). *Metadata for Digital Media, Special Issue of SIGMOD Record*, 23 (4), ACM Press, December.
- Mudumbai, S. (1997). *ZEBRA Image Access System: Customizable, Extensible Metadata-based Access to Federated Image Repositories*. M.S. Thesis, LSDIS lab, Computer Sc. Dept., Univ. of Georgia, May.
- Pentland, A., Picard, R.W., and Sclaroff, S. (1996). *Photobook: Content Based Manipulation of Image Databases*, Chapter 2, Kluwer Academic Publishers.
- Shah, K., Sheth, A., and Mudumbai, S. (1997). *"Black Box Approach to Image Feature Manipulation used by Visual Information Retrieval Engines,"* The Second IEEE Metadata Conference, September.
- Shah, K. and Sheth, A. (1998). *"Logical Information Modeling of Web-accessible Heterogeneous Digital Assets"*, Proc. of the Forum on Research and Technology Advances in Digital Libraries (ADL'98), Barbara, CA, April.
- Sheth, A. and Kashyap, V. (1996). *"Media-independent Correlation of Information: What? How?"* Proceedings of First IEEE Metadata Conference, April.
- Sheth, A. and Klas, W., Eds., (1998). *Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media*, McGraw-Hill.
- Shklar, L., Sheth, A., Kashyap, V. and Shah, K. (1995). *InfoHarness: Use of Automatically Generated Metadata for Search and Retrieval of Heterogeneous Information* in Proceedings of CAiSE-95, June.
- The SIMS system, <http://www.isi.edu/sims/>
- The TSIMMIS system, <http://www-db.stanford.edu/tsimmis>

Appendix A: Results when using a null image based BBA for content-based image data access in the VisualHarness system

We show two kinds of results based on the null image strategy³. Even though this is a naïve strategy, the results in many cases have been good. One set of results deals with content-based retrieval focusing on the BBA retrieval compared to Virage's VIR. The second set of results shows the refinement of user queries using the combination of different access strategies to obtain better quality results.

A1: Content-based Retrieval

The first image in the sequence is the input query image. Results have been tested on the four properties of the image- color, composition, texture and

³ Black and White images in the printed version of the paper may not demonstrate color-sensitive results adequately. An on-line version of the paper in color is available from the LSDIS lab's library on the Web.

structure. So far, the results have been promising and additional work with larger image repositories is in progress. The null image used for achieving the following results is a full white image.

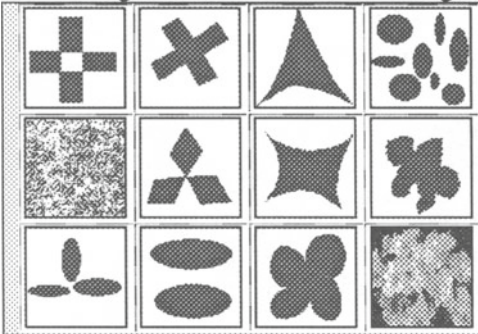


Figure A-1
COLOR (HR = 91.7%)

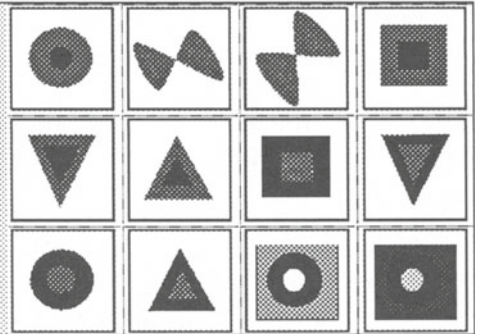


Figure A-2
COMPOSITION (HR = 75%)

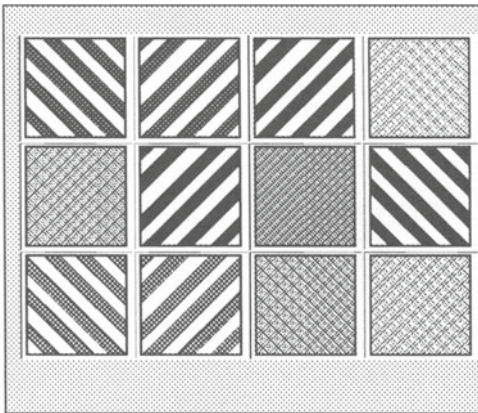


Figure A-3
TEXTURE (HR = 100%)

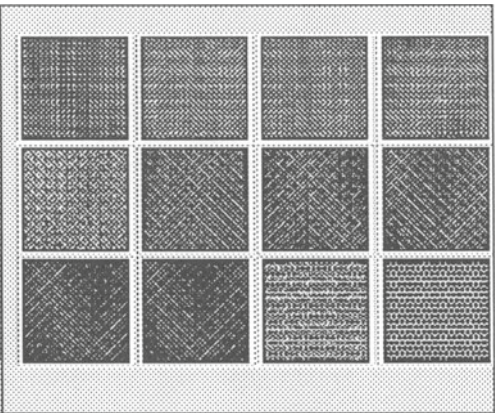


Figure A-4
TEXTURE (HR = 91.7%)

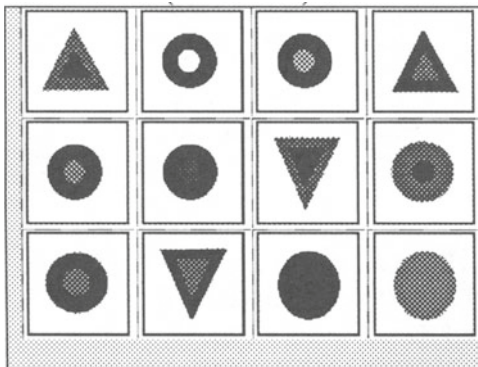


Figure A-5
STRUCTURE (HR = 91.7%)

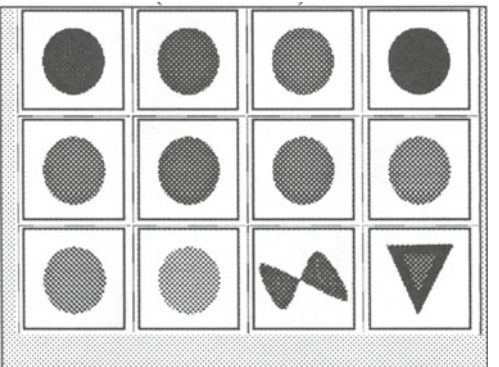


Figure A-6
STRUCTURE (HR = 83.3%)

To compare our results with the VIR, we calculate a hit ratio which is the proportion overlap between the result set returned by VisualHarness and the VIR for n top hits. If A is the set of top n hits returned by VisualHarness and B is the set of top n hits returned by the VIR (Virage in this case) then:

$$\text{Hit ratio (HR)} = ((A \cap B) / n) \text{ where } n \text{ is 12 in our examples.}$$

A2: Combination of keyword, attribute, and image based access strategies

We now discuss some results that involve combining image based and attribute based access strategies. First a random set of images is chosen from a given collection and an icon click is done on the 10th image in the given set with all image properties having equal user weights.

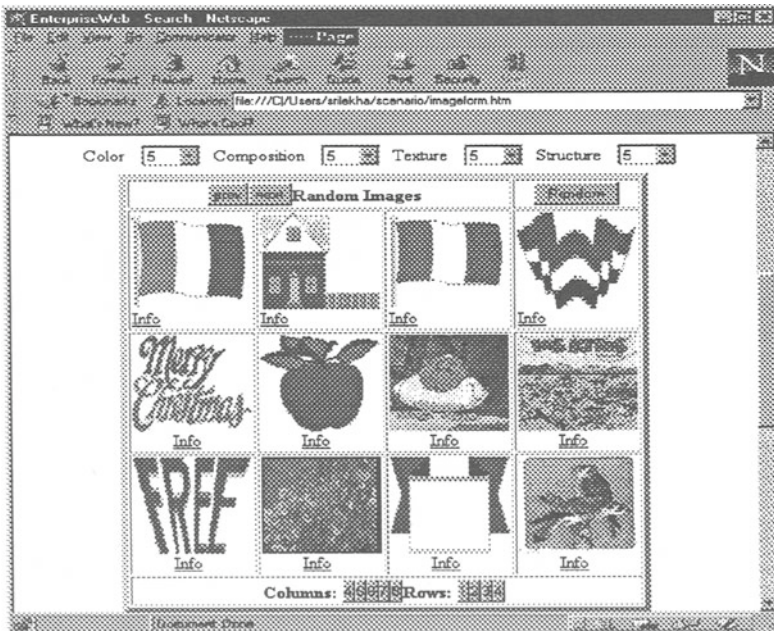


Figure A-7

Figure A-8 shows the results from choosing the 10th image in Figure A-7. Figure A-9 gives the results for the icon click on the first image of Figure A-9 in addition to attribute name, value pairs namely 'Imagename: *flowers*' and 'Color: *green*'. The results indicate how a given query can be refined using combination of different access strategies (image content based and attribute based access).

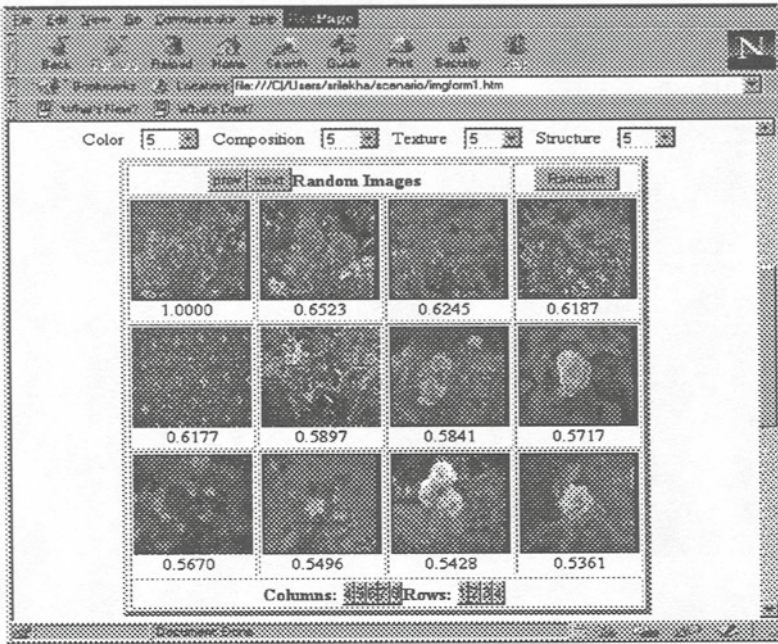


Figure A-8

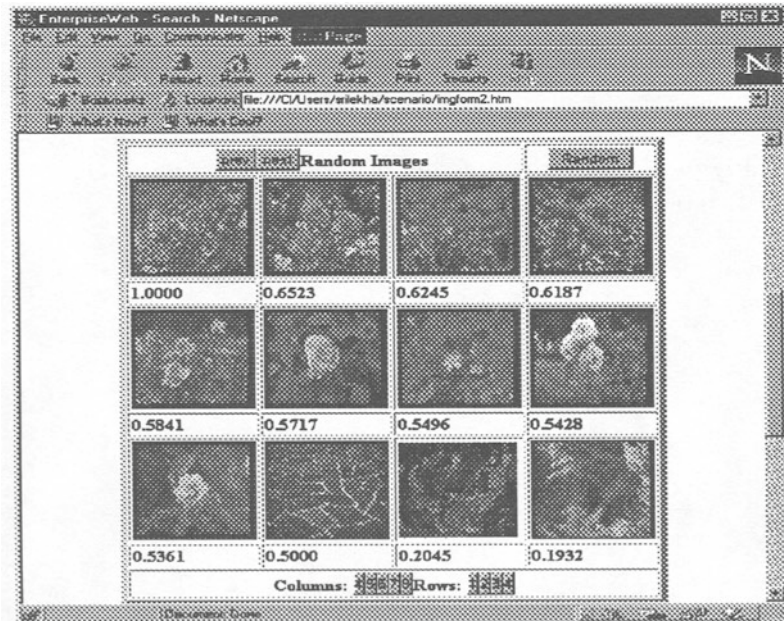


Figure A-9

Figure A-10 gives the results of only attribute based access on the attribute name, value pairs namely 'Imagename: *flowers*' and 'Color: *green*' as applied above.

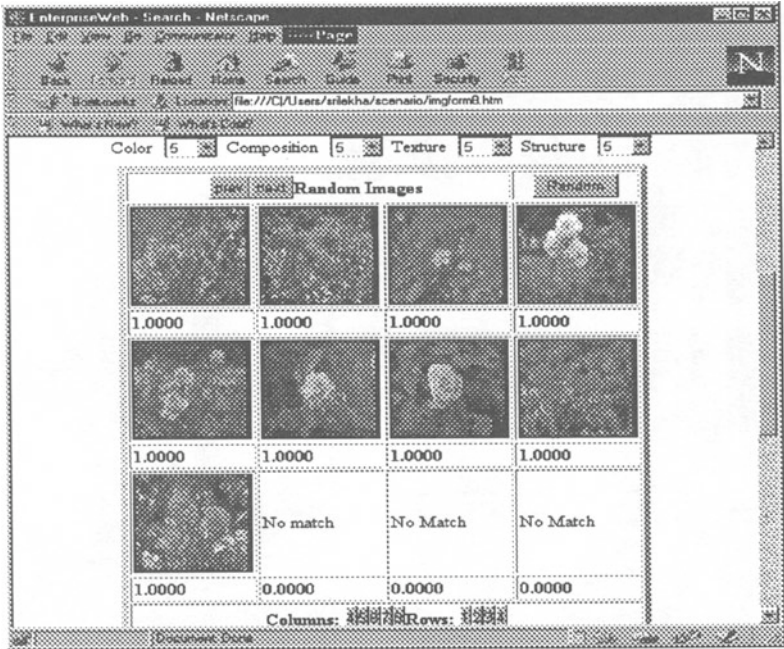


Figure A-10

A different ordering and ranking of images can be noticed at this point. This shows how a user can refine the queries and hence achieve better quality results. This is clearly indicated in the ordering of images in attribute access only and a combination of both attribute and image content based access.