Development of a Multimedia Information System for an Office Environment

S. Christodoulakis, J. Vanderbroek, J. Li, T. Li, S. Wan, Y. Wang, M. Papa**, E. Bertino**

Computer Systems Research Institute
University of Toronto
10 King's College Road
Toronto, M5S 1A4

* visiting from CRAI, Italy
** visiting from CNR, Italy

Abstract
We describe an experimental multimedia information system for an office environment which is being developed in the University of Toronto. Multimedia messages are composed of text, image, voice and attribute information. We discuss issues related to internal representation, presentation and communication with the outside world, content addressability in the various data types, user interface and access methods.

1 Introduction
There is a growing interest among computer science researchers in office information systems that handle complex data such as text, attributes, graphics, images, and voice ([VLDB 63a], [VLDB 83b]). We will call the unit of multimedia information a multimedia message. Multimedia messages are composed of attribute, text, image and voice information. Some of the functions that these systems may provide are filing of multimedia information, content addressability of multimedia messages, and multimedia message transmission and reconstruction in a different site.

There are several important problems associated with the development of such systems. Some of these problems are identified next:

1. Query environment
Queries in this environment may be different than queries in traditional Data Base Management System Environments. Users may only have a vague idea of what they are looking for. Their understanding of what they want and how to specify it may increase as they look at other messages. Their query may prove to be inadequate. In that case they may want to reformulate it. Some other users may want to enhance their retrieval capability by specifying some characteristics of the messages that have to do with the presentation form of the messages rather than the content. Finally queries on the image and text part of messages are not often handled by traditional DBMS’s.

2. Content addressability in various data types:
Content addressability in these diverse data types presents serious problems.

Content addressability in messages containing attribute value and text information can be achieved by allowing the user to specify expressions involving the attribute values of the message as well as regular expressions of words appearing within the text message ([Aho et al. 78], [Tsichritzis and Christodoulakis 83]). Structures for efficient retrieval of formatted data from single and multi-file environments have been studied extensively for various retrieval request types and frequencies ([Teorey and Fry 82], [Wiederhold 83], [Ullman 83], [Christodoulakis 84], [Christodoulakis 83a], [Christodoulakis 83b]). Content retrieval from text files has also been studied for various environments and efficient methods have been described ([Salton and McGill 83], [Rijssbergen 79], [Floyd and Ullman 80], [Haskin 81], [Tsichritzis and Christodoulakis 83], [Christodoulakis and Faloutosas 84]).

Content addressability of the image message and voice message part is much harder. One reason is that the current technology on image and voice recognition has serious limitations.

Picture recognition involves very expensive pattern recognition routines ([Tou and Gonzalez 73], [Duta and Hart 72]). In addition picture recognition of generic type pictures is still remote ([Ballard and Brown 82], [Rosenberg 76], [Fu 83], [Pavlidis 77]). Existing experimental and commercial systems based on high power machines (array processors) can be successful only when much knowledge about the scene presented in a picture is available. A second, equally important problem related to pictures is that it is very difficult for the user to specify precisely the picture content of the pictures that he wants.

Speech recognition presents similar problems ([Fu 83], [Reday 73], [Reddy 70], [Erman et al. 80], [Electronics 83]). Currently only speaker dependent, discrete speech, voice recognition devices with a limited vocabulary of words exists in the market.

3. Information Organization and Access:
In an office information environment files are seldom static. The information is usually diverse, the world changes fast, people do not like to spend time for organization and reorganization of information. In addition the good naming, structuring, consistency and quality of information which is assumed in data base environments is not easy to maintain in an office information system. Information may be inserted using error prone methods and the quality of information and the consistency of naming may be
very difficult to maintain. The query capability and the access methods used should be able to cope with these problems. Of course performance is of very high importance in this environment. By specializing the application environment we expect to provide better services (e.g. contentaddressability) and usable systems. Performance may become easily unbearable due to large volume of data and to the flexibility of the query environment.

4. Query Interfaces

Users of office information systems may have very diverse backgrounds varying from the sophisticated type to the naive type. The precise syntax required by many DBMS may not be appropriate for this environment. The high quality screens, voice inputoutput and other sophisticated devices that are now available have the potential for providing very effective interfaces. In this environment it would also be desirable that the query interface facilitates the user to express his queries in a better way.

5. Information extraction and internal representation:

In a multimedia message environment several possible ways of message creation exist. Multimedia messages may be interactively generated in a given station and sent to another station via communication lines. In the receiving station additional editing of the message may take place. Alternatively messages or parts of messages (pictures) may be in a paper form in which case a powerful image segmentation and OCR capability may be used for extracting the information from the messages.

Automatic extraction of information from documents is a difficult task. Segmentation into text regions and image regions is an easier problem and the performance of existing techniques is adequate [Wong et al. 82]. In addition optical character recognition techniques perform well for a variety of fonts. Thus existing techniques can be used for the automatic recognition of the text part of documents.

However, in many cases this may not be adequate. The information contained in various images may contain much redundancy [Pratt et al. 80]. For example if an image of a document contains a simple graph, this graph may be encoded in an internal representation form with much reduced storage requirements. Thus an internal representation may be used to reduce storage requirements as well as telecommunication costs. The internal representation may need to be different from system to system depending on the availability of devices of various types, as well as on the workload characteristics of the system (e.g. CPU bound versus IO bound system, capacity of the communication medium).

6. Data presentation

Data base management systems have traditionally emphasized data organization and manipulation. Presentation of data in devices with diverse capabilities has traditionally been deemphasized [Gray 83]. This problem is particularly important in the presence of bit map display capabilities with different number of gray levels, colors, display sizes, as well as facsimile input/output devices [Horsk 83]. In addition the large storage requirements of image information and the large cost of transmitting this information through communication lines may impose an internal representation of some image types which is different than the presentation form of these images. Finally, voice input and output devices also present similar problems.

In this report we describe an approach for the development of an office information system which handles multimedia messages. We present aspects related to internal representation, presentation, content addressability, access method, user interface, message formation and information extraction.

2. A framework for multimedia messages

In this section we present a conceptual framework for multimedia messages [Christodoulakis 84a]. The conceptual framework describes the logical components of multimedia messages and their interrelationships. The framework is useful for describing the capabilities of the system to users.

The logical components of a multimedia message are shown in figures 1a and 1b. Multimedia messages have a type associated with them and they are composed of one or more of the following: a set of attributes, a voice message, a text message and a set of images. In addition multimedia messages may have an annotation part. The message type contains a minimal common information (a set of common attributes) in a large number of messages.

Attributes have an attribute name, a type, and a value. The value may be a repeating group of values.

The text message is composed of text sections. Each text section is composed of text paragraphs. Each text paragraph is composed of text words. Each text word is composed of overlapping parts of words. This structuring of the text message allows queries to restrict retrieval based on the proximity of words within the text message as well as to associate annotation with each of the text components.

An image is composed of an image type, a vector form, a raster form, a statistical form, and a text part.

The image type can be: graph if it contains at least one graph, pie chart if it contains at least a pie chart, histogram if it contains at least a histogram, table if it contains at least a table, statistical (any of the previous), and picture (anything else).

The vector form represents the image as a set of image objects. An image object is composed of a set of ordered points and an object caption. Points are pairs of values indicating the position of a point within an image. Points of an object may be connected to form lines, polygons, polylines, .... The object caption is composed of object caption words. Object caption words are of type text and they are composed of parts of words.

The raster form represents the image as an ordered set of pixels in two dimensions. The raster form of an image may contain possibly overlapping raster objects which are sets of adjacent pixels. Each raster object corresponds to a distinct vector object of the same picture which is a polygon. The implication is that the set of pixels composing the raster object is defined by the boundaries of the vector object when it is superimposed on the raster form of the image.
The statistical part of the image is composed of a set of tables. Each table has a set of attributes. Attributes have a name, a type, and a set of values. Tables within an image are independent of each other. We do not allow joins among tables. Tables are used internally to store the statistical information contained in images of type graph, pie chart, histogram or table.

The **image text part** is composed of image text words. Image text words are composed of parts of words. The image text part is text related to a given image. The text part is formed by the following:

- The image caption of a given image.
- Text paragraphs related to the image.
- Text annotation.
- Object caption words of objects within the image.
- Attribute names of attributes in the statistical part of the image.
- Attribute values of attributes of the type text in the statistical part of the image.

The **voice message** is composed of **voice words**.

Annotation is composed of **text annotation** and **voice annotation**. Text annotation is composed of **text annotation sections** and voice annotation is composed of **voice annotation sections**.

Annotation may be associated with a text message, text sections, text paragraph, text words, and images. (The lines from voice annotation are not shown in the figure.) Annotation is a further informal explanation about the contents of a message, paragraph, word, image or image object.

### 3. Internal Representation and Presentation Form of Multimedia Messages

The presentation form of the constituents of a message may be different from the internal representation of the message.

The **internal representation** of an image does not have to have both an object form and a raster form. It may only have one of the two. An example of an image where both forms exist in the internal representation is a photograph where objects have been identified and stored in the object form for enhancing content retrieval. An example of an image having only a raster internal representation is an uninterpreted photograph. An image having only an object form as internal representation can be an engineering design. (At the presentation level however, the object form may be used to display the design in a raster display.)

The internal representation of the **object form** of an image is a collection of objects. With each object is stored information related to its type (polygon, circle, . . .), its name, name display specifications (font, size, position of display), shading information, and the coordinates of a set of points or other information specific to object type (radius . . .). This information enables the reconstruction of the set of points which compose an object.

The **internal representation** of statistical type images (graphs, pie charts, histograms, tables) is a collection of tables. This information is not displayed and it is in fact a duplication of information since the information about the objects composing the presentation of these images in a specific device is also maintained. However, the information duplication is not very large usually, and the approach facilitates both answering queries on the image contents and presenting the image in a different form, or the same form but with different parameters (different coordinate system say), at a later point in time. In addition it can be used to display the contents of the image in devices which do not have graphics or bit map display capability.

The presentation of a multimedia message in an output device is called a **physical message**. With a physical message we associate some default information (such as font, size, line spacing, . . .) which is used for displaying the message in an output device.

The structure of a physical message is shown in figure 2. A physical message is divided into **physical pages**. Each physical page is composed of rectangles. A rectangle can be a text rectangle or an image rectangle. Rectangles are identified by their **location** within a physical page and their **size**.

Image rectangles correspond one to one to images of a multimedia message.

Text rectangles may contain some information that is used for displaying messages in an output device (alternative font, alternative size, . . .). Since sequences of words may be displayed in a different way we also use **word sequences rectangles** which are contained within text rectangles.

Finally the voice message and the annotation message part of a multimedia message are not displayed in the physical message. However, the voice part of the message, **voice annotation sections**, and **text annotation sections** are mapped one to one to image rectangles and paragraph rectangles of the physical message. An **indication** of their existence is a special symbol associated with the relevant rectangle, which may be optionally displayed in the output device. The indication symbol can denote **voice indication**, **voice annotation section indication**, or **text annotation section indication**.

A descriptor is associated with each created multimedia message. The descriptor indicates the parts of the message, the internal form for each part and its mapping to a physical message.

**Compression information** may also be encoded in the message descriptor. As we mentioned before, the compression to be used in such an environment depends on the system workload and the devices used. In addition since there may be a variety of techniques that can be used (and none is best for all cases [Gonzalez and Winz 77]) the particular method used (if not the default) and its parameters may be encoded within the descriptor. This may be more important for the image part rather than the text or attribute part of multimedia messages due to the large amount of information in images. The simplest case (which has been implemented in our prototype) is the encoding of an image as a set of objects (and regions of uniform shade). The image is expanded to a complete bitmap at presentation time.
Figure 1a: Multimedia message structure

Figure 1b: Multimedia message structure (Image part)
4. Content Addressability

In our system multimedia messages are retrieved by specifying message content information instead of a unique message identifier. The user will have some idea of what is the content of messages that he wants to see (or not see) and he will specify this information in his query. The system will try to return to him all relevant messages.

We would like to avoid the general pattern recognition problem associated with images in our system, and still provide as much content addressability as possible. In some cases converting image recognition problems to attribute and text recognition problems provides us with a powerful alternative. The system should provide some basic but powerful support for content retrieval.

Image content addressability can be achieved by specifying conditions on the image text part, on the image statistical part, as well as similarity and spatial relationships among image objects.

Retrieving messages based on conditions on the image text part is logically and physically different than specifying conditions on the text part of the message. The former specifies that the user wants to see a message that has an image related to the condition specified while the latter specifies a message related to the condition specified. The search is also limited to the image text part.

An image in our system may contain a number of statistical objects (graphs, pie charts, histograms, tables). Each one of those has an internal representation in the form of a table. In our system the user can focus his attention to only one of the statistical objects at a time. We do not allow content addressability based on relationships among tables. We follow this approach because we believe that it will be confusing to the user to remember which statistical objects belong to the same image. In addition, conditions on single tables may be very selective so that the size of the response is limited. However, the presentation of a message allows that more than one statistical objects (graphs, tables, . . . ) to appear in the same picture.

Finally similarity and spatial relationships of objects in pictures may be found useful in restricting the size of response for non-statistical images. The specification of the query will be done interactively by the user using a set of menu options and the graphics editor [Christodoulakis and Elles 84]. We have not implemented this option yet.

Some examples of possible queries and the way that they can be formulated follow:

Ex1: Give me any documents with images that have to do with IBM

..image type statistical
IBM exists in the text part of the image

Ex2: Give me any documents that have statistics on IBM

..image type statistical
IBM in the text part

Ex3: Give me any documents that have a graph related to IBM

..image type graph
IBM on text part

Ex4: Give me any documents with statistical figures relating IBM sales and year

..image type statistical
IBM sales and year are attribute names (the user may specify partial match on the attribute names if he is not sure about the exact name of the attribute)

Ex5: Give me all documents which have graphs where all IBM sales where greater than CDC sales

..image type graph
.attribute values of IBM sales greater than attribute values of CDC sales

In our system the user is able to specify the following for message retrieval:

0. Message type
1. Conjunctions of attribute values and attribute ranges (=, >, <, ...)
2. Conjunctions of disjunctions of words or parts of words appearing within the text message, text section, text paragraph.
3. Existence of voice.
4. Existence of images
5. Approximate location of an image
6. Conjunctions of words or parts of words appearing within the text related to the image (text, paragraphs, image caption, names of objects, attribute names, attribute values of attributes of type text). This capability may be useful when the user wants to ask queries about an image that he has not seen before. In this case he may not know if the word IBM for example appears as attribute name or value.
7. For statistical images (pie charts, graphs tables...) existence of attributes, attribute values, relationships of attribute values.
9. Existence and relationships of image objects for non-statistical images. Such relationships contain, intersect,... We plan to incorporate in this option some aspects of fuzzy retrieval of images. The user will draw an image and the system will search for images that contain objects similar to it. This option has not yet been implemented.

10. Conjunctions of the above.

3. User Interface and Query Reformulation

The important task of the user interface is to support in a uniform and integrated way the various data forms and activities.

In order to ask a query the user has to specify a filter. The specification of the filter is done using menus in a by-example fashion. This approach allows the user to specify his selection non-procedurally using a set of options and thus it presents advantages for non-expert users [Vanderbroek 84].

The screen is divided into two regions. The left displays a message template. A message template has two main components: 1) A set of fields that correspond to the attributes of the message, 2) the message body. The various components of the template are filled in during the process of query formulation. The right region of the screen is the menu area and displays the available options for definition of restrictions on the voice and image part of the message.

For restrictions on voice the options that can be specified are present, absent, or no restriction.

For image restrictions several options are available:

Location on/off
When the location is on, the user indicates the location by pointing to the approximate position within the message body. A small frame is outlined to indicate the position chosen. A corresponding frame is displayed in a fixed area of the screen called repository. This frame is always used (even if location has not been specified) to display information which has been so far specified about the image.

Image type
The following image types are available: picture, pie chart, histogram, graph, table, statistical, picture. The statistical type permits the user to query for images containing statistics without specifying their representation. The user chooses one of the above types by pointing to the icon representing the image type. The information that the user specifies about the image is inserted in a frame for the image. Frames are displayed below the menu options in the right hand side of the screen. The location of the image within a page (if specified by the user) is used to outline a dotted frame where the approximate location of the image exists. A new menu is displayed indicating the various options for that image type. The user can either quit going back to the previous menu, or specify further details.

The options for each image type for statistical images can be divided into two classes: 1) image type attributes such as column names 2) image attribute relationships. These options allow the user to specify relationships among the attributes of an image type. For example if the type is graph the menu options presented describe objects related to graphs (e.g. axis, lines, ...). This is expected to be desirable for a user that has a particular presentation in mind (possibly because he saw the image recently) and he specifies the type of the image. However, all queries (independent of type) are examined versus a common internal representation (table) [Vanderbroek 84]. Thus the user has the option to specify the type of the image to be statistical if he does not remember (or has not seen) the presentation form of an image which contains statistical information.

After the query specification by the user the system starts the search for finding the desirable messages. Very often the user will not be able to specify a very tight filter. Thus he will often describe a filter to which a superset of the messages that he wants to see qualify. A browsing capability is used to assist the user in identifying the appropriate messages. Miniatures and voice abstractions are used for better browsing [Tsichritzis et al 83].

Miniatures are realistic visual abstractions of messages which are displayed for the user during browsing. Several miniatures can fit on the screen at the same time. Miniatures are automatically generated and stored for each message.

Voice abstractions are voice excerpts associated with a message and highlights the meaning of the message. Voice abstractions may be generated either manually by inserting a voice abstraction message or automatically by using an automatic indexing technique. In the case that an automatic indexing technique is used the abstraction can be stored in a binary form to save space. When the miniature is displayed, the voice output capability uses the binary form to produce voice. The voice abstraction can be turned on or off by the user. We have not implemented the automatic indexing option yet in our system.

In a multimedia information system environment it may often be the case that the user cannot exactly describe the information that he wants. This is not typical of a data base environment where the information is well structured and named, and attributes take values from a fixed set of attribute values. An example from a text retrieval environment which demonstrates that this may not be true in a more general information retrieval environment is synonyms, words with similar meaning. There is a need for query reformulation in this environment.

Dynamic query reformulation in image messages is even more important. The reason is that the information extraction process may fail to name all the existing objects within an image. This may be due to several reasons

1. The person extracting the information was not careful or patient.
2. There is too much information in a picture.
3. Certain objects may not be very clear within a given picture. This will affect both manual and automatic extraction of information. Some of this problem may be avoided by allowing a powerful set of similarity measures and transforms for raster objects (or user defined, query specific transforms and similarity measures).
4. Certain objects are not known or important at
the time that the images were inserted in the
multimedia message repository. Again this is
true for both manual and automatic extraction
of information.

It is possible that the user will feel the need for
query reformulation at some point in time as he
browses through the messages. One reason is that
something that he saw in these messages may have
prompted his memory to a better specification of his
query. Another may be that he has decided that he
is receiving too many documents back. The query
reformulation may restrict the number of qualifying
documents further, it may expand the query with a
disjunctive term, or it may completely change the
query. In our system we allow options for query
expansion using an environment dependent
thesaurus (not implemented yet), query modification
(more restrictions) and continuing the search for-
wards or backwards, or changing the query and res-
tarting without seeing the documents seen so far.

For images the query reformulation capability
should allow the user to extract a part of an image
and to use it for expanding his filter. This will be
useful when a user as he browses through qualifying
messages, sees an object of a class of objects that he
wants. It will be easier for him to extract this infor-
mation from the image itself instead of redrawing
the image. It will also probably result in a more
accurate specification of the query.

A possible different scenario would be that the
user is not able to draw or specify his image objects
very well. Thus he starts by using some text words to
select messages that possibly contain an image
similar to the one that he wants use in his filter. When
he finds one, he extracts the information that he
wants and he uses it as a new filter. One objective of
the interactive image editor that is described below is
to facilitate this type of queries.

Thesaurus mechanisms have been traditionally
used in information retrieval for replacing one word
in a query with its synonyms. An expansion of the
thesaurus idea would be to use thesaurus mechan-
isms that associate words with their pictorial
representation. This type of thesaurus mechanism
can be environment dependent so that excessive
overhead is avoided.

6. Access Method

Multimedia messages coming into a station are
stored in general files. At a later point in time a user
of the station may want to view these messages or
extract some information from these messages to
form a new message. An access method based on
abstractions is used to achieve fast response time in
user queries. An abstraction of the multimedia mes-
sage is much smaller than the multimedia message
itself and restricts the attention to a small number
of qualifying messages.

Information stored in the abstraction file con-
tains abstractions of text image and voice data. The
text abstraction scheme is based on superimposed
coding [Christodoulakis and Faloutsos 84]. A fixed
length block signature is created for each block of
text data. Originally all the bits of the block signa-
ture are set to zero. The signature is constructed by
taking each non-trivial word in the text message
splitting it into successive overlapping triplets of
letters and hashing each triplet into a bit position
within the block signature. The bits are set to one.
If the word is too short, additional bit positions are
created by using a random number generator, which
is initialized with a numeric encoding of the word.
Thus a constant number of bits corresponds to each
non-trivial word. These bits are set to one. The size
of the signatures and the number of bits per word
have been determined in such a way that the perfor-
mance of the system is optimized [Christodoulakis
and Faloutsos 84].

To examine if a given word appears within a logi-
cal block of the message, the signature of this block
is examined. The same transformation is performed
on the word and the bits determined by the transforma-
tion are examined. If they are all one, the word is
assumed to appear in the text message. This access
method retrieves supersets of the qualifying mes-
sages. The browsing capability described before
allows the user to pinpoint the relevant messages.
Parts of words can also be specified in queries. More
complicated query patterns (including conjunctions
and disjunctions of words) can be examined versus
the signature in an obvious manner. Information
related to attribute values is also abstracted using a
signature technique. The only difference is that
order preserving transformations are used in order
to answer inequality queries. Some further evalua-
tion which was done by [Rabitti and Zizka 84] shows
that the approach is more appropriate for an infor-
mation system environment than word signatures
[Tschirritzi and Christodoulakis 83], [Larson 83], or
rigid indexing techniques like IBM STAIR.

Important information regarding images like the
image type and approximate location is also inserted
in the abstraction file. In addition information
related to the objects of a picture as well as an
abstraction of the image text part and the statistical
part is also inserted in the abstraction file.

Finally the only information related to voice
that is inserted in the abstraction file is information
related to the absence or existence of a voice section
in the message.

As with text information, the information used
for answering queries involving attribute picture and
voice which is stored in the abstract file guarantees
that a superset of the qualifying messages to a given
request is retrieved. The blocks of the access file are
accessed sequentially. The sequentiality for access,
the use of large blocking factors, and the small size
of the access file result in a small cost of the access
method.

7. Message Formation and Information Extraction

Messages may be interactively created using a
bit map display capability of a workstation. The text
formatting software may provide the same basic
features seen in traditional formatters [Futura et al.
82]. Alternatively, the formatting software may be
integrated with the filing capability so that new mes-
sages are synthesized from old ones [Christodoulakis

Proceedings of the Tenth International
Conference on Very Large Data Bases.

Singapore, August, 1984

267
Interactive creation of certain image types (graphs, diagrams, pie charts, histograms, tables, drawings) seems easier for the user and the system. The user has an immediate feedback of what his images look like and he will be able to directly edit them and see the result on the screen. The system will possibly avoid some overhead coming from the fact that user encoded specifications do not map exactly to what the user wanted to see, or even if they do, the resulting figure when displayed is not satisfactory.

The interactive image editor formatter in our system is used to assist the user in creating images interactively, extracting information from other images already in the system, or manually editing digitized images, extracting the information in them and possibly discarding the raster form of the image which is expensive to store.

We plan to use the image editor also for specifying or reformulating queries that refer to non-statistical images. Thus the image editor becomes an important part of the management system for multimedia messages: It is useful for message formation, change of presentation form, query specification, query reformulation, information extraction for achieving content addressability, and information compression (when the raster form of the incoming messages is discarded). The image editor should be powerful enough to support these functions. In addition it would provide a nice interface to the user.

The communication with the image editor in our system is based on menus. The right part of the screen is used for communication with the user during interactive design. Menus are small and well identified.

The user gives commands to the system by utilizing mainly a single communication medium (mouse). He uses another medium (keyboard) only when there is no alternative. This in combination with good descriptive menu options, relatively short menus, confidence on where exactly the user is at a particular time in his navigation through the menu options and modes, confidence on what exactly the system expects from him and what the system is doing, are the main principles used for designing a good user interface for interactive image creation.

The image editor provides capabilities for creation and manipulation of individual objects and for creating an internal representation of the image in a vector form. It can also be used to create special objects (graphs, pie charts,...). For these objects an internal representation in a table form is maintained.

The image extraction mode of the image editor extracts information from images of other messages stored in the system, tables and picture bitmaps in order to synthesize a new picture. Information that can be extracted from images of other documents is the raster form of an object or a set of objects, the vector form of an object or a set of objects, and an attribute or a set of attributes. Extracted information may be stored into a subpicture. Thus a new picture can be synthesized from subpictures that have been selected from different pictures. The sub-picture creation mode may be used to adjust the relative positioning of images. This is basically a cut and paste form of creating new images from old ones.

Frequently multimedia messages that have been prepared with this editor formatter may arrive in the system. The system may have to deal with both: Multimedia messages that have been prepared using formatters that are known to the system [Christodoulakis et al. 84a] as well as multimedia messages that have been prepared using unknown formatters. The later is similar to the problem of general document conversion to an internal form [Wong et al. 82].

Information may be extracted from pictures in a raster form using the interactive object form edit mode of the image editor to draw objects on top of the raster form. The image extraction mode is used to read the raster form of the picture first. For special images (graphs, histograms,...) the required attribute information is also extracted at the same time by specifying the values of the parameters that were used in the presentation (minimum and maximum values in the axis, scale, ...). The vector form of images requires in general much less storage space. The raster form of these images may or may not be maintained after the creation of the object form of the image.

Objects within pictures in a bit map form will be defined manually in this mode. The user will indicate on the screen a point or an ordered set of points which composes an object. He may also introduce type, caption and/or annotation information. (This information may be the only information needed since not all images have a raster form.)

In the future the object description for pictures in raster form can become semi-automatic. An edge detector could be used to detect edges and the user may name the objects and annotate. He may also correct mistakes of the edge detector. The edge detector need not be fancy. It will only see the dominant objects of the picture. For simple images specific to our environment (graphs, diagrams,...) the information extraction process may be completely automated.

The image editor is a part of the document formation unit that currently exists in the system. Simple text formatting capabilities complement the unit.

8. Concluding Remarks

In this report we have presented issues related to the development of a multimedia information system for an office environment. Messages in this environment are retrieved based on content. The user can specify attribute value relationships as well as words or parts of words that appear in the text part of multimedia messages. Image content retrieval is achieved by allowing queries on the image text part, queries on the internal table representation and queries on similarity and spatial relationships among image objects. Some aspects of the presentation of messages may also be specified in queries. We presented issues related to the user interface, query reformulation, access method, image creation, information extraction. We described the internal representation and the
presentation form of multimedia messages and the mapping between them.

We are implementing a prototype for a multimedia information system for an office environment based on the framework described in this paper. Multimedia message filing and retrieval have been completed [Vanderbroek 84]. Most of the image formatter is complete at the time that we write this paper (except the information extraction mode). Spatial and similarity retrieval for images has not been implemented yet.

Some relevant or complementary research has been done in other places. Aspects of document segmentation in an office environment are discussed in [Wong et al. 82]. Image compression is discussed in [Pratt et al. 80]. Text retrieval hardware is discussed in [Haskin 82]. Attribute and text retrieval aspects are also discussed in [McLeod 81] and [Stonebraker et al. 83]. Text retrieval for an office environment is also discussed in [Larson 83]. Several systems for image processing and pattern recognition have been implemented in the past ([Kunii and Harada 80], [Christodoulakis 80], [Kulick et al. 75], [Chang and Fu 81], [Economopoulos and Lochofsky 83]). Finally an ambitious design for a system with attributes, text, and pictures is described in ([Penny and Picard 83]).

References

[Christodoulakis 80]

[Christodoulakis et al. 84a]

[Christodoulakis et al. 84b]

[Date 82]

[Erman et al. 80]

[Feiner et al. 81]

[Gray 83]

[Haskin 81]

Proceedings of the Tenth International Conference on Very Large Data Bases.

Singapore, August, 1984


Kulick et al. 75

Kunii and Harada 80

Larson 83

McLeod 81

McLeod 83

Pavlidis 77

Penny and Picard 83

Pratt et al. 80

Rabitti and Zizka 84

Reddy 75

Reddy 76

Rijsbergen 79

Robertson et al. 81

Rosenfeld 76

Salton and McGill 83

Stonebraker et al. 83

Teorey and Fry 82

Tou and Gonzalez 74

Tsichritzis 83

Tsichritzis 83b

Tsichritzis and Christodoulakis 83

Tsichritzis et al. 83

Tsichritzis and Lochovsky 83

Ullman 83

Vanderbroek 84

Verning 77

VLDB 83a
"Panel on Complex Data Objects: text, voice, images: Can DBMS manage them?" Proceedings VLDB83.

VLDB 83b
"Panel on Office Information Systems: What is our role?" Proceedings VLDB 83.

Wiederhold 83

Proceedings of the Tenth International Conference on Very Large Data Bases.

Singapore, August, 1984

[Wong et al. 82]