ERJ
Engineering Research Journal
Faculty of Engineering
Minoufiya University

AN AGENT-BASED IMAGE DESCRIPTION FRAMEWORK

Tarek Helmy[2*] Mohamed T. Faheem Saidahmed [1*] Ahmed Al-Nazer[2]

[1]*College of Engineering, Taif University, Taif, Saudi Arabia email: m_t_faheem@yahoo.com*

[2]*College of Computer Science and Engineering, King Fahd University of Petroleum and Minerals,*
Dhahran 31261, Mail Box # 413, Saudi Arabia, email: [helmy, alnazer]@kfupm.edu.sa

** On leave from College of Engineering, Department of Computers Engineering & Automatic Control, Tanta University, Egypt*

ABSTRACT
This paper presents high quality graphical computer software that helps the visually impaired people to interact with the graphical computer software. We note that almost every document contains some kind of images which make a problem for special users like visually impaired people to interact and understand the context of the document. As we know that there is a support for text to be converted into voices but there exist no support for an image to be converted into a voice. This issue has motivated us to develop a framework that can give a text description of the images, where its real description significantly aids the work of both visually impaired and professionals. Based on this idea, we have developed an Agent-based Image Description Framework (ABIDF) that enables visually impaired people to recognize most common geometrical shapes in addition to some normal images. The framework produces phrases to describe the recognized shapes, and finally submits the generated text description to the voice synthesizer to be converted into speeches. This process will help the visually impaired users to identify these charts and make the required information more available to them. To support the usefulness of our work, experimental applications are introduced and the obtained converted images into voices showed successful and encouraging results of the proposed framework.

نظرا للتطور الهائل والسريع فى برامج الرسم العالية بحيث أصبحت معظم ألمستندات تحتوي تقريباًعلى صور ورسومات توضيحية مختلطه مما قد يسبب مشكلة لبعض المستعملين مثل ذوى ألإحتياجات الخاصة ممن فقدو أبصارهم عند التعامل مع هذه الصور لفَهْم تلك ألمستندات وحيث أنه من المعلومّ وجود برامج دعم لتحويل النَصوص ألمكتوبة إلى صوت مسموع ولكن ليس هناك دعم لتحويل الصور وألرسومات التوضيحية إلى صوت مسموع لذا ومن هذا المنطلق فإننا نقدم هذا البحث لمعالجة هذه الظاهرة. فى هذا البحث فإننا نقوم بتقديم طريقة يتم من خلالها تصنيف الوثيقة طبقاً للنصِ الموجود بها بإستخدام نظم ألتنقيب عن ألبيانات ألحديثةً وكذا التشابه مع بعض النماذج التى تخص فقط ذلك المجالِ التى صنفت إلية الوثيقة. وقد تم استخدام عدد من ألملامح ألجديده للتعرف على الأنماط وتمييزألأنواع ألمختلفة من الصورالإحصائية والبيانية والغير هندسية ومن ثم يقوم بعمل وصف نصى مسموع ذو معنى مكتمل للصور.

Keywords: Human Computer Interaction, Multi-Agent, Image Description.

1. INTRODUCTION

Recently, it is well known that Personal Computer (PC) becomes more user-friendly and popular due to intensive research on Human Computer Interaction (HCI) that increases the accessibility of information to become available at the hand of disabled people which posses the right to exercise the great advantages of the PC like the other human beings. According to estimates that have been carried out by the United Nations http://www.worldbank.org/, there are about 10% of the world's populations are disabled. They are diversified groups and have different kinds of challenges. For the purposes of defining solutions to 'usability for disability', several broad categories should be taken into consideration the following factors: Sensory impairment (e.g. sight and hearing loss), mental impairment (e.g. learning difficulties, aphasia, and rehabilitation requirements), motor impairment (e.g. limited movement, co-ordination difficulties), and visually impaired (e.g. blind users). We note that, working for these disable people is not just only a humanitarian work but also a legal obligation that researchers and their colleagues have to pay for them. It becomes an important concern to include sight disabled users in the new technological revolution which is evidenced through the increasing awareness of the requirements and rights of these users. In this paper, we will mainly

focus on sight disabled people. With higher capability of PC, documents that are not just a combination of bunch of text with some hazy picture but rather becomes a cocktail of several media that can be shown in Figure 1 [22] which contains different contents of a typical modern document. To proceed, we introduce a model that will increase accessibility to multipart documents to be sighted by disabled users. The proposed model is inherited to include some areas that have not received a lot of potential attention in the literature such as: "a voice interpretation of pictures" specially the standard geometrical pictures. We note also that, the main important tasks of most recent computers are to handle a cocktail of several multimedia documents with high probability of having combinations of bunch of text as well as hazy pictures.

It is well known that, lots of researches in the literature are being focused on the field of manipulating documents with multimedia. Such field of interest has become more attractive specially in making some presentations with animation, searching information from scattered documents, searching multimedia data, and etc. There are several strong search engines (e.g. Google, Yahoo, etc.) which basically work on text parts of the documents and extract and store information from them. But still there is no significant solution for automated content description of images that a good content of visual information for blind users. Therefore, describing images for blind people has become an important area of researches. As shown in Figure 1, it can be seen that most of the presentations contain two types of graphical elements:

(a) Pictorial (e.g. photos, digital art, and etc. from various domain)
(b) Regular geometrical shape (e.g. bar chart, pie chart).

We find that it is possible to make fine description for geometrical images where shapes are related (e.g. Pie charts or Bar charts). The rest of this paper is organized as follows. Section 2 covers a literature review of some existing models that help sight disable users and show their incapability. Design and implementing details of the proposed framework is given in Section 3 where different parts of the system (e.g. Text analysis module, Image repository module and Image analysis module) with their internal structure dependencies and relationships are introduced. Section 4 shows the detail construction of the experimental results with some comments and discussion. Finally, Section 5 summaries the conclusion and future works.

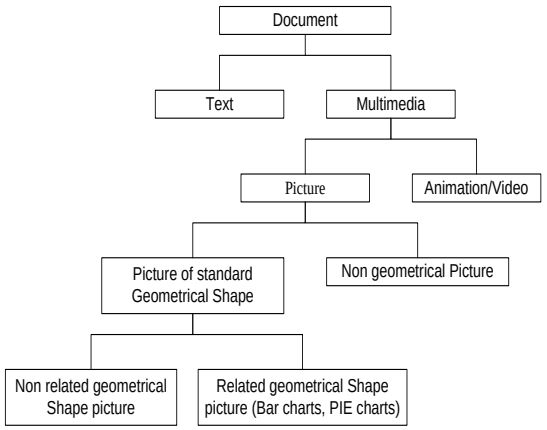

Fig. 1 Parts of a modern document

2. LITERATURE SURVEY

We know that blind people may not have sense of computer pictures or graphics and for these reasons they need to deal with some documents to get some information from these documents that may contain pictures or graphics. In most systems that deal with the blind users they should have some predetermined rules that any graphical or pictorial presentation existing within some documents can be eliminated or converted into voice. It may be possible for simple information to be implemented locally for some kind of institution. But for the vast area of Internet, it is impossible to force all of them to make a dual (text base) version for their presentation. It is well known that most of the developments that have been introduced in computer science literature are targeted for the general users. But still there are some special other kinds of users like visually impaired should receive much more attentions to satisfy their needy life circumstances. Recently, there are several models or systems which have the capability to implement voice that can be operated on the computer systems for blind people. They naturally use voice reorganization system or Brail key board or touching mechanism for collecting user's input that give the output in the form of speech by using some kind of text to speech synthesizer [1, 2]. There are also some other models that can help the blind people to use computers exactly as they are using "Brail Keyboard" [7] or "Voice/command system" [8] to give their instruction to the computers. Nowadays, almost every OS supports the Voice/speech recognition so that the user can give instructions to the computer to carry on their executions.

There is no common solution for all disable people to access information technology. For each kind of disable people there is a specialized area of research. As an example for quadriplegic people, an important area of research is controlling and giving input to a

system with easy and less amount of movement [3, 4]. For example, a wisifd ("windows gadget" see Figure 2) [3] based software system is operated by a single switch. The developed HCIs (Human Compute Interface) are constructed with the frame-wifsid concept and the employment of the scanning technique. The scanning process works independently by scanning the wifsid in several modes and stages, capturing the user's inputs and deciding when a single wifsid has been selected, so as to pause the scan and allow the wifsid to proceed with the rest of the input's handling.

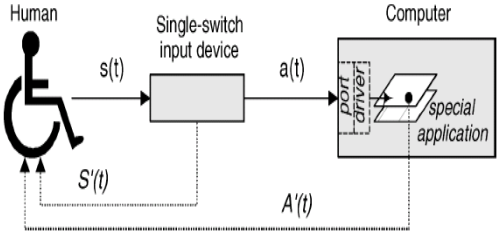

Fig. 2 The human-computer interaction for a quadriplegic person

There are systems for guiding blind people to move around like "Dristi" [5], see Figure 2. There are different models that implement voice for operating the computer for blind people. They usually use voice reorganization system or Brail keyboard or this type of touching mechanism [6, 7] for collecting user input and giving output as a form of speech by using some kind of text to speech synthesizer [8, 9]. Recently after the revolution of WWW accessibility of information becomes a key issue for visually impaired people. A simple approach is to organize the web documents in such a way that will be more meaningful when synthesized by a voice reader [10]. This is a very good approach for text oriented documents but not enough for documents which contain images. An interesting solution is to develop special purpose devices in which they can touch and feel the shapes in an image [11]. For mathematical charts or curves and binary or low color images can be represented in this way but complication arise for critical images which consist of shades and colors. Due to the cost as well as the performance, it does not look a viable solution. In another approach annotation (text or direct voice) can be a way to describe pictures [12]. It is possible to have some kind of automatic annotation of pictures [13] but still annotation is dependent on the manual effort and they are not suitable for special users. Recently Internet becomes a huge source of documents. There are some significant works have been done to increase accessibility for this domain. According to [14], solutions to the problem of Web accessibility for the blind user fall into one of four categories:

- Reliance on a conventional Web browser and a screen reader: Microsoft's Internet Explorer (MSIE) is used by 95% PC all over the world. It becomes a standard for web clients, and most of the web pages designed by the Web developers targeted at MSIE. Using MSIE and a screen reader or magnifier guarantees as an assistive technology that a maximum of Websites will work for the disable user. The problems with this approach are the inaccessibility of content displayed by the browser and the complexity of the user interface.
- Utilizing the accessibility features of HTML and existing Web clients: This approach takes advantage of the principle of HTML which separating content and presentation. Most of the web clients possess the abilities to present content in a way desirable to the user.
- Using proxy servers to convert Webpage HTML into a more accessible format: Requests for WebPages from servers are made not to the servers themselves but to an intermediate server, a proxy, which fetches the page, converts it according to a set of rules, and returns the converted page to the requesting client.
- The final approach is to use a dedicated Web browser designed for visually-impaired or blind people. There are two tactics employed: the first, exemplified by the Home Page Reader from IBM, is a self-voicing application that provides a complete audio interface to Web pages. The second is to render the content of a webpage as a text only flat document and permit the user to access this accessible content using their normal assistive technology, typically a screen reader. Developing a dedicated web browser affords the maximum flexibility in approach, but requires the developer to carry more responsibility for the presentation of Web content.

These are all helpful approaches for visually-impaired people. The problems with these approaches are that they fail to address a range of problems related to overly-complex interfaces (tables and page layout are generally still preserved, so the user must still search over the page for content of interest) and the needs of users without any degree of functional vision. To improve this situation we introduce a platform that will help blind people to get information from pictures in a document by content analysis and similarity matching with previously annotated pictures.

3. FRAMEWORK DESIGN

This section presents the overall framework design aspects, and brief functional description with key concepts. It provides a top-level description of the framework and its major external interfaces to aid the reader in understanding what the framework is to accomplish.

3.1. Subsystem Decomposition

During the subsystem decomposition, we divide the framework into smaller subsystems with a strong coherence. The different subsystems should have a loose coupling. The system will be decomposed based on the use cases and the different actors we have defined. The decomposition shows the existence of the following subsystems, (Figure 3):

- User Interface Subsystem
  - Image Categorization Interface (ICI),
  - Normal Image Descriptor Interface (NIDI),
  - Neural Document Classifier Interface (NDCI),
  - Image Descriptor Interface (IDI).
- Text Analysis Subsystem
  - Weighted Key Repository for several domains
  - Fuzzy Agent for Document Type Identification
- Image Repository Subsystem
  - Templates for Normal Images (NI)
  - Templates for Geometrical Images (GI)
  - Template Search Agent
- Image Analysis Subsystem
  - Image Categorization Agent
  - General Image Agent
  - Geometrical Image Agents
    - Bar Chart Agent
    - Pie Chart Agent
    - Line Chart Agent

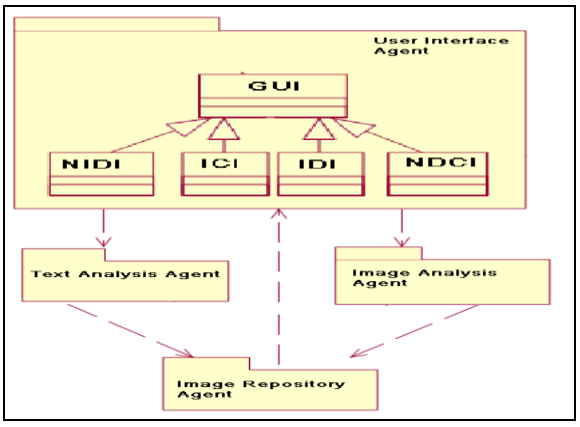

Fig. 3 User Interface Subsystem

3.1.1. User interface subsystem

The user has to load the test image, for which the similarity matching is to be done, by typing the file name or browsing the system repository. User also has to provide the associated text in the text box. By clicking on the 'Search' button, the user can start the similarity searching. The system produces the result according to the selected options, and shows the most similar picture. A similar description will be shown, and the test image's information can be saved or added to the database. The operation for GI is similar to non-geometric images, except that the image file should be GI (e.g. a bar-chart, a pie-chart). The user does not need to choose the image type, as it can be automatically selected by the image categorization agent. Moreover, the system categorizes the document and trains the NDCI. Then the rest of steps will be done automatically to give the meaningful description.

3.1.2. Text analysis subsystem

In this module, the documents will be categorized based on their text contents. We use a Fuzzy Model to implement this module. First, we do supervise training with some pre-categorized documents. In this process, we first extract the tokens of each domain, and then give them some weights according to their frequency of presence in the documents. After the training, we reduce the number of tokens by removing the low-frequency domains, and we make the number of tokens the same for all categories. Figure 4 shows the model, where 'S' denotes the subject, and 'W' the token weight in each subject. It takes the extracted features information from different module agents such as document analysis, image analysis, and image categorization agents. It has two different repositories of images: one for the general images, and one for the geometrical images. According to the image type obtained from the image analysis, the module selects one of these repositories. In the searching process, it uses the extracted features from the image as well as the document to reduce the search domain and to find the most similar template. The image repository module is the main source to get semantic descriptions for the NIs, and it enriches the GIs' content descriptions with semantic meaning.

3.1.3. Image analysis subsystem

The image analysis is the core module of the proposed framework and other modules are supporting components for this module. It mainly consists of different component such as:

(1) Image categorization agent
(2) Geometrical image agent
(3) General image agent

These agents implement several related concepts, including image categorization, simple image preprocessing, shape recognition, feature extraction, etc. We choose the agent base paradigm as it combines knowledge and intelligence with modularity and extensibility [19, 21] and due to our experience in this area [36].

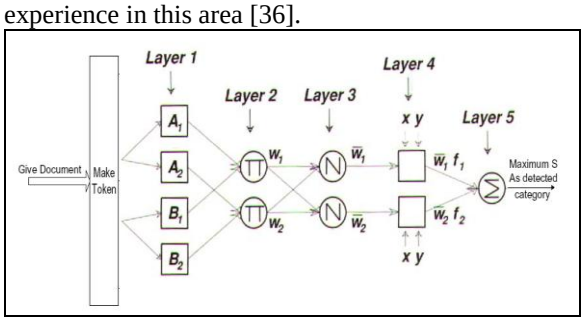

Fig. 4 Text Analysis Subsystem

3.1.4. Charts Recognition Model

The charts (scientific charts) are special kind of graphs with a collection of graphical elements and textual elements which have sort of arrangement. These charts have the graphical representation of some statistics or studies. For example, the bar chart contains x-axis with specific category and y-axis with some values along with set of bars between the x-axis and the y-axis. Each bar has a height which represents the value of the category with respect to y-axis. Because these charts are simple and easy to understand, people find these graphical representations very useful for their analysis and sometimes one chart summarizes pages of reports. There are a lot of charts generation tools and hence there are different representation pattern. We assumed in our study that the charts are represented as an image. There are eight steps required in our model to describe an image with a chart. First, we start with pre-processing the image and then recognize the chart type. After that, we separate the graphic elements from the text elements. Then, we detect the edges of the graphic elements followed by retrieve the text information. Then, we start re-construct the graph again in order to describe it. This starts with association text to related graph and ends with text description of the whole chart.

3.2. Image Pre-Processing

The objective of this step is to pre-process the input image which contains a chart of the three types: (1) Bar Chart, (2) Pie Chart and (3) Line Chart. We assume the user will have a colored image as a chart input to the prototype system. So, the pre-processing step will convert the image to binary image (0s and 1s) in order to process it and recognize the chart type in it. Figure 5 shows the image before and after the pre-processing.

3.3. Text and Graphics Separation

Typically, any chart contains two types of information: graphical and textual. The graphical represent the chart itself whether it is pie chart, bar chart or line chart. The text illustrates the labels on that chart which represents the categories of the columns and (if found) the values of these categories with respect to the y-axis. The objective of this step is to separate the chart from the text. So, we will have one big graph for the chart and multiple text elements which are used as labels on the chart. This is achieved by component separation that is built in any scientific tool like MATLAB which is responsible to separate the component of a figure. The biggest component is representing the chart and the other small components represent the text labels surrounding the chart. Figure 5 illustrates an example of this separation.

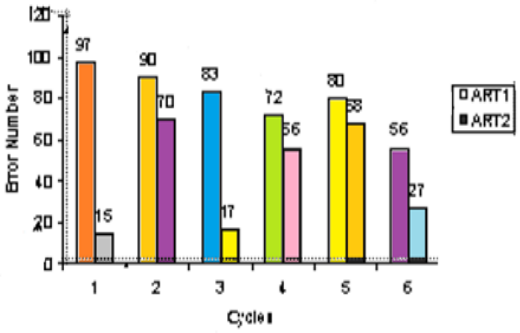

Fig. 5-a Neural Network Training Result

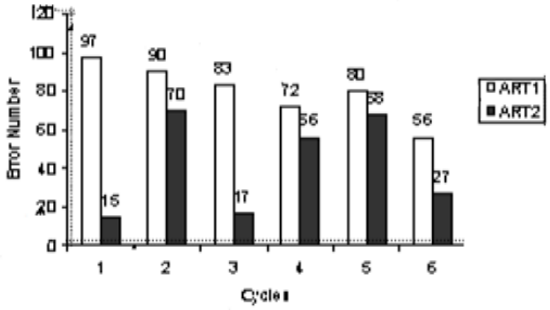

Fig. 5-b Neural Network Training Result

Fig. 5 Example of Image Pre-Processing (a before) and b after)

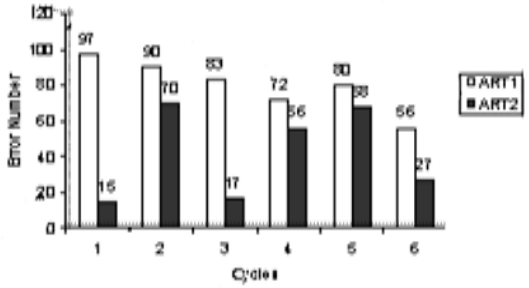

Fig. 6-a Original Image

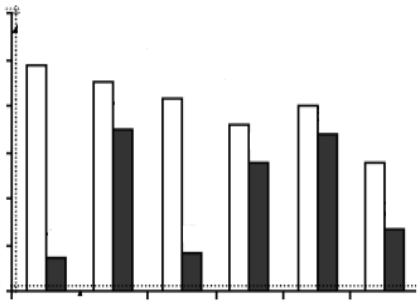

Fig. 6-b Graphics Extracted

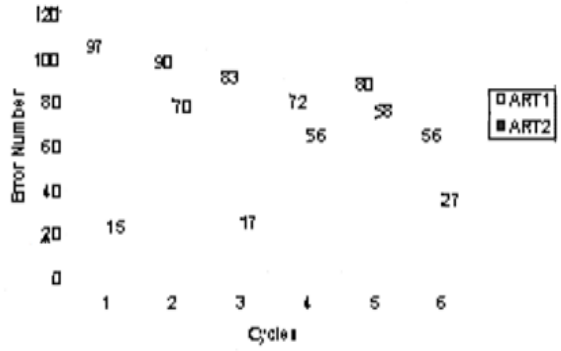

Fig. 6-c Text Extracted

Fig. 6 (a~c) Text and Graphics Separation

3.4. Text Information Retrieval

The chart normally contains wealth text information to help the reader understanding the chart itself. The text is scattered through the chart whether it is above the bars (in bar chart), inside the pies in the pie chart or above the line in the line chart. Also, the y-axis contains textual labels for the values of the x-columns (in bar and line charts. In this step, we extract all the text found in the graph with its x and y coordinates (which will help later when associating the text to the graphical elements of the chart) and then using OCR to extract the text and map it to the parts of the graph (Figure 6).

3.5. Chart Type Recognition

The objective of this step is to classify the input charts into three types (Figure 7): (1) Bar Chart, (2) Pie Chart and (3) Line Chart. Using the x-axis and y-axis feature with projection technology for classifying the chart type. A projection is the transformation of points and lines in one plane onto another plane by connecting corresponding points on the two planes with parallel lines. The branch of geometry dealing with the properties and invariants of geometric figures under projection is called projective geometry. First, with projection of all points in the graph, we can have two classification categories:

(1) Bar Chart and Line Chart,
(2) Pie Chart.

Clearly the first category is obvious because both the bar chart and the line chart have x-axis and y-axis and some graphical elements in between. This step will separate the pie chart because its projection will never exceed a pre-defined threshold that identify if it is not bar chart non line chart. Second, we need to classify and identify the bar chart from the line chart. This can be done through another projection with different threshold where the bar chart contains bars that will projected as parallel to the y-axis (each bar has two y-columns) while the line chart does not have these columns and have only one line drawn in the area between the x-axis and the y-axis.

3.6. Edge Detection

Any chart has graphical components that construct the whole chart. For example, bars construct the bar chart and connected points (with a line) construct the line chart. There is a need to identify and detect the edge between each graphic component (e.g. bars). In pie chart, the intersection between the two lines in the circle center is identifying one area with an arc and angle. In the bar chart, two vertical lines connected in the top points with a horizontal line are determining the bar. In line chart, the point connected to another point by a line is identifying an edge. Figure 8 shows some examples.

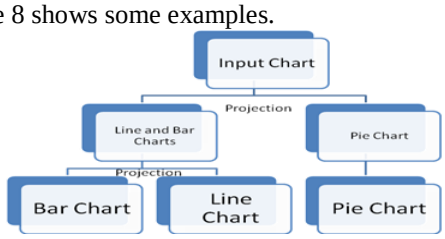

Fig. 7 Chart Type Recognition

Fig. 8 Detecting Edges in Bar Chart and Pie Chart

3.7. Chart Model Construction

The objective of this step is to reconstruct the input chart with the information extracted and the features found in our analysis steps. This will re-build the chart with this information and put structural output (like a table) summarizes all found information. The chart re-construction will play a role to verify and compare the resulting output to the input and find the variance if any to help in analyzing the results and finding the error rate.

3.8. Text-Graphics Association

Previously, we have separated the text from the graphics elements and in this step we join them back. Now, we have identified the chart type, the elements of the chart and their values so we need to assign these graphics elements to their categories. Going back to the text information retrieval step, we have identified all labels with their x and y coordinates. In this step, we correlate these labels to the graphics elements of the chart. Using the geometric properties of the graphical elements of the chart, we find the nearest label whether it is along with the x-axis or the y-axis and then assign it to the proper one. As in pie-chart, the label is either inside the pie elements or surrounding the pie and the same applies on the line chart and the bar chart.

3.9. Text Description

The objective of this step is to generate the final text that represents the description of the input chart. To meet this objective, we have created a template that is filled and used in this step. The template merges the information from the previous steps and put all of this information in one paragraph describing the whole chart. First, it states the type of the chart itself. Then, it reads the caption if exists that is the title of the chart. Another way to make up the title if the caption is not available is to correlate the x-axis title with the y-axis title to make the chart title. Next, we describe the graphical elements of the chart by assigning the category (on x-axis) a value within the y-axis. If the value is already labeled or it can be calculated by projecting a line to the y-axis and then calculate the relative value represents this category (in bar chart and line chart). In the pie chart, we can calculate it from the angle of the arc that represents a category. Figure 9 shows an example of output description of the chart. The highlighted text is changeable based on the input chart.

> The input figure is bar chart that is showing the average revenue for the past 5 years for company X. It shows $500,000 for 2005, $350,000 for 2006, $450,000 for 2007, $570,000 for 2008 and $650,000 for 2009.

Fig.9 Text Description Template

3.10. Image Analysis

Image analysis is a vast area. In Figure 10 a flow diagram of such a system is shown. It contains several steps and each step needs its own kind of specialization to work with. We mainly work on feature extraction process to find some suitable features that can be used for image categorization and similarity matching. Image categorization is an important area of research. It can be used for specialized field like medical image processing [15] or simple content based classification [16] for image library. Features in this process can be varied differently. In general they are classified as local or global features [17]. We use some common features like color histogram [18, 22, 23] or contour of a shape for further processing. We have devised some new features especially for image categorization. Features extraction for special purpose is more suitable and accurate than general image features [16, 20, 22] where authors use concavity, horizontal color projection for script categorization that are more useful. Finding suitable feature is a very challenging task. In our case the main hurdle were to find representative features for different classes and prove those features capability as a distinguisher. In general, automatic categorization as a mapping of images into pre-defined classes involves three basic principles:

(i) Representation, i.e. the extraction of appropriate features to describe the image content,
(ii) Adaptation, i.e. the selection of the best feature subset regarding discriminative information, and
(iii) Generalization, i.e. the training and evaluation of a classifier.

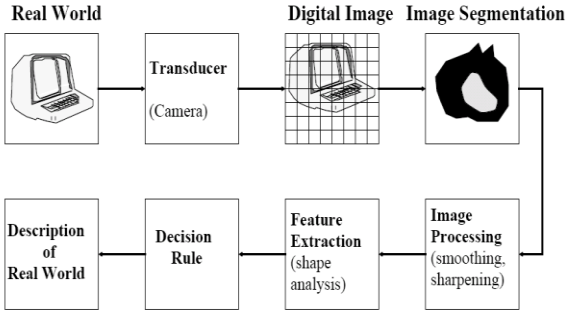

Fig. 10 A flow diagram of a typical image analysis system

We use six features for classification purpose [22] in which four are defined by us. In Section 4, we describe these features in details. We also use some very domain specific features in geometrical image analysis (e.g. the length and width of a bar in bar charts).

3.11. Shape Detection

After categorizing sample images into two categories as "Normal Image" (NI) or "Geometrical Image" (GI), we go for two different approaches to describe them. We need to find geometrical shapes from GI to describe them individually or collectively (as a related GI). So shape recognition with their extractable parametric information (e.g. radius and center position of a circle or length, width and position of a rectangle) is one of our major concerns. There are some advanced techniques for finding shapes from a picture [22]. Jagadish in [13] detects a shape in rectilinear form which can have very different optimal descriptions based on general rectangular cover. Yan and Chew in [25, 26] showed a method based on 'Modified Probabilistic Hough Transform' algorithm for parallel lines clusters detection to detect and recognize bar charts in a document image. It is also possible to use techniques based on neural network for this purpose. For simplicity we use typical method to extract different objects from the figure; we take color difference from background as an indication of shape. Relative positioning of objects in an image is also an important feature. We use row or column scanning to get the relative positions of objects [27]. We only use this technique for geometrical images. There are some models already developed which combine the above concepts for image retrieval. An extensive survey on these types of systems is performed by the authors of [28] (Figure 11). Most of the systems are based on either annotation or conceptual and semantic text retrieval with extracted image features information. In our approach we would rather emphasize on simple categorization by using statistical neural classifier on associated text. In our model, we determine the type of the documents (e.g. scientific, business, officials etc.) by word tokens and use it to choose similar pictures on that domain that increases the probability of similarity.

3.12. Document Analysis

A document contains valuable text that has some context meaning for images. Some investigation on text content can be fruitful to find more refined search of images. We analyze text for categorization purpose. It is the assignment of natural language texts to one or more predefined categories based on their content [28]. Text categorization problems possess several distinct characteristics. First, text categorization problems normally involve high dimensionality. Because the features usually represent key-words or tokens derived from the textual content of documents which can contain thousands of words and their different metaphors. Second, the high dimensional space leading to a very sparse data representation which increases the search space but each document contains only a small number of features. Third, the number of potential relevant features is very large, but only a few occur in a particular document. Features (Key words/tokens) duplication is a common phenomenon as keywords can simultaneously involve in several categories. So finding the bonding strength or tendency to a group is critical. Finally, in general, the overlap between features in documents is quite small that means small number of tokens as well as less effort in training will not find the pattern properly [29].

There is some learning or mining techniques to do the job. There are other methods of document categorization; one of the popular approaches is LSI (Latent Semantic Indexing) [30, 31]. LSI is capable of automatically extracting the conceptual content of text items. One of the facilities of this approach is that it can process arbitrary character strings and it is not restricted to work with words only. We did not use this method as it mainly targets extensively on information collection. Authors of [32, 33] propose a model for text indexing and textual similarity based on creating a representation in terms of conceptual word-chains. Some extensive survey in this issue has already been done [33]. Most of the work in this area targets the similarity matching between documents according to their conceptual meaning and semantics.

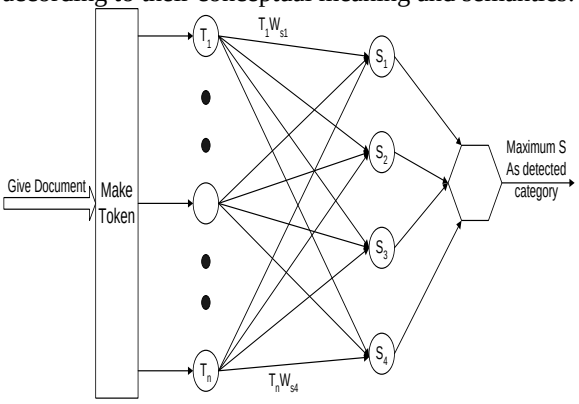

Fig. 11 A diagrammatic view of neural document classifier

3.13. Similarity Matching

Information matching can grossly be divided in two categories- Exact matching and Similarity matching. Information matching in images is basically similarity matching when it is used for image retrieval. Similarity searching is an excellent approach for getting information from subjective materials like images or videos. Popular imager search engine in WWW incorporate this process for retrieval of queried image, Figure 12 shown a general

image handling approach. On the other hand exact matching is used in recognition purpose like in OCR (Optical character recognition) or barcode reading. We mainly use similarity matching for our purpose. Image matching process can be done by global features or local features [17, 24] extracted from images. There are some general features common to all kinds of images like-color histogram (global feature) or shape of silent objects (local feature). But domain specific knowledge gives more meaning to these features. Although there are some generally developed systems QBIC, KMeD [17] for similarity matching, it looks more meaningful when domain specific knowledge is incorporated with such systems. There are some good works are done in medical sector, satellite images [35] and some other domains [34]. Most of the above systems are based on content base image retrieval (CBIR) approach. In this system some standard features and their extraction process are selected to represent the content of an image. When an image is stored in database its features information is extracted by previously defined process then this information is stored with some indexing schema. Indexing process itself is a big research issue which is very important for faster image retrieval from mammoth image database. In searching steps the example image is analyzed first and its content information is extracted as a form of features information and then these features are compared with the stored image's features and retrieve those images which have similar feature values [15]. There is some variation in searching by drawing approach where query is given in a form of interactive manual drawing which is mostly outlines of targeted shapes. We follow query by example approach for similarity matching in our implementation.

In the proposed framework we implement similarity matching as the key for describing an image in text. We develop two sections where in one section we have no domain specific knowledge (for Normal Images) on the other hand we use very domain specific knowledge to formulate similarity matching (for Geometrical Images).

4. IMAGE CATEGORIZATION INTERFACE

An Image Categorization Interface (ICI) is developed to support the image categorization agent. A picture is analyzed here by extracting the following features [18, 22].

- Number of colors
- Percentage of background color
- Rate of horizontal color discontinuity
- Rate of vertical color discontinuity
- Rate of horizontal color overlap
- Rate of vertical color overlap

We have to go through an empirical testing for the above features. So we design this module to work in two phases as:

- Testing phase (for empirical proof)
- Automatic description phase(work as an image categorization agent)

In the testing phase, it analyzes the picture and shows the result in six text boxes labeled as:

- Number of Colors
- Percentage of Background Color
- Horizontal Colors Discontinuity
- Vertical Colors Discontinuity
- Horizontal Color Overlaps
- Vertical Color Overlaps

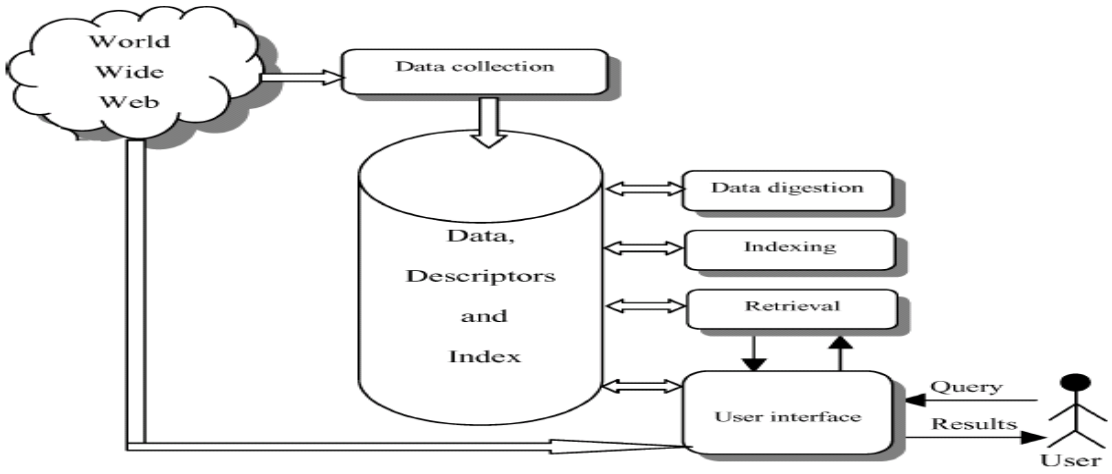

Fig. 12 General structure of a Web image search engine and its main tasks

Tester has to select the picture type. Then it supports to save the result in a MS Access table by clicking "Save Result" button. We ensure that no picture information is duplicated by checking the path of image before saving it. Description of all associated fields of the Access table is given in Table 1. We use this result later for analysis purpose to find a suitable combined threshold value to discriminate between different kinds of pictures. In the automatic description phase, the user needs to open the image file and then click on "Start Automatic Description" button. It will analyze the picture and calculate the combined feature value according to the following formula 1 and 2 [18, 22]:

$$cv = \frac{(\eta + \frac{1}{\beta} + \alpha_1 + \alpha_2 + \theta_1 + \theta_2)}{\lambda \times \omega} \quad (1)$$

Another formula for this purpose which we found more balancing for diversified image set as:

$$cv = \frac{\eta}{256} + (1 - \frac{\beta}{100}) + \frac{(\alpha_1 + \alpha_2 + \theta_1 + \theta_2)}{\lambda \times \omega} \quad (2)$$

The description of the notation is as follows

- Number of colors – η
- Percentage of background color – β
- Rate of horizontal color discontinuity – α_1
- Rate of vertical color discontinuity – α_2
- Rate of horizontal color overlap – θ_1
- Rate of vertical color overlap– θ_2
- Image Height– λ
- Image width– ω
- Combine value- *cv*

Finally it moves to either NIDI or GIDI depending on the threshold value we set for discrimination purpose. We set it as .4 after extensive testing when we use equation (1). It works as an image categorizers in this phase.

4.1. Normal Image Descriptor Interface (NIDI)

We have designed NIDI agent to describe Normal Images too. It combines three different modules as:

- Image Repository Module (IRM)
- NI agent
- NDCI.

Now we will describe NIDI functionality. We need a mechanism to store feature information and textual description of the presented images. This information will be used later for similarity matching that finally yields the image description. So this module has to support information storing process for IRM. We built it in such a way that it can work in two phases as:

1. IRM information storing phase
2. Automatic description phase

4.2. Image Repository Module

In this phase we create an image repository database. To store an image as a template in the repository, it has to go through the following steps:

- The user has to load a template image by typing the file name and its path or browsing the system.
- The user has to select the domain of the image by choosing an area from dropdown box in the upper left pan of the window.
- The user can also provide a textual annotation/description.
- Then the user has to click on the 'Add this figure' button. It will initiate the image analysis process by NI agent which will extract and process the features information and send them to the IRM database.

Table 1, ICI Information Storing Schema

Field Name	Data Type	ICI bindings	Description
ID	AutoNumber		Automatic record number provided by access itself
Ptype	Number	'Selected Picture Type': Dropdown Combo Box	Stores picture type as number according to tester choices (e.g. 0 for NI, 1 for GI bar charts etc.)
Nc	Number	'Number of Colors': Text Box	Stores number of distinct color presented in the image
Pbck	Number	'Percentage of Background Color': Text Box	Stores background color percentage of the image
Nhcd	Number	'Horizontal Colors Discontinuity' : Text Box	Stores total horizontal color discontinuity
Nvcd	Number	'Vertical Colors Discontinuity'	Stores total vertical color discontinuity
Nholp	Number	'Horizontal Color Overlaps': Text Box	Stores total horizontal color overlaps
Nvolp	Number	'Vertical Color Overlaps': Text Box	Stores total vertical color discontinuity

5. RESULTS AND DISCUSSION

In this section, we will describe the results obtained while testing the proposed model in various aspects. We implement all the modules in such a way that they can act autonomously which facilitate us to test them individually; like testing NDCI and Categorization agent independently. We also tested their performance in combined form when they produce output by collaboration according to our conceptual model. We also show the results graphically. In testing phase we targeted the six features individually that we mentioned earlier to:

- Check their different combinations to find an optimal feature vector
- Find a suitable threshold value as well as a precise and combined formula to calculate it which can be used later for classification purpose.
- Test neural document classifier's performance
- Verify the similarity matching improvement with NDCI.

In addition to the experiments we conducted and presented earlier. We have tested the system with different chart diagrams and we will show only two complete examples: one with bar chart (Figure 13 (a) ~ 13 (d)) and one with pie chart (Figure 14 (a) ~ 14 (c)) as followings.

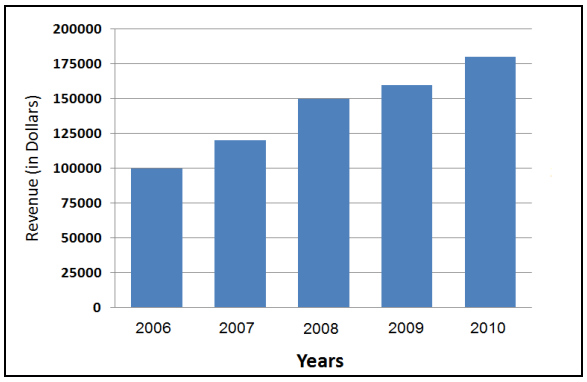

Fig. 13 (a) Input Bar Chart

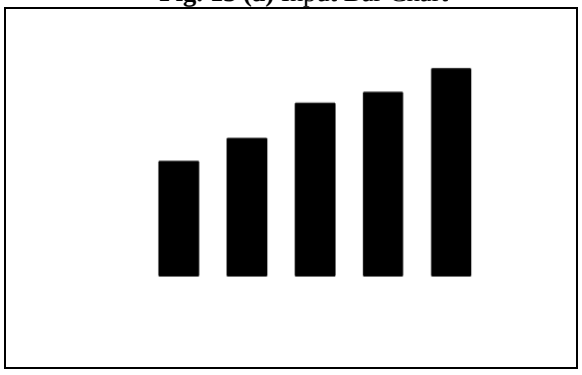

Fig. 13 (b) Isolated Bar from Text

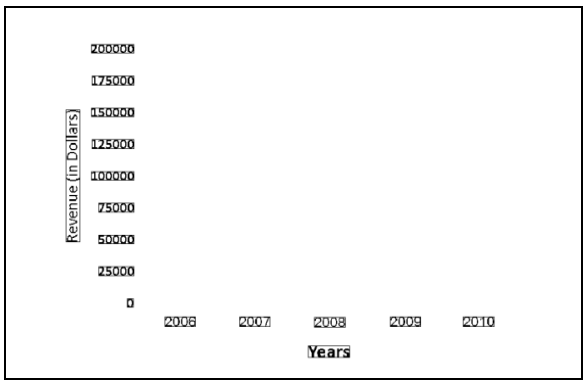

Fig. 13 (c) Isolated Texts

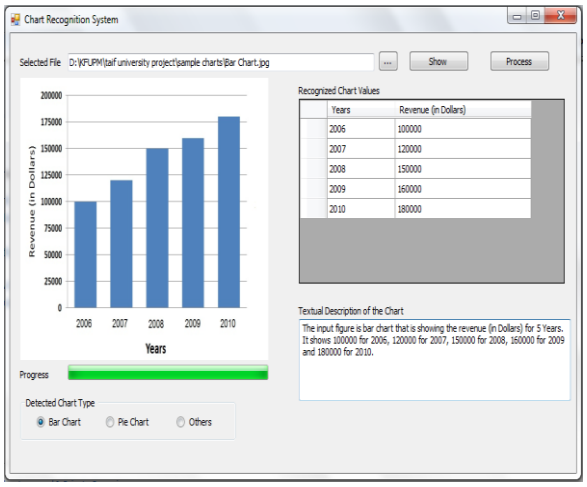

Fig. 13 (d) System Screen for the Bar Chart

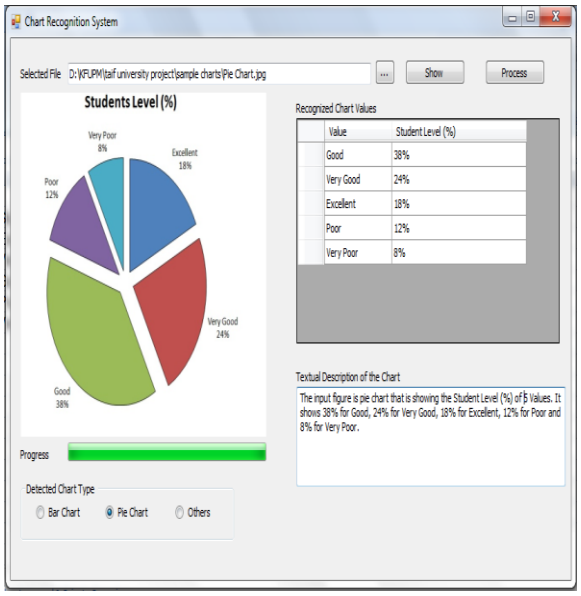

Fig. 14 (a) System Screen for the Pie Chart

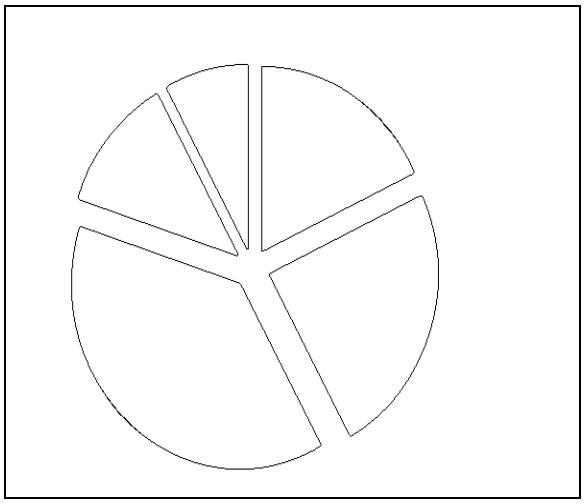

Fig. 14 (b) Isolated Pie Components from text

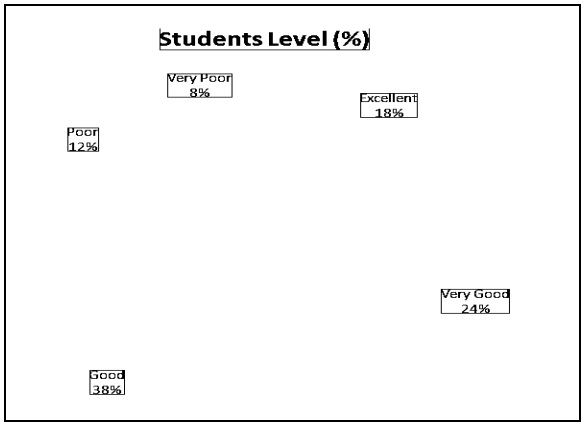

Fig. 14 (c) Isolated Texts

6. CONCLUSION AND FUTURE WORKS

The main objective of the presented framework is to recognize pictures of standard geometrical shapes (i.e. bar chart and pie chart) and then describe them in a text format. The resulting text then can be input to any text-to-speech system to produce a voice synthesized form to help the visually impaired users. The processing includes recognize the shape, describes it and then prepares it for any text-to-voice tool to make sound description of the shape. Similarity searching is a very popular theme in image searching, but we did not find any significant research specialized on statistical or geometrical images for blind users. In our model we use similarity searching to enhance textual description of geometrical images. Describing an ordinary image is a tough issue as the possibilities are unbounded. So in most of the case the best way is to categorize the image domain and then device some appropriate approach for each category. We have devised new features especially for image categorization. We test their different combination to build a suitable classifier. We find the category of an image by analyzing associated text. We have devised a statistical approach for this purpose where we implement a document classifier that can categorize documents after learning the capability by supervised training. In the framework we use mathematical equation of geometrical shapes to build the shape agent that searches the shapes in a picture to fit them with a pre-recognized shape model. Later we enhance the idea by implementing more composite agents that can recognize standard charts. To enhance the description we introduce specialized similarity matching techniques. On the other hand low matching can be achieved by checking the color sequence only. After matching we get the similar description of a presented image. We have also implemented a simple similarity scheme for finding previously annotated images from ordinary image database. This is done by following segmented color histogram approach for similarity matching. To improve the situation we incorporate domain search. This reduces the search space which yields closely similar images that brings more semantic meaningful description. As we plan for making a full image descriptor system, our future work will focus on having some kind of semantic meaningful description and procedure to describe ordinary images.

7. ACKNOWLEDGEMENT

We would like to thank Taif University and King Fahd University of Petroleum and Minerals for providing the financial support and the computing facilities that have been introduced to support and develop this work. Special thanks go to anonymous reviewers for their insightful comments and feedback, resulting in a significant improvement in the quality of this paper.

8. REFERENCES

[1] S. Christodoulakis, "A mixed-mode message system " , Proceedings of the 7th International ACM conference on Research and development in information retrieval, 1984, pp.1 - 20

[2] William A. Barry, John A. Gardner, and Randy Lundquist, "Books for Blind Scientists: The Technological Requirements of Accessibility", 1994.

[3] Christopher R. Murphy, "Computers Assisting The Handicapped"http://ei.cs.vt.edu/~cs3604/lib/Disabilities/murhpy.at.html

[4] Christopher M. Bisop, "Neural Networks for Pattern Recognition", Oxford University, Inc, 1995

[5] Xu-Hong Xiao and Graham Leedham, "Signature Verification by Neural Networks with Selective Attention", Applied Intelligence, Vol. 11, Number 2, September 1999, pp. 213–223

[6] AK Jain, L Hong, S Pankanti, R Bolle, "An Identity-Authentication System Using Fingerprints", Proceedings of the IEEE, VOL. 85, NO. 9, September 1999.

[7] Thomas G. Kieninger, "The Growing up of Hyper Braille, an office workspace for blind people", ACM Symposium on User Interface, pp.67-73.

[8] B. Manaris, R. McCauley, V. MacGyvers, "An Intelligent Interface for Keyboard and Mouse Control Providing Full Access to PC Functionality via Speech", Proc. of 14th International Florida AI Research Symposium (FLAIRS-01), 2001.

[9] Wasfi Al-Khatib, Y. Francis Day, Arif Ghafoor and P. Bruce Berra, "Semantic Modeling and Knowledge Representation in Multimedia Databases", IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, January 1999.

[10] Euripides G.M. Petrakis and Christos Faloutsos, "Similarity Searching in Medical Image Databases", IEEE Transactions On Knowledge And Data Engineering, Vol. 9, No. 3, 1997

[11] Glenn Heaiey and Amit Jain, "Retrieving Multispectral Satellite Images Using Physics-Based Invariant Representations", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 18, No. 8, August 1996

[12] Mihael Ankerst, Hans-Peter Kriegel and Thomas Seidl, "A Multistep Approach for Shape Similarity Search in Image Database", IEEE Trans. On Knowledge & Data Engin., Vol. 10, No.6, 1998

[13] H. V. Jagadish, "A Retrievable Techniques for Similar Shapes", Proceedings of the 1991 ACM SIGMOD international conference on Management of data, Colorado, United States, pp. 208 – 217

[14] Alberto Del Bimbo, Enrico Vicario and Daniele Zingoni, "Symbolic Description and Visual Querying of Image Sequence using Spatio-Temporal Logic", IEEE Transactions On Knowledge & Data Engin., Vol. 7, No. 4, 1995

[15] Mohammad M. Hassan, S.M.S Islam, Md. Golam Kaosar, "Virtual Ear: A Center Based Visual Clustering Approach", Proceedings of IEE International Conference on Intelligent System, Kuala Lumpur, Malaysia, 8-D-1.

[16] Tarek Helmy, Mohamed M. Hassan, "Graph Descriptor" An Approach to Convert Standard Geometrical & Statistical Figures into Text and Voice", IEE Proceedings of International Conference on Intelligent Systems (ICIS-2005), pp. 180-186, December 1-3, 2005 in Malaysia.

[17] Stuart Russell and peter Norvig, "Statistical Learning Methods", Artificial Intelligence: A Modern approach, Prentice Hall, 2003.

[18] Mohamed M. Hassan, Tarek Helmy, Mohamed Sarfraz, "Geometrical versus Non-Geometrical Image Categorization Using Horizontal and Vertical Color Features", Book chapter: 17 of a book titled "Geometric Modeling and Imaging, modern techniques and application", pp. 102-107, published by IEEE, 2008.

[19] Giovanni Casella et.al., "An agent-based framework for sketched symbol interpretation", Journal of Visual Languages & Computing, Volume 19, Issue 2, April 2008, Pages 225-257.

[20] Ahmed Bashir and Latifur Khan, "A Framework for Image Annotation Using Semantic Web", University of Texas at Dallas, 2005.

[21] Sabyasachi Saha and Sandip Sen, "Agent Based Framework for Content Based Image Retrieval", American Association for AI, 2004.

[22] Tarek Helmy, Mohammad M. Hassan, Muhammad Sarfraz, "A Hybrid Computational Model for an Automated Image Descriptor for Visually Impaired Users", Elsevier Journal of Computers in Human Behavior, Vol. 27, Issue 2, 2011, pp. 677-693.

[23] Huang, Wei-Hua, Tan, Chew Lim, Leow, Wee Kheng, "Associating text and graphics for scientific chart understanding", ICDAR05 (II: 580-584), 2005.

[24] Weihua Huang, Chew Lim Tan, Wee Kheng Leow, "Model-Based Chart Image Recognition", GREC 2003: 87-99, 2003.

[25] Yan Ping Zhou, Chew Lim Tan, "Chart Analysis and Recognition in Document Images", ICDAR 2001: 1055-1058, 2001.

[26] Yan Ping Zhou, Chew Lim Tan, "Bar Charts Recognition Using Hough Based Syntactic Segmentation", Diagrams 2000: 494-497, 2000.

[27] Alberto Del Bimbo, Enrico Vicario & Daniele Zingoni, "Symbolic Description and Visual Querying of Image Sequence using Spatio-Temporal Logic", IEEE Transactions On Knowledge And Data Engineering, Vol. 7, No. 4, Aug. 1995 Pages: 609 – 622.

[28] M. L. Kherfi And D. Ziou, & A. Bernardi, "Image Retrieval From the World Wide Web: Issues, Techniques, and Systems", ACM

Computing Surveys, Vol. 36, No. 1, March 2004, pp. 35–67

[29] S. Dumais, J. Platt, D. Heckerman, and M. Sahami. Inductive learning algorithms and representations for text categorization, Proce. of ACM-CIKM'98.

[30] Wai Lam & Yiqiu Han," Automatic Textual Document Categorization Based on Generalized Instance Sets and a Meta-model", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, May 2003, pp.628-633

[31] Stuart Russell and peter Norvig, "Statistical Learning Methods", Artificial Intelligence A Modern approach ,Prentice Hall, 2003, pp.736-748

[32] Anthony Zukas & Robert J. Price, "Document Categorization Using Latent Semantic Indexing" , http://www.contentanalyst.com/

[33] Charu C. Aggarwal and Philip S. Yu, "Effective Conceptual Indexing and Similarity Search in Text Data", Proceedings First IEEE International Conference on Data Mining, USA, pp.3-10, 2001.

[34] Yiming Yang, "An Evaluation of Statistical Approaches to Text Categorization", Information Retrieval , Kluwer Academic Publishers, 1999

[35] Glenn Heaiey & Amit Jain, "Retrieving Multispectral Satellite Images Using Physics-Based Invariant Representations", IEEE Transactions On Pattern Analysis & Machine Intelligence, Vol.18, No.8, Aug. 1996, pp. 842–848.

[36] Tarek Helmy, Ali Bahrani and Jeffery Bradshaw, "Agent-Oriented Service Model for Personal Information Manager", Lecture Notes in Computer Science, Springer, Vol.5907, pp.24-40, 2009.