Interactive Multimodal Visual Search on Mobile Device

Priyanka Pakhare
Department of IT
GSMCOE, PUNE

Karishma More
Department of IT
GSMCOE, PUNE

Prof. Pradnya Velhal
(Project Guide)
Department of IT
GSMCOE, PUNE

Karan Nagane
Department of IT
GSMCOE, PUNE

Abstract— This paper gives information about interactive multimodal visual search on mobile device. We mainly focus on the color features, color presents an interesting problem (namely the 'cross-talk' of features). Multimedia includes a combination of text audio, still images, animation, video or interactivity content form. The searching method will search a database of multimedia object to locate objects that match a query object exactly. Multimedia search queries as textual request and through other media. And this process is often referred to as search by example, because the typical interaction consists submitting a piece of information (e.g.., video, an image or piece of audio) at the purpose of finding similar multimedia items

Keywords— Visual search, Cross-talk, Query object, Multimodal search interface.

I. INTRODUCTION

Multimedia refers to content that uses a combination of different content forms .multimedia is usually recorded and the played, displayed or accessed by information content processing device, such as computerized and electronic devices, but can also be part a live performance. Multimedia system should have the capability to store, retrieve, transport and present data with varies heterogeneous characteristic such as text, images, graphs and sounds. Multimedia query language should provide predicates for expressing conditions on the attribute, the content and the structure of multimedia object.

II. REQUIREMENTS OF MULTIMEDIA

A. PREDICATE

Query predicate can be classified as:

1) Attribute predicates
2) Structural predicates and
3) semantic predicates.

In this paper,
(1) Attribute predicates concern the structured content attribute of the multimedia object. Examples of attribute are the speaker of an audio object, the size of an object and the type of an object.

(2) Structural predicates concern the structure of data being consider. Example of this type is "Find all multimedia object containing at least one image and a video clip".

(3)Semantic predicate concern the semantic content of the queried data depending on the features that have been extracted and stored for each multimedia object. Example of this type is "Find all the object containing the word OFFICE". The main difference between attribute predicates and semantic predicates is that in semantic predicate an exact match cannot be applied. i.e.., there is no guarantee the object retrieve are 100% correct or precise. The result of a query involving semantic predicates is a set of objects which has an associated degree of relevance with respect to the query.
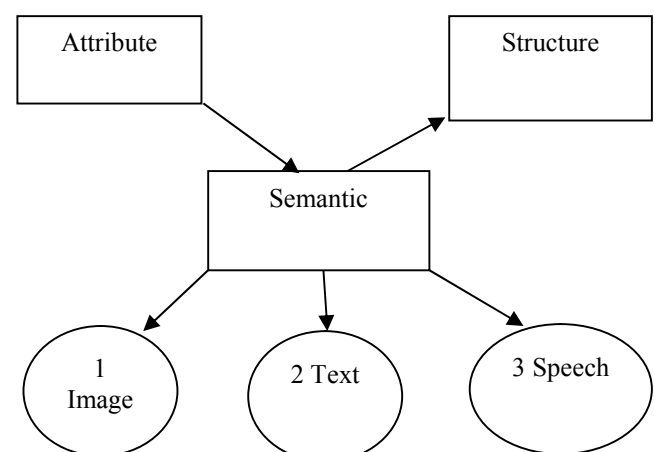

Fig 1:Inputs and variables.

B. SEARCHING METHODS

We focus on here is the design of fast searching method that will search a database of multimedia objects to locate object that matches a query object, exactly or approximately. Object can be two dimensional color images, gray-scale medical images in 2-D or 3-D, one dimensional time series, digitized voice or music, audio clips, etc. A typical query by content would be, e.g., 'in a collection of color photographs, find once with the same color distribution as a sunset photograph. 'Google, Yahoo, Bing provide image search as part of their service. In all this image search engine the query specified textually; no visual input i.e. called as text based image search. The most of the major search engines gives the user the opportunity to search for audio and pictures. But most of them only allow to search with a mono modal interaction, based on textual keyword search.

.

C. APPLICATIONS

Multimedia finds its application in various areas including, but not limited to, advertisements, engineering, medicine, mathematics, business, scientific research and spatial temporal applications. Several examples are engineering, Industry and document imaging. Multimedia models for screening assessment of long-range transport potential and overall persistence. Multimedia system that generalize the concept of QoS to all layers of its software architecture. A number of multimedia files representative of video or images information are stored in a set of directories, which are each characteristics of a predetermined surface area required to display the image or video information contained within the multimedia files. Application of multimedia CD-ROMs in school Multimedia technology in manufacturing a review Multimedia information networking.
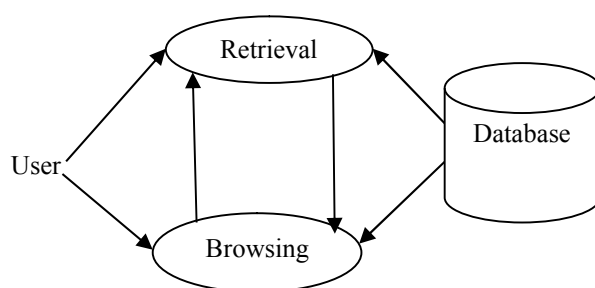

Fig 2:Interaction of user and database.

D. COMPARISIONS OF VARIOUS METHODS

Sr NO	Steganography Techniques	Cover Media	Advantages
1.	Text Technique	Document	Alterations not visible to the human eye
2.	Image Hiding: 1) LSB (Least Significant Bit	Image	Simple & easiest way of hiding Information.
3.	DCT 2) (Direct Cosine Transform)		Hidden data can be distributed more evenly over the whole image in such a way as to make it more robust
4.	Video Technique	Video Files	The scope for adding lots of data is much greater

Fig 3: Interaction of user

III. PROPOSED WORK

The data model, the query language, the access and storage mechanisms of a multimedia information system supports object with very complex structure. While traditional system or conventional system deals with data type such as strings or integers. Multimedia system handles multimedia data while traditional system handles textual unstructured data. Traditional systems are unable to support the mix of unstructured and structured data and different kinds of media. Traditional system does not support metadata information such as that provided by data base schema which is fundamental component in a database management system. A multimedia information retrieval system requires some form of database scheme because multimedia applications need to structure their data at least partially.

Multimedia information retrieval system requires handling metadata which is crucial for data retrieval. Whereas traditional information retrieval system don't have such requirement [10].Traditional system handles attributes bases queries i.e. set of attributes. But multimedia information retrieval system answer attribute based as well as content based queries i.e. set of features. In traditional system object retrieved by query processing are exact and precise. While in multimedia information retrieval system exact match cannot be applied means there is no guarantee that the object retrieved by this type of predicate are 100% correct or precise. Queries in traditional type system are 'Retrieve all names having Ids between EMP10 to EMP100'. While in multimedia information retrieval system queries are of the

type 'Retrieve all the cars manufactured by same company and with different color' [5].

A.   ALGORITHMS AND EQUATIONS

1. *Porter stemming algorithm:*

The Porter stemming algorithm (or 'Porter stemmer') is a process for removing the commoner morphological and in flexional endings from words in English. Its main use is as part of a term normalization process that is usually done when setting up Information Retrieval systems.
Points of difference from the published algorithm:
There is an extra rule in Step 2,
(m>0) logi → *log so archaeology* is equated with *archaeological* etc.
The Step 2 rule
(m>0) abli → able   is replaced by
(m>0) bli → ble   so *possibly* is equated with *possible* etc.
The algorithm leaves alone strings of length 1 or 2. In any case a string of length 1 will be unchanged if passed through the algorithm, but strings of length 2 might lose a finals, so *as* goes to *a* and *is* to *i*.
These differences may have been present in the program from which the published algorithm derived. But at such a great distance from the original publication it is now difficult to say [4].
It must be emphasized that these differences are very small indeed compared to the variations that have been observed in other encodings of the algorithm [1].

2. *Stemming algorithm*:

There are several types of stemming algorithms which differ in respect to performance and accuracy and how certain stemming obstacles are overcome. A simple stemmer looks up the inflected form in a lookup table. The advantages of this approach is that it is simple, fast, and easily handles exceptions. The disadvantages are that all inflected forms must be explicitly listed in the table: new or unfamiliar words are not handled, even if they are perfectly regular (e.g. iPads ~ iPad), and the table may be large. For languages with simple morphology, like English, table sizes are modest, but highly inflected languages like Turkish may have hundreds of potential inflected forms for each root [2].
A lookup approach may use preliminary part-of-speech tagging to avoid over stemming. The production technique
The lookup table used by a stemmer is generally produced semi-automatically. For example, if the word is "run", then the inverted algorithm might automatically generate the forms "running", "runs", "runned", and "runly". The last two forms are valid constructions, but they are unlikely [3].

3. *Suffix-stripping algorithms:*

Suffix stripping algorithms do not rely on a lookup table that consists of inflected forms and root form relations. Instead, a typically smaller list of "rules" is stored which provides a path for the algorithm, given an input word form, to find its root form. Some examples of the rules include: if

the word ends in 'ed', remove the 'ed' if the word ends in 'ing', remove the 'ing' if the word ends in 'ly', remove the 'ly'
Suffix stripping approaches enjoy the benefit of being much simple to maintain than brute force algorithms, assuming the maintainer is sufficiently knowledgeable in the challenges of linguistics and
Morphology and encoding suffix stripping rules. Suffix stripping algorithms are sometimes regarded as crude given the poor performance when dealing with exceptional relations (like 'ran' and 'run'). The solutions produced by suffix stripping algorithms are limited to those lexical categories which have well known suffixes with few exceptions. This, however, is a problem, as not all parts of speech have such a well formulated set of rules. Lemmatization attempts to improve upon this challenge. Prefix stripping may also be implemented. Of course, not all languages use prefixing or suffixing.

Equation:
The cosine of two vectors can be derived by using the Euclidean dot product formula:

$$a.b = ||a||\,||b|| \cos\theta$$

Given two vectors of attributes, *A* and *B*, the cosine similarity, *cos(θ)*, is represented using a dot product and magnitude as [8].
The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity. For text matching, the attribute vectors *A* and *B* are usually the term frequency vectors of the documents. The cosine similarity can be seen as a method of normalizing document length during comparison. In the case of information retrieval, the cosine similarity of two documents will range from 0 to 1, since the term frequencies (tf-idf weights) cannot be negative. The angle between two term frequency vectors cannot be greater than 90° [6] and [7].
Cosine similarity is related to Euclidean distance as follows. Denote Euclidean distance by the usual,$(||A-B||)$ and observe that $||A-B||^2 = (A-B)^T \; (A-B) = ||A||^2 + ||B||^2 - 2A^TB$ by expansion. When *A* and *B* are normalized to unit length [9],
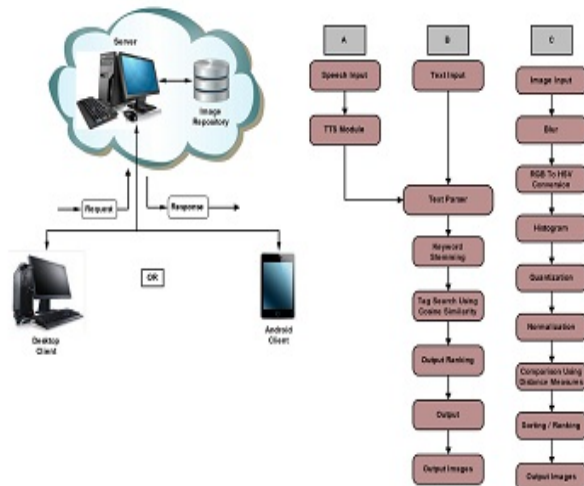$||A||^2 = ||B||^2 = 1$ so the previous is equal to $2(1 - \cos(A - B))$

Fig 4:Interactive multimodal flowchart.

IV.   EXPERIMENTAL RESULT

Experiments are performed in three stages. In the first stage settings in that one million images from commercial search engine. In the second stage objective evaluations are performed, first 100 test queries are determined then normalized discounted cumulative gain and find out response time. The third stage is increase the usability of user. One successful search with single component requires around 30 seconds every failed trial or extra component will increases the time by about 20 seconds.
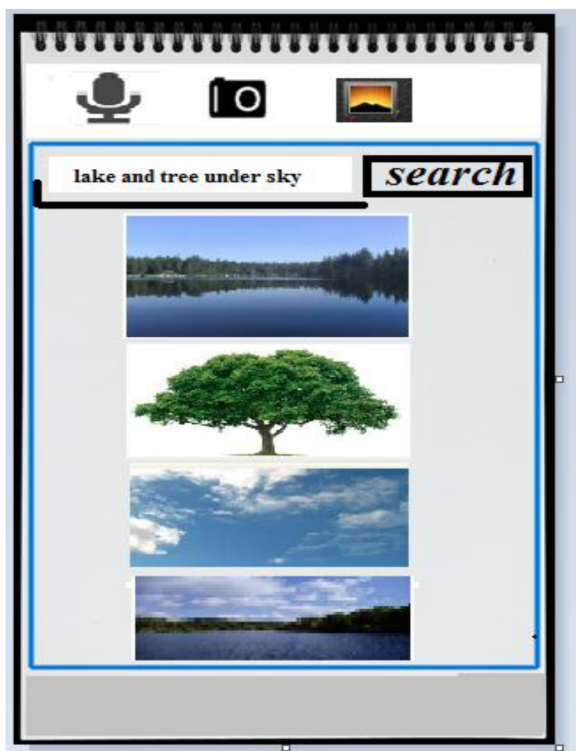

Fig 5:Image Result

V.   CONCLUSION

Introduce a new interactive visual search system on mobile. Propose a visual search method for this application. Deployed The system on a WP7 mobile phone. Improve the efficiency of Visual search. Handle relative positions between objects. Future studies can be directed to investigate the distinctive characteristics of multimedia needs and searching depending on the types of multimedia and lead to an information behavior model for multimedia resources.

VI.   REFERENCES

1.   Apache Open NLP (http://opennlp.apache.org/index.htm) includes Porter and Snowball stemmers
2.   Snowball on C# (http://www.iveonik.com/blog/2011/08/snowball-stemmers-on-csharp freedownload/)- port of Snowball stemmers for C# (14 languages)
3.   Python bindings to Snowball API (http://snowball.tartarus.org/wrappers/guide.html)
4.   Christopher McKenzie, who contributed the Javascript stemmer, demonstrates his program (and therefore demonstrates the Porter stemming algorithm) at qaa.ath.cx/porter_js_demo.html
5.   D. L. Hall and J. Llinas, "An introduction to multisensory data fusion," *Proc. IEEE,* vol. 85, pp. 6–23, Jan. 1997.
6.   ^ Singhal, Amit (2001). "Modern Information Retrieval: A Brief Overview". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4): 35–43.
7.   ^ P.-N. Tan, M. Steinbach & V. Kumar, "Introduction to Data Mining", Addison-Wesley (2005), ISBN 0-321-32136-7, chapter 8; page 500.
8.   ^ Cross validated:Distribution of dot products between two random unit vectors in RD(http://stats.stackexchange.com/questions/85916/distribution-of-dot-products-between-two-random-unit-vectors-in-mathbbrd)
9.   ^ Ochiai A. Zoogeographical studies on the soleoid fishes found japan and its neighboring regions.2//Bull.Jap.Soc.sci.Fish.1957. V. 22. No 9. P. 526-530.
10.   M. Gales and S. Young, "The application of hidden Markov models in speech recognition," *Foundations and Trends in Signal Process.*, vol.1, no. 3, pp. 195–304, 2008.