Mining MultiMedia Data*

Osmar R. Zaïane Jiawei Han Ze-Nian Li Jean Hou
Intelligent Database Systems Research Laboratory, School of Computing Science
Simon Fraser University
Burnaby, BC, Canada V5A 1S6
E-mail: {zaiane, han, li, jhou}@cs.sfu.ca

Abstract

Data Mining is a young but flourishing field. Many algorithms and applications exist to mine different types of data and extract different types of knowledge. Mining multimedia data is, however, at an experimental stage.

We have implemented a prototype for mining high-level multimedia information and knowledge from large multimedia databases. MultiMediaMiner has been designed based on our years of experience in the research and development of a relational data mining system, DBMiner, in the Intelligent Database Systems Research Laboratory, and a Content-Based Image Retrieval system from Digital Libraries, C-BIRD, in the Vision and Media Laboratory.

MultiMediaMiner includes the construction of multimedia data cubes which facilitate multiple dimensional analysis of multimedia data, and the mining of multiple kinds of knowledge, including summarization, classification, and association, in image and video databases. The images and video clips used in our experiments are collected by crawling the WWW. Many challenges have yet to be overcome, such as the large number of dimensions, and the existence of multi-valued dimensions.

Keywords: Data Mining, Data Warehousing, Data Cube, Multimedia, Image Analysis, Information Retrieval, World-Wide Web.

1 Introduction

Substantial progress in the field of data mining and data warehousing research has been witnessed in the last few years. Numerous research and commercial systems for data mining and data warehousing have been developed for mining knowledge in relational databases and data warehouses [10]. Despite the fact that Multimedia has been the major focus for many researchers around the world, data mining from multimedia databases is still in its infancy. While one of the first dominant and referenced papers in the field of knowledge discovery by Fayyad et al.[8, 9] describes discovering patterns from satellite pictures, multimedia mining still seem shy on results. Many techniques for representing, storing, indexing, and retrieving multimedia data have been proposed. However, rare are the researchers who ventured in the multimedia data mining field. Most of the studies done are confined to the data filtering step of the KDD process as defined by Fayyad et al. in [29]. In [6], Czyzewski shows how KDD methods can be used to analyze audio data and remove noise from old recordings. Chien et al. in [5] use knowledge-based AI techniques to assist image processing in a large image database generated from the Galileo mission. Others use multimedia to complement data mining systems. Bhandari et al. [2], for instance, marries a data mining application with multimedia resources. His application does not claim to mine a multimedia database, but uses video

*Research is supported in part by the Natural Sciences and Engineering Research Council of Canada, the Canadian Network of Centres of Excellence (IRIS:HMI-5 and TL:NCE5.2), and MPR Teltech Ltd.

clips to support the knowledge discovered from a numerical database. More recently Tucakov and Ng in [33] used a method for outlier detection to identify suspicious behaviour from videos taken by surveillance cameras.

Multimedia data mining is a subfield of data mining that deals with the extraction of implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia databases. Multimedia data mining is not limited to images, video or sound, but encompasses text mining as well. There has been interesting research in text mining from text documents[11, 12] and Web or semi-structured data querying and mining[37, 20, 7, 26]. The availability of affordable imaging technology is leading to an explosion of data in the forms of image and video. Many relational databases are now including multimedia information, such as photos of customers, videos about real estate, etc. The proliferation of huge amounts of multimedia data is becoming prominent. Global information networks like the Internet are filled with a variety of multimedia, necessitating means to retrieve, classify and understand this data. Moreover, with the popularity of multimedia objects in extended and object-relational databases, it is becoming important to mine knowledge related to both multimedia and relational data in large databases, and maybe, to deal with them in the same manner.

Most of the recent work on multimedia systems has concentrated on transmission, synchronization and management of continuous data streams of audio, video and text. Other fields, no less important, are authoring, coding, indexing and retrieving of media data. The last focused area has drawn the attention of many. Researchers, for instance, try to "summarize" video clips in one image. Salient stills were introduced in [32], in an attempt to represent an abstract of a video clip in one still image. The salient stills reflect aggregates of temporal changes that occur in a moving image sequence. Stills are created automatically or with user intervention by combining affine transformation and multiple frames of the image sequence. Taniguchi et al.[31] use "mosaicing" to glue overlaping video frames to create a panoramic still image representing the video sequence. Despite the fact that representing a video clip in one still image summarizes in a way video clips, it is hard to claim that this is data mining from video.

With huge amount of multimedia data collected by video cameras and audio recorders, satellite telemetry systems, remote sensing systems, surveillance cameras, and other data collection tools, it is crucial to develop tools for discovery of interesting knowledge from large multimedia databases.

Recent advances in the research on multimedia databases [19, 4, 28, 13] enable creation of large multimedia databases which can be queried in an effective way. These advances, in combination with the research into multimedia database and advances in data mining in relational databases [10], created a possibility for the creation of multimedia data mining systems.

As an integrated team from two research labs, we have been working on multimedia data mining and especially spatial data mining for several years. Based on our previous research into relational data mining [15, 14, 16], spatial data mining [21, 25, 27], and content-based image and video retrieval from multimedia databases [22], we have extended the DBMiner system [14, 17, 15] and the C-BIRD system [24, 23] to manipulate and interpret multimedia data for knowledge discovery purposes.

The current MultiMediaMiner system, which was demonstrated at the SIGMOD98 conference, includes a module for *characterization* of knowledge in image and video databases, a module for *classification* of multimedia data, and a module for detection of *association* between multimedia features.

A more detailed description of the MultiMediaMiner system is presented in Section 2. The challenges and obstacles that we encountered with mining multimedia data, and the turn- arounds for our prototype implementation are presented in Section 3. Section 4 summarizes our on-going research.

(a) (b)

Figure 1: C-BIRD Web user interface.

2 A database mining system prototype

The MultiMediaMiner system is based on our experiences in the development of an on-line analytical data mining system, DBMiner, and C-BIRD, a system for Content-Based Image Retrieval from Digital libraries.

The DBMiner system, demonstrated in SIGMOD'96, KDD'96/97, CASCON'96/97, etc. (some of the function modules can be played on the Internet interactively via http://db.cs.sfu.ca/DBMiner), currently contains the following five data mining functional modules: *characterizer, comparator, associator, predictor*, and *classifier*. A general description of these functional modules is in [15]. Several additional functional modules, especially with time-related data, clustering, and visual data mining, are at the research and development stage. DBMiner applies multi-dimensional data-base structures [15], attribute-oriented induction, [14] multi-level association analysis, [16], statistical data analysis, and machine learning approaches for mining these different kinds of rules in relational databases and data warehouses. C-BIRD system, demonstrated in CASCON97 (some of the function modules can be played on the Internet interactively via http://jupiter.cs.sfu.ca/cbird/), contains four major components: (i) Image Excavator (a web agent) for the extraction of images and videos from multimedia repositories, (ii) a pre-processor for the extraction of image features and storing precomputed data in a database, (iii) a user interface, and (iv) a search kernel for matching queries with image and video features in the database. C-BIRD allows searches by conjunctions and disjunctions of keywords, colour histograms, colours with illuminance invariance, colour percentage, colour layout, edge density, edge orientation and texture coarseness. In particular, C-BIRD is characterized by its ability to cope with significant changes in image chrominance and to search by object model. The database used by C-BIRD is an addition to the image repository and contains mainly meta-data extracted by the pre-processor and the Image Excavator, like colour, texture, and shape characteristics and automatically generated keywords. MultiMediaMiner, the general architecture of which is shown in Figure 2, inherits the CBIRD database.

Figure 1(a) shows the C-BIRD Web user interface using Netscape to browse the image repository. Figure 1(b) shows a querying interface for a search by colour layout.

The Image Excavator and the pre-processor have been enhanced to collect and pre-process more information necessary for the MultiMediaMiner. Video clips are segmented after cuts have been detected. Each video segment is represented by one or more video frames which are later traited and processed by the system like images. For each image collected, the database contains some

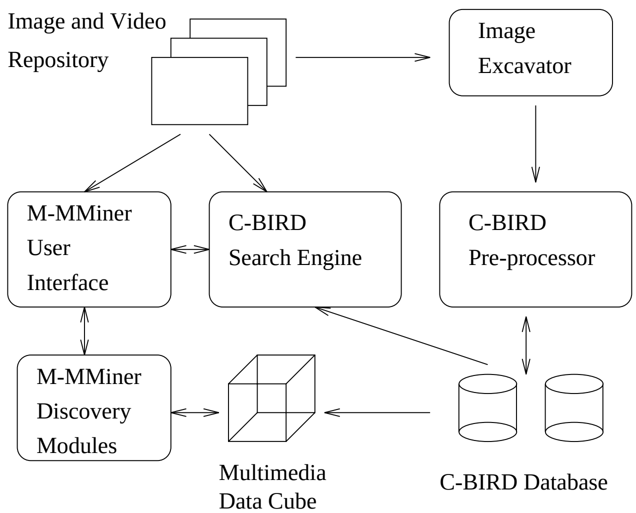

Figure 2: General Architecture of MultiMediaMiner.

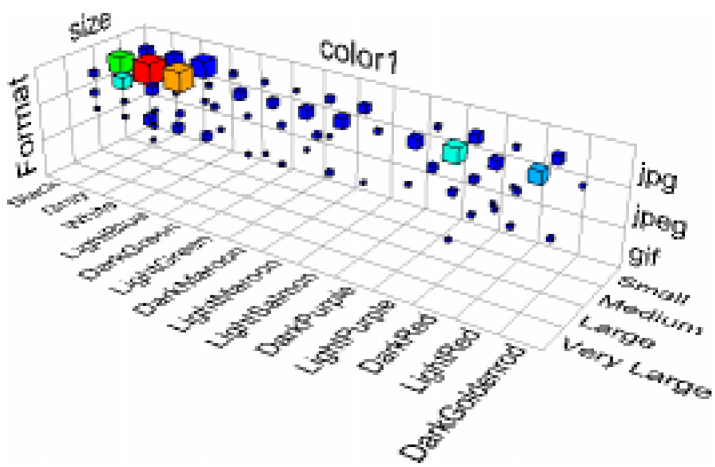

Figure 3: Browsing 3 dimensions of the multimedia data cube.

descriptive information, a feature descriptor, and a layout descriptor. The original image is not directly stored in the database; only its feature descriptors are stored. The descriptive information encompasses fields like: image file name, image URL, image and video type (i.e. gif, jpeg, bmp, avi, mpeg, ...), a list of all known web pages referring to the image (i.e. parent URLs), a list of keywords, and a thumbnail used by the user interface for image and video browsing. The feature descriptor is a set of vectors for each visual characteristic. The main vectors are: a colour vector containing the colour histogram quantized to 256 colours (all colours are represented in the RGB space by 8 values in red, 8 values in green and 4 values in blue), MFC (Most Frequent Colour) vector, and MFO (Most Frequent Orientation) vector. The MFC and MFO contain 5 colour centroids and 5 edge orientation centroids for the 5 most frequent colours and 5 most frequent orientations (the edge orientations used are: $0°$, $45°$, $90°$, $135°$). The layout descriptor contains a colour layout vector and an edge layout vector. These vectors allow matching with user-defined layouts as in the user interface shown at the right of Figure 1(a). Regardless of their original size, all images are assigned an 8×8 grid. The most frequent colours for each of the 64 cells are stored in the colour layout vector and the number of edges for each orientation in each of the cells is stored in the edge layout vector. Other sizes of grids, like 4×4, 2×2 and 1×1, can be derived easily. These colour layout grids can be used for spatial relationships between colours at different levels of resolution.

The Image Excavator uses image contextual information, like HTML tags in web pages, to derive keywords. For example, image file name and path if it contains a word or recognizable words, ALT field in the IMG tag, HTML page title, HTML page headers, parent HTML page title, hyperlink to the image from parent HTML page, and neighbouring text before and after the image, META tag placed in the HEAD element of the HTML page, can disclose valuable keywords related to an image. The set of all words collected this way, is reduced by eliminating "empty" words like articles (i.e. the, a, this, etc.) or common verbs (i.e. is, do, have, was, etc.), or aggregating words from the same canonical form (e.g., clearing, cleared, clears, clear) as presented in [36]. There are 400 frequently found words in English, defined in [34], that can be considered of low semantic information content and thus, can be eliminated (stopwords). The automatically generated keyword list is later normalized and filtered and used to build a concept hierarchy as explained in Section 3. The hierarchy of keywords with its hypernymy and hyponymy relationships allows one to browse the image and video collection by topic. In Figure 4, for example, thumbnails of commercial airplanes pertaining to the aircraft manufacturer Boeing are displayed. This user interface also allows the selection of a multimedia data set to be mined. The hierarchy of keywords on the left of Figure 4 is a section of the concept hierarchy automatically generated by visiting some web sites containing aircraft images.

The mining modules of the MultiMediaMiner system include three major functional modules, *characterizer, classifier*, and *associator*. Many data mining techniques are used in the development of these modules, including *data cube construction and search* [3], *attribute-oriented induction* [15], *mining multi-level association rules* [16], etc.

The functionalities of these modules are described as follows:

- MM-Characterizer: This module discovers a set of characteristic features at multiple abstraction levels from a relevant set of data in a multimedia database. It provides users with a multiple-level view of the data in the database with roll-up and drill-down capabilities. Figure 5 describes in a histogram graph the general characteristics for two dimensions: the size of the media in bytes and the Internet domain from which the media were extracted. For this example, only three Internet domains were considered, while the sizes were "rolled-up" to a higher concept of media size (i.e. small, medium and large). With this user interface, it is possible to visualize any two dimensions at a time, and drill-down or roll-up along a given dimension to find characteristics on more concrete values or specialized concepts.

- MM-Associator: This module finds a set of association rules from the relevant set(s) of data in an image and video database. An association rule shows the frequently occurring patterns

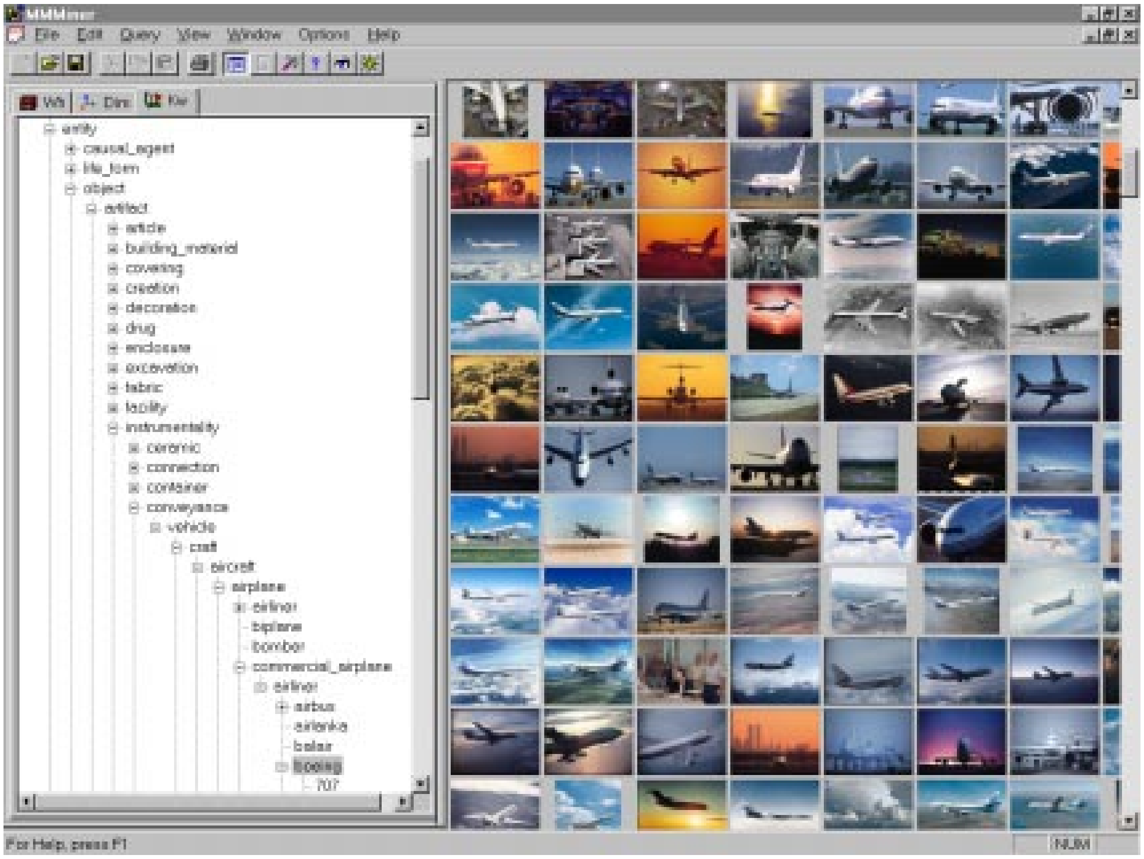

Figure 4: Selecting (and browsing) data sets of images using keyword hierarchy.

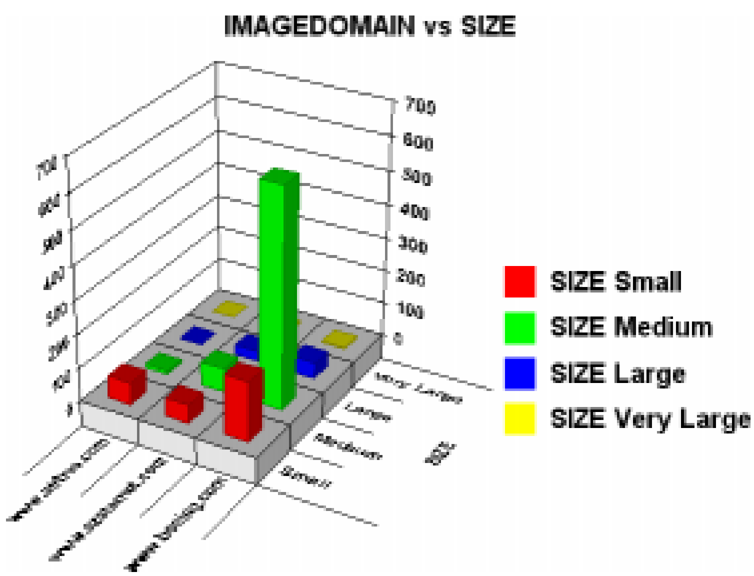

Figure 5: Snapshot of MultiMediaMiner Characterizer

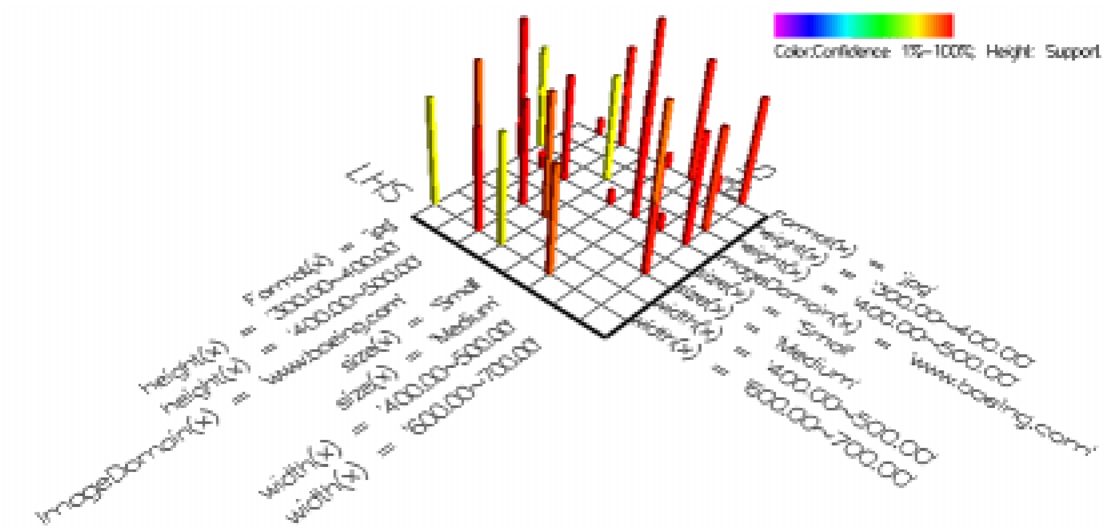

Figure 6: Visualization of association rules.

(or relationships) of a set of data items in a database. A typical association rule is in the form of "$X \rightarrow Y[s\%, c\%]$" where X and Y are sets of predicates, $s\%$ is the support of the rule (the probability that X and Y hold together among all the possible cases), and c% is the confidence of the rule (the conditional probability that Y is true under the condition of X). For example, the module mines association rules like: "*what are relationships among still images, the frequent colours used in them, their size and the keyword '*sky*'?*" One possible association rule among many to be found is "*if image is big and is related to* sky*, it is* blue *with a possibility of 68%*" or "*if image is small and is related to* sky*, it is* dark blue *with a possibility of 55%*". Figure 6 shows a visualization of some association rules. The existence of a column on the grid represents an association between the left-hand side parameters and the right-hand side parameters. The height of the column depicts the support of the rule it represents while the colour of the column describes the confidence of the rule.

- MM-Classifier: This module classifies multimedia data based on some provided class labels. The result is an elegant classification of a large set of multimedia data and a characteristic description of each class. This classification represented as a decision tree can also be used for prediction. Figure 7 shows an output of this module where a classification of images and frames based on their topic, with reference to the distribution of image format, is made for a given Web site. By clicking on a class, it is possible to drill through to the raw data. A window displays the images pertaining to the class (ex. book, animal, flower in Figure 7).

The user interface of all these modules allow drilling and rolling-up along the different concept hierarchies defined on the dimensions, and thus, allow interactive mining. It is also possible to drill through right to the raw data. In our case the raw data are images and videos stored on the Web. MultiMediaMiner calls a Web browser and displays the original image in its original size or even the web pages that contain the image. This gives an opportunity for information retrieval from the Web, based on the data mining results.

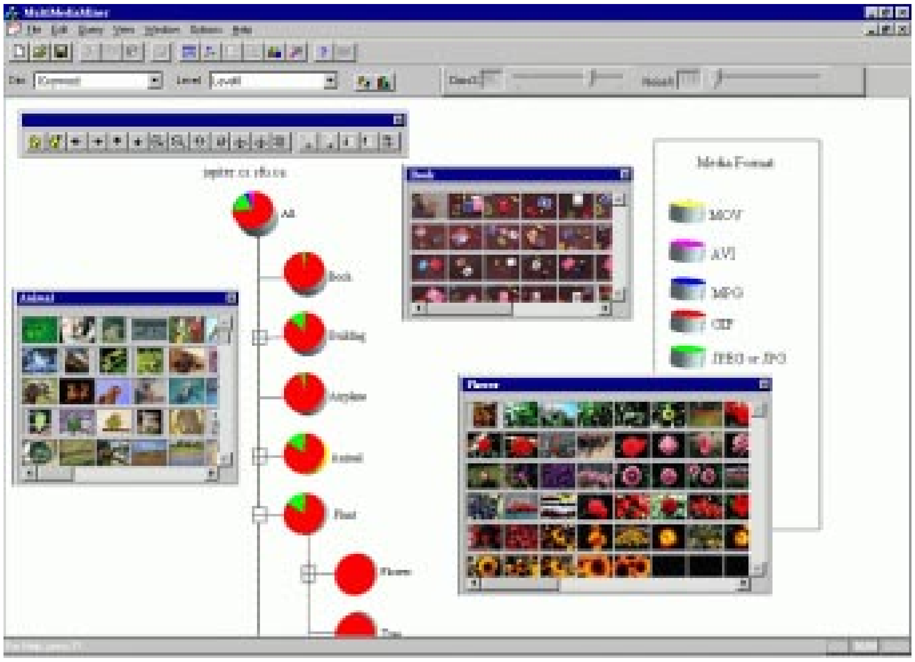

Figure 7: MultiMediaMiner Classifier user interface.

3 Obstacles and Challenges with Multimedia Mining

The first problem with mining multimedia databases is gaining access to significantly large multimedia data sets. This may seem trivial, but getting access to CT scans from hospitals, for instance, is not easy due to privacy issues. CT scans would have been an interesting application for the discovery of association rules based on colours in these scans. We chose the World-Wide Web as our image and video source because it is free, available and has a reasonably large collection. Another advantage of using the Web as our source for images and video is that we can use the context of the images to automatically extract additional information like the keywords from the pages containing the image, the popularity of the image (i.e. how many pages use the same image), the Internet domain of the image, etc. All this information was added to the already dimension-rich database. Moreover, by saving the URL of images, we avoid the need for large storage space for the images and videos. The World-Wide Web is used as the repository. This, however, requires regular validation due to the World-Wide Web dynamic nature. Indeed, some images may disappear and some new ones appear in the web pages already visited by our crawler. If images disappear from the Web, they are discarded form our database. If the images change, they are processed again and the descriptors in our database replaced while the changes are propagated to the data cube structure. In addition, by saving the URLs of the images and the URLs of the pages that contain the images, it is possible to do information retrieval and resource discovery from the Web by drilling through the results of the data mining process.

3.1 Keyword hierarchies

Keywords describing images are very important and useful when dealing with large collections of images. However, automatically associating keywords to images is not easy, while manual keywording is definitely not scalable. As mentioned in Section 2, we take advantage of the semi-structure of the web pages and the syntax of the URLs to extract candidate keywords that after normalization and

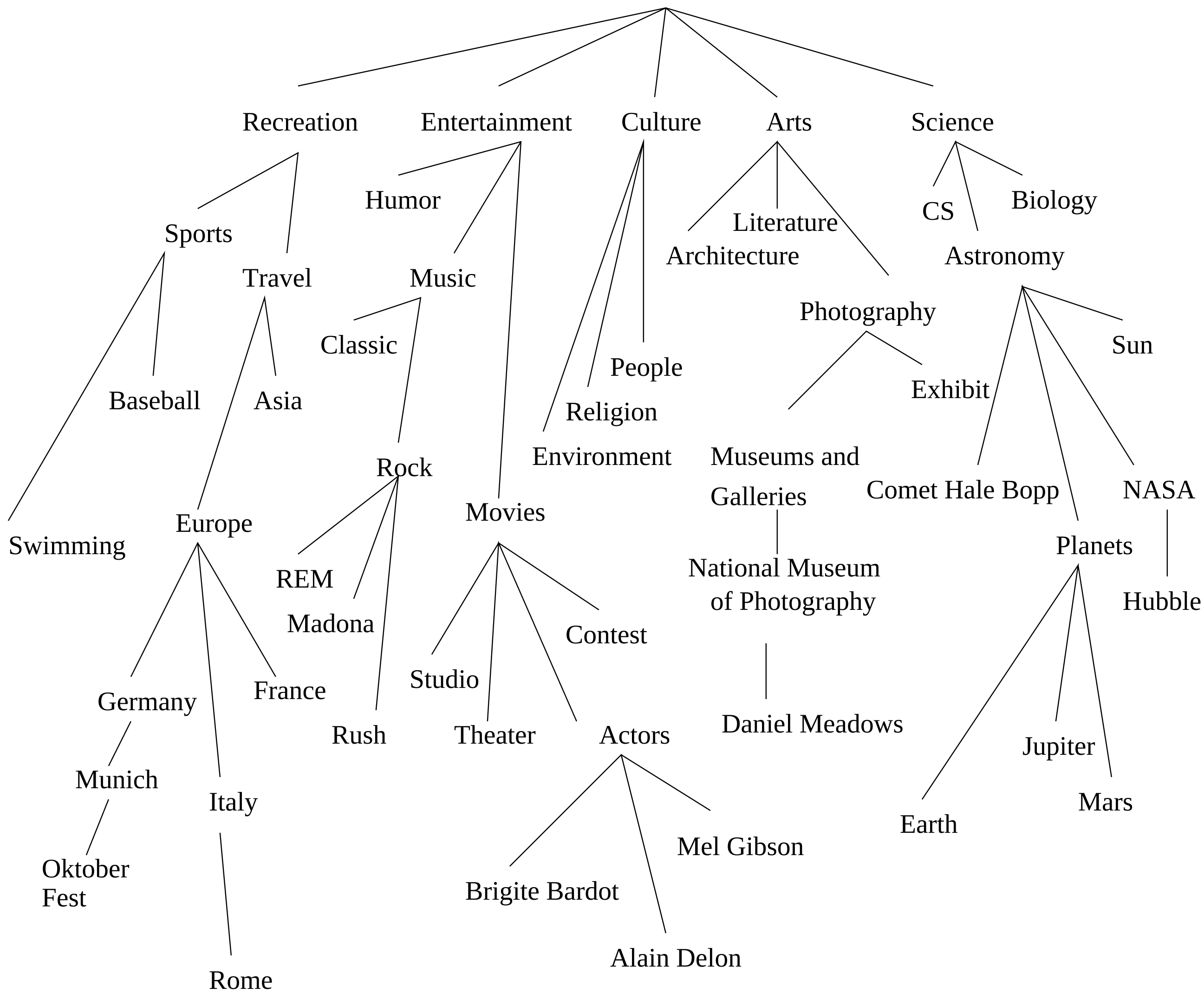

Figure 8: Portion of the keyword hierarchy generated by traversing the Yahoo directories.

filtering are associated to the images. The normalization process uses morphological analysis to draw forth the canonical forms of words, while the filtering process uses a list of stopwords and WordNet lexical database to eliminate illicit or unwanted words. While the candidate keyword selection and the keyword filtering eliminate most of the unwanted words, the list of keywords per image still remains large. This can be reduced by adding new stopwords and/or use natural language heuristics to eliminate oultiers.

For On-Line Analytical Processing (OLAP), concept hierarchies are needed to drill-down and roll-up along the dimensions defined on the data. These hierarchies are also important for multi-level mining in order to specialize or generalize the knowledge discovered. Thus, organizing the keywords in a concept hierarchy is pertinent for multimedia mining. However, building a concept hierarchy of natural language words is difficult because of the controversies it may generate. We had to build an explicit representation of the set of keywords in the form of concept hierarchy that most people (users) would agree upon. The solution was to use existing word hierarchies that are widely and extensively used and accepted. Our first attempt was to automatically build a concept hierarchy

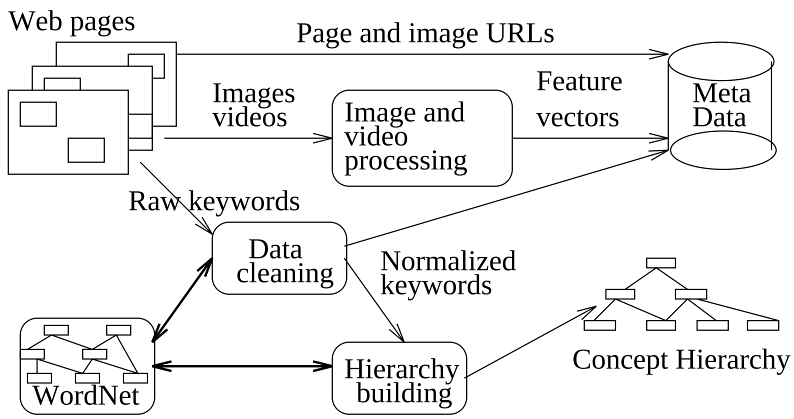

Figure 9: Keyword Normalization and Concept Hierarchy building using WordNet.

by traversing a manually-built and widely-used on-line directory structures. By traversing the on-line Yahoo directory, for instance, one can build a general hierarchy with all nodes of the directories. Figure 8 shows a portion of the keyword hierarchy generated by traversing the Yahoo directories and mapping the directories to keywords. Unfortunately, this hierarchy tends to be too general and is not flexible enough to accommodate new terms. In other words, the hierarchy generated is shallow, narrow and not flexible.

Ultimately, we opted to use the on-line dictionary, thesaurus, and semantic network WordNet developed at the University of Princeton [35, 1] and used by many researchers in linguistics and cognitive science. WordNet version 1.6 contains approximately 95 600 different word forms organized into 71 100 word meanings interconnected with links representing subsumptions. Unfortunately, WordNet's word list does not contain specific words like "Boeing 747" or "fighter F15" that were extracted from the Web sites our crawler visited. After consulting the list of words rejected by the filtering process, some words were selected and added to enrich WordNet's semantic network with these new domain related terms.

Finally, the subsumption connections in the enhanced WordNet semantic network were used to build a concept hierarchy with all (and only) the keywords extracted and accepted from the web pages. This hierarchy is used to classify images by topic and browse the image and video collection. Figure 4 shows a portion of such hierarchy starting from the node "entity" of the enhanced WordNet network.

Figure 9 illustrates the use of WordNet for keyword filtering and word hierarchy building.

3.2 The curse of dimensionality

A data cube is a particular structure for storing multi-dimensional data and handling queries that aggregate over some of these dimensions at different levels of abstraction. This structure can be stored either in main memory or on disk.

The multimedia data cube we use has many dimensions. The following are some examples: (1) The size of the image or video in bytes with automatically generated numerical hierarchy. (2) The width and height of the frames (or picture) constitute 2 dimensions with automatically generated numerical hierarchy. (3) The date on which the image or video was created (or last modified) is another dimension on which a time hierarchy is built. (4) The format type of the image or video with two-level hierarchy containing all video and still image formats. (5) The frame sequence duration in seconds (0 seconds for still images) with numerical hierarchy. (6) The image or video Internet domain with a pre-defined domain hierarchy; Each image or video collected has a unique URL

(Unified Resource Locator) that indicates the location (Internet domain) where the image or video is stored. (7) The Internet domain of pages referencing the image or video (parent URL) with a pre-defined domain hierarchy; When an image or video is located in a web page, a reference to that page (parent URL) is stored with the image meta-data in our database. (8) The keywords with a term hierarchy defined as described above; (9) A colour dimension with a pre-defined colour hierarchy; colours are quantized and indexed in a range between 0 and 255. A colour hierarchy is defined from specific colours to more general colours. An image or a video is considered containing a given colour if the percentage of pixels in that colour exceeds a given threshold. (10) An edge-orientation dimension with a pre-defined hierarchy, etc. An image is considered containing a certain edge orientation if the percentage of edges in the orientation in the image exceeds a given threshold. (11) The popularity of an image or video with a numerical hierarchy. The popularity of an object is the known number of pages that reference that object. (12) The richness of a web page with a numerical hierarchy. The richness of a web page is the number of multimedia objects referenced in the page.

Using these different dimensions and their respective concept hierarchies, it is possible to build a multi-dimensional data cube that aggregates the values for all attributes in each dimension domain. Figure 3 shows a visualization tool used to browse such multi-dimensional data cubes, 3 dimensions at a time. The concept hierarchy defined on each dimension allows drilling-down and rolling-up along any given dimension. This type of data cube browsing gives a big picture of the content of the database and even allows to see rough clustering of data values. Selecting a sub-cube from the view drills through it up to the raw data, and one can see the set of multimedia items in the selected sub-cube and even the web pages that contain them.

Unfortunately, it is very difficult, if not impossible, to have more than a given number of dimensions in a physical data cube. This is not due to the visualization or conceptualization as it may seem, but it is due to the fact that the size of the data cube grows exponentially with the number of dimensions. Each time a dimension is added, the size of the data cube is multiplied by the number of distinct values in the new dimension. This is the curse of dimensionality. In [30] Ross illustrates how the number of dimensions in a data cube is physically limited due to the physical size of the memory.

The colour attribute of an image has 256 dimensions, for instance. Each of the dimensions counts the frequency of a given colour in images. This already goes beyond the limit of most data cube-based systems. Even after quantizing the colours to 64 values, the number of dimensions is still too large for MultiMediaMiner to handle. In order to reduce the number of dimensions, we decided to collapse and pivot the 64 colour dimensions into one. One previous colour dimension represented a colour and the values were frequencies of that colour in an image. With the collapsed dimension, the values represented are colours and the colour frequencies are discarded. This loss of information is a compromise to reduce to dimensionality. The same principal was applied for the dimensions of the attribute texture. This brings up yet another challenge: the problem of multi-valued attributes. The collapsed colour dimension represents all the colours, however, an image or a video frame has more than one colour. If all the colours of an image are represented in the same dimension, the aggregate values in the aggregated layers of the data cube become wrong and meaningless. To solve this problem, a colour dimension for each colour present in an image is needed. However, this contradicts the goal of reducing the dimensionality. In our implementation, we have chosen to represent only the three most frequent colours of an image with 3 colour dimensions. This reduces the colour representation from 256 dimensions to 3.

As might be expected, colour is not the only multi-valued dimension. An image has many textures, is described by many keywords, and can be present in many web pages. In other words, the dimension texture, the dimension keyword, and the dimensions related to the web page (page richness and parent page Internet domain) are all multi-valued. For our prototype implementation, we had to compromise by choosing to represent only the most frequent texture in an image, only the first parent web page of an image found by our crawler, and we chose not to represent the keywords

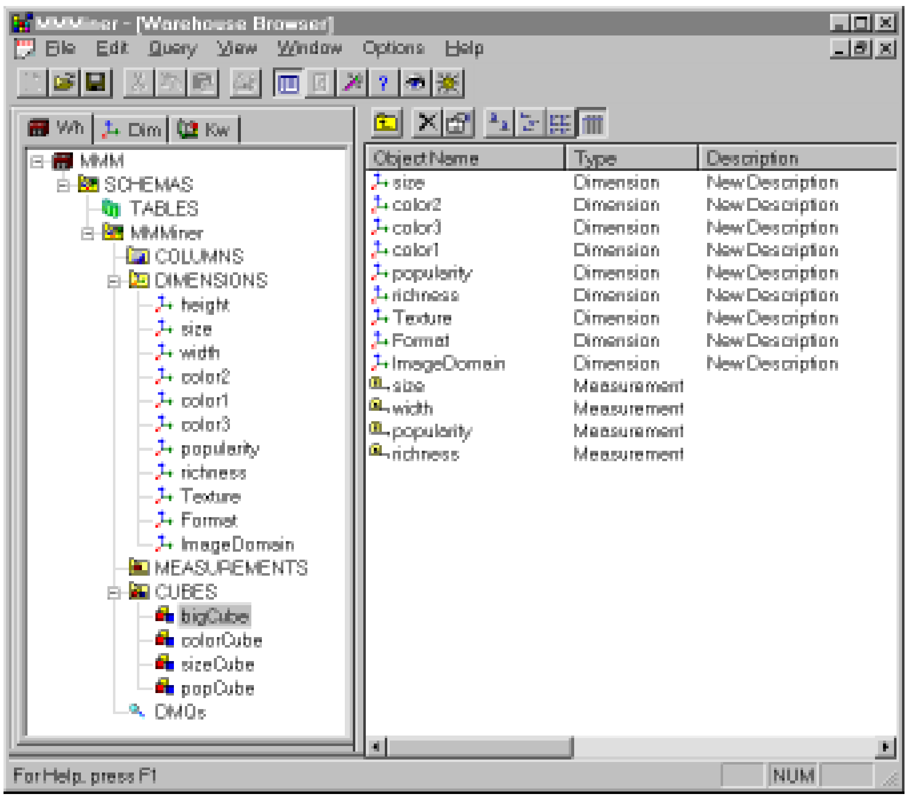

Figure 10: MultiMediaMiner data warehouse with cubes and dimensions.

in our data cube. Not only it is not significant to select only one keyword by image or video since the keywords can not be ranked effectively, but the keyword dimension has also a very large number of potential values formed from words and phrases. This would cause the size of the cube to rapidly exceed the physical available limit.

Despite the fact that keywords are not represented in our cube, we use the keywords as a data set selection attribute to select a set of images on which to build our data cube. Thus, the aggregate values in the data cube pertain to the multimedia objects that are associated with the keyword used for the selection. By doing so, the selection keyword can be appended as a predicate to any rule discovered by our data mining modules based on the constructed data cube. Figure 4 shows the selection process using the keyword concept hierarchy. This selection is used for browsing images and for data set selection for data cube construction. When a keyword is selected, all keywords subsumed by it are also selected. This allow generalization and specialization along the word hierarchy.

Although we reduced the number of dimensions, the number is still large. For the implementation of the MultiMediaMiner prototype, we have chosen to create not one cube, but a set of different data cubes with different (overlapping) dimensions. Figure 10 shows the user interface of the MultiMediaMiner data warehouse with 4 data cubes and the dimensions and measurements defined in one of them. Separating the data cube into smaller ones is a limitation. This restriction brought up new challenges. It is not trivial to choose which dimension should be represented in which cube, when we have our data materialized in separate cubes. It is important to mention that the OLAP interaction and the data mining algorithms operate on one given cube at a time. Thus, it is not possible to discover correlations, for example, between two dimensions in different data cubes. Moreover, merging rules discovered form two cubes that do not overlap, is not possible.

In [18] selective materialization of data cubes is proposed to select the appropriate cuboids for materialization rather than materializing all the views. This approach, using a lattice that expresses dependencies among views and contains cube materialization costs, is intended to optimize the data cube construction based on the needs dictated by the user queries. In our implementation, as mentioned above, we chose to materialize 4 cuboids and pre-compute them after the user selects a data set using the keyword hierarchy. The cubes are built on-the-fly and can easily be built in parallel. There are some heuristics regarding the selection of the dimensions in the different cuboids, some based on the access frequency and some based on the size of the dimensions themselves. We opted for a more semantic approach. The set of dimensions was divided into 3 sub-sets: a content-based dimension set (colour and texture), a size-based dimension set (size, width, height, etc.), and a resource-based dimension set (Internet domain, popularity, etc.). Each set was materialized in a different cuboid. In addition, a fourth cuboid was materialized with dimensions from the 3 dimension sets. In order to create an overlap between the cuboids, the Internet domain and the size dimensions were repeated in all 4 cuboids.

Each cell of a data cube can contain aggregate values (i.e. measurements) like a count, a sum, etc. Because measurements are not expensive in memory size, we decided to materialize numeric attributes (like size, richness, popularity, etc.) as measurements, rather than as cube dimensions, whenever the attribute is not selected as dimension effectively present in the cuboid. This allows the consideration of values of that attribute, however, without the possibility to drill-down or roll-up along the dimension it represents.

A new model for data cube materialization is under study. In this model, called MDDB for Multi-Dimensional DataBase, we conceptualize the entire data cube in a database with a special-purpose structure. The structure contains all dimensions and the aggregation of interesting values in preparation for cube materialization. The structure is not a data cube *per se*, but the "definition" of the hypercube which helps speed up the materialization of cuboids. Cuboids are then materialized on-the-fly depending on the dimensions needed by the query. Moreover, borrowing from the multi-layered database technology presented in [37], a cuboid can generalize a set of cuboids along the hierarchies of its dimensions. A cuboid would join the dimensions of other cuboids at a higher conceptual level. This model allows the creation and manipulation of data cube with an unrestricted

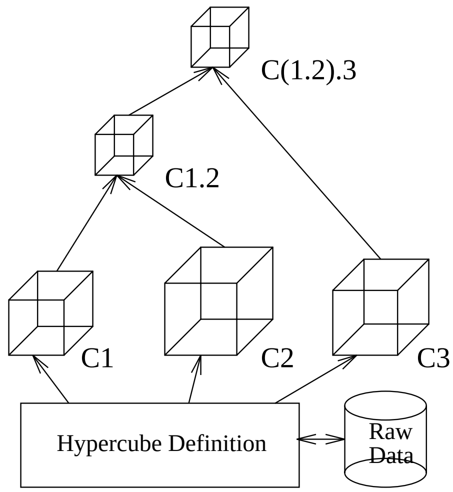

Figure 11: Muli-Dimensional Database model with materialization of cuboids.

number of dimensions, and allows multi-dimensional selection on raw data. Figure 11 shows the cuboid materialization path from a hypercube definition.

Multi-dimensional data cubes are created in order to reduce the response time when querying large databases for decision support or data mining. Typically, all the dimensions are aggregated in the cube. However, it is not always necessary to represent all the dimensions in the cube. Depending upon the application and the user needs, we can choose not to materialize some dimensions and keep them as raw data in the database. For example, if colour is considered unnecessary for some applications, we can avoid materializing the colour dimensions and keep the colour information in the database. This obviously reduces the size of the data cube. However, if for some queries colour is required, we need to build on-the-fly a new data cube with colour dimensions directly from the raw data. This can be very costly. Another approach would be to adapt the data mining algorithms to use simultaneously the aggregations in the data cube and the raw data in the database without materializing the portion of the data that is still in the database. This is acceptable if the queries accessing the non materialized portions are scarce.

4 On-going work and Conclusions

We have designed and developed an interesting multimedia data mining system prototype, MultiMediaMiner, with the following features: (i) a multidimensional multimedia data cube, (ii) multiple data mining modules, including characterization (or summarization), association, and classification, and (iii) an interactive mining interface and display with Web information retrieval capabilities. Our preliminary experiments demonstrate that multimedia data mining may lead to interesting and fruitful knowledge discoveries in multimedia databases.

There are some major tasks calling for further research into the design and development of the MultiMediaMiner system.

The design and construction of multimedia data cube can be improved by integrating the MDDB model or by using a virtual composite data cube that has some of its dimensions not materialized but in the database. The current design of the multimedia data cube, though works, produces a huge multimedia data cube, due to the big size of two numeric dimensions: *colour* and *texture*. Most relational data cubes contain only categorical dimensions each having a relatively small number of distinct values. However, since we would like to support search from colour and edge-orientation, it is necessary for the data mining algorithms to have access to the data either materialized in a cube or directly from the database. Our current implementation supports only a limited number of intervals on these two dimensions in the data cube. The search along these dimensions with finer granularities than those currently supported has to access the C-BIRD database, which degrades the performance but can be improved by using the hypercube structure of the MDDB model.

Another task is to enhance our data mining algorithms to take advantage of the MFC and MFO centroids pre-processed and stored in the C-BIRD database. The centroids can help in order to discover interesting spatial relationships within an image or between frames of a video clip. We are defining spatial primitives like *next_to, ontop_of* and *under* to describe relationships between colours or colour segments in an image. These primitives and colour layout grids extracted by the preprocessor can help discover association rules about colours within an image or a video clip. In [23] we define the notion of localization or locales which are rough colour and texture segments in an image. We are studying the option to use these locales, rather than all the colour and texture of an image, to describe the colour features of the image or objects within the image, since they are perceptually more accurate.

There are plans to add new data mining functionalities into the system, like a clustering module which would group images into different clusters based on their multiple dimensional features, including both multimedia features, such as colour and edge-orientation, and relational features, such as keywords, URL information, and duration.

We have used the keyword hierarchy for browsing our image collection and selecting a data set for mining. In other words, the selection of images to mine is done based on keywords. We plan to use the content-based image retrieval features of C-BIRD to also select the images for mining.

With the introduction of multimedia data, there are some research issues to be solved before successful construction of these two new data mining modules. Further developments of multimedia data cubes and multimedia data mining modules in the MultiMediaMiner system will be reported in the future.

Acknowledgements

The authors would like to express their thanks to Gang Liu and Eli Hagen for their invaluable help in the implemetation of the system.

About the Authors

Osmar R. Zaïane is a Ph.D. candidate in Computing Science at Simon Fraser University, Canada. He holds a Master's in Computer Science from Laval University, Canada, and a Master's in Electronics from the University of Paris XI, France. He has worked in a variety of research areas, such as data mining, web mining, multimedia databases, information retrieval, web technology, natural language processing, distance education and collaborative learning, and smartcard technology. Osmar will integrate the Computing Science departement at the University of Alberta, Canada, as an Assistant Professor in July 1999.

Jiawei Han (Ph.D. Univ. of Wisconsin at Madison, 1985), Professor of the School of Computing Science and Director of Intelligent Database Systems Research Laboratory, Simon Fraser University, Canada. He has conducted research in the areas of data mining and data warehousing, spatial databases, multimedia databases, deductive and object-oriented databases, and Web technology, with over 100 journal and conference publications. He is the project leader of an NCE/IRIS-3 project "Building, Querying, Analyzing, and Mining Data Warehouses on the Internet" (1998-2002), and has served or is currently serving in the program committees of over 30 international conferences and workshops, including ACM-SIGMOD'99/96, ICDE'97-99, VLDB'96, KDD'96-99 (KDD'96 PC co-chairman), SSD'97/99. He has also been serving as an editor for IEEE Transactions on Knowledge and Data Engineering, Journal of Intelligent Information Systems, and Data Mining and Knowledge Discovery.

Ze-Nian Li (Ph.D. Univ. of Wisconsin at Madison, 1986), Professor of the School of Computing Science and Director of the Multimedia and Vision Laboratory, Simon Fraser University, Canada. His current research interests include computer vision, pattern recognition, parallel vision machines and algorithms, and content-based retrieval in multimedia systems. He has over 70 publications in international journals and conferences. Recently, he has lead the developement the C-BIRD system (Content-Based Image Retrieval in Digital-libraries) as part of the "Multimedia Databases" project in the Network of Centres of Excellence (NCE-Telelearning).

Jean Hou is a second year M.Sc student at Simon Fraser University Computing School. Currently her research topic is Clustering. She has completed her Hon. Co-op B.Math degree at the University of Waterloo with major in Computer Science and minor in Statistics (1991 - 1996).

References

[1] R. Beckwith, C. Fellbaum, D. Gross, K. Miller, G.A. Miller, and R. Tengi. Five papers on WordNet. *Special Issue of Journal of Lexicography*, 3(4):235–312, 1990. Also available from ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

[2] I. Bhandari, E. Colet, J. Parker, Z. Pines, and R. Pratap. Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, 1(1):121–125, 1997.

[3] S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26:65–74, 1997.

[4] S. Chaudhuri, S. Ghandeharizadeh, and C. Shahabi. Avoiding retrieval content for composite multimedia objects. In *Proc. 21st Int. Conf. on Very Large Data Bases*, pages 287–298, 1995.

[5] S. Chien, F. Fisher, H. Mortensen, E. Lo, and R. Greeley. Using artificial intelligence planning to automate science data analysis for large image databases. In *Proc. Third Int. Conf. on Knowledge Discovery and Data Mining*, pages 147–150, 1997.

[6] A. Czyzewski. Mining knowledge in noisy audio data. In *Proc. Second Int. Conf. on Knowledge Discovery and Data Mining*, pages 220–225, 1996.

[7] O. Etzioni. The world-wide web: Quagmire or gold mine? *Communications of ACM*, 39:65–68, 1996.

[8] U. Fayyad and P. Smyth. Image database exploration: Progress and challenges. In *Proc. Knowledge Discovery in Databases Workshop*, pages 14–27, Washington, D.C, 1993.

[9] U. M. Fayyad, S. G. Djorgovski, and N. Weir. Automating the analysis and cataloging of sky surveys. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 471–493. AAAI/MIT Press, 1996.

[10] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.

[11] R. Feldman and I. Dagan. Knowledge discovery in textual databases (KDT). In *Proc. 1st Int. Conf. Knowledge Discovery and Data Mining*, pages 112–117, Montreal, Canada, Aug. 1995.

[12] R. Feldman and H. Hirsh. Mining associations in text in the presence of background knowledge. In *Proc. 2st Int. Conf. Knowledge Discovery and Data Mining*, pages 343–346, Portland, Oregon, Aug. 1996.

[13] M. Flickner, H. Sawhney, W. Niblack, and et al. Query by image and video content: The QBIC system. *IEEE Computer*, pages 23–32, September 1 1995.

[14] J. Han, Y. Cai, and N. Cercone. Data-driven discovery of quantitative rules in relational databases. *IEEE Trans. Knowledge and Data Engineering*, 5:29–40, 1993.

[15] J. Han, J. Chiang, S. Chee, J. Chen, Q. Chen, S. Cheng, W. Gong, M. Kamber, G. Liu, K. Koperski, Y. Lu, N. Stefanovic, L. Winstone, B. Xia, O. R. Zaïane, S. Zhang, and H. Zhu. DBMiner: A system for data mining in relational databases and data warehouses. In *Proc. CASCON'97: Meeting of Minds*, pages 249–260, Toronto, Canada, November 1997.

[16] J. Han and Y. Fu. Discovery of multiple-level association rules from large databases. In *Proc. 1995 Int. Conf. Very Large Data Bases*, pages 420–431, Zurich, Switzerland, Sept. 1995.

[17] J. Han and Y. Fu. Exploration of the power of attribute-oriented induction in data mining. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 399–421. AAAI/MIT Press, 1996.

[18] V. Harinarayan, A. Rajaraman, and J. D. Ullman. Implementing data cubes efficiently. In *Proc. 1996 ACM-SIGMOD Int. Conf. Management of Data*, pages 205–216, Montreal, Canada, June 1996.

[19] S. Khoshafian and A. B. Baker. *Multimedia and Imaging Databases*. Morgan Kaufmann Publishers, 1996.

[20] D. Konopnicki and O. Shmueli. W3QS: Aquery system for the world wide web. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 54–65, 1995.

[21] K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In *Proc. 4th Int'l Symp. Large Spatial Databases (SSD'95)*, pages 47–66, Portland, Maine, Aug. 1995.

[22] Z.N. Li and B. Yan. Recognition kernel for content-based search. In *Proc. IEEE Conf. on Systems, Man, and Cybernetics*, pages 472–477, 1996.

[23] Z.N. Li, O. R. Zaïane, and Zinovi Tauber. Illumination invariance and object model in content-based image and video retrieval. *Journal of Visual Communication and Image Representation*, 1998. Submitted for review.

[24] Z.N. Li, O. R. Zaïane, and B. Yan. C-bird: Content-based image retrieval in digital libraries using chromaticity and recognition kernel. In *International Workshop on Storage and retrieval Issues in Image and Multimedia Databases, in conjonction with the 9th International Conference on Database and Expert Systems (DEXA'98)*, Vienna, Austria, August 1998.

[25] W. Lu, J. Han, and B. C. Ooi. Knowledge discovery in large spatial databases. In *Proc. Far East Workshop Geographic Information Systems*, pages 275–289, Singapore, June 1993.

[26] A. Mendelzon and T. Milo. Formal models of web queries. In *Proc. 15th ACM Symp. Principles of Database Systems*, pages 134–143, Tucson, Arizona, May 1997.

[27] R. Ng and J. Han. Efficient and effective clustering method for spatial data mining. In *Proc. 1994 Int. Conf. Very Large Data Bases*, pages 144–155, Santiago, Chile, September 1994.

[28] B. Özden, A. Biliris, R. Rastogi, and A. Silberschatz. A low-cost storage server for movie on demand databases. In *Proc. 20th Int. Conf. on Very Large Data Bases*, pages 594–605, 1994.

[29] G. Piatetsky-Shapiro, U. Fayyad, and P. Smith. From data mining to knowledge discovery: An overview. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 1–35. AAAI/MIT Press, 1996.

[30] K. Ross and D. Srivastava. Fast computation of sparse datacubes. In *Proc. 1997 Int. Conf. Very Large Data Bases*, pages 116–125, Athens, Greece, Aug. 1997.

[31] Y. Taniguchi, A. Akutsu, and Y. Tonomura. PanoramaExcerpts: extracting and packing panoramas for video browsing. In *Proc. ACM Multimedia 97*, pages 427–436, 1997.

[32] L. Teodosio and W. Bender. Salient video stills: Content and context preserved. In *Proc. First ACM Int. Conf. on Multimedia*, pages 39–46, 1993.

[33] V. Tucakov and R. Ng. Identifying unusual spatio-temporal trajectories from surveillance videos. In *Proc. of 1998 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'98)*, Seattle, Washington, June 1998.

[34] W.F. Williams. *Principles of Automated Information Retrieval.* The Business Press, Elmhurst, Illinois, USA, 1965.

[35] WordNet - a lexical database for english. http://www.cogsci.princeton.edu/~wn/, 1998.

[36] O. R. Zaïane, A. Fall, S. Rochefort, V. Dahl, and P. Tarau. On-line resource discovery using natural language. In *Proceedings, RIAO'97*, Montreal, Canada, June 25-27 1997.

[37] O. R. Zaïane and J. Han. Resource and knowledge discovery in global information systems: A preliminary design and experiment. In *Proc. 1st Int. Conf. Knowledge Discovery and Data Mining (KDD'95)*, pages 331–336, Montreal, Canada, Aug. 1995.