ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2 Issue Special 1 Jan-March 2014
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

Performance Enhancement of an Image Retrieval by Integrating Text and Visual Features

[I]Najimun Nisha S, [II]Mehar Ban K.A
[I]Final Year ME, Dept. of CSE, S.Veerasamy Chettiar College of Engg & Tech, Puliangudi
[II]Assistant Professor ,Dept. of CSE, S.Veerasamy Chettiar College of Engg & Tech, Puliangudi

Abstract
An image retrieval system is a computer system for browsing, searching and retrieving images from a huge database of digital images. Most traditional and ordinary methods of image retrieval use some method of adding metadata such as captioning', keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. There are two types of image retrieval such as text-based image retrieval and content-based image retrieval. Content-based image retrieval (CBIR) – the application of computer vision to the image retrieval. CBIR aims at avoiding the use of textual descriptions and instead retrieves images based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query image or user-specified image features. Both techniques individually have both advantages and disadvantages. In our work, a vertical search engine is proposed to improve image retrieval performance by integrating text and visual features. It is implemented as vertical product search engine for apparels shopping by discovering relationship between textual features and image features. User intention gap and semantic gap is bridged by integrating textual and visual features. Visual features are reweighted according to their significance to the query terms. Hence, precision and recall of content based image retrieval is improved.

Keywords
CBIR, Vertical Search Engine, Semantic gap, Annotation, user-supplied query image

I. Introduction

Recent years have seen a rapid increase in the size of digital image collections. Everyday, both military and civilian equipment generates giga-bytes of images. A huge amount of information is out there. However, we cannot access or make use of the information unless it is organized so as to allow efficient browsing, searching, and retrieval. Image retrieval has been a very active research area since the 1970s, with the thrust from two major research communities, database management and computer vision. These two research communities study image retrieval from different angles, one being text-based and the other visual-based. Avery popular framework of image retrieval then was to first annotate the images by text and then use text- based database management systems (DBMS) to perform image retrieval.

Many advances, such as data modelling, multidimensional indexing, and query evaluation, have been made along this research direction. However, there exist two major difficulties, especially when the size of image collections is large (tens or hundreds of thousands). One is the vast amount of labour required in manual image annotation. The other difficulty, which is more essential, results from the rich content in the images and the subjectivity of human perception. That is, for the same image content different people may perceive it differently. The perception subjectivity and annotation impreciseness may cause unrecoverable mismatches in later retrieval processes. In the early 1990s, because of the emergence of large-scale image collections, the two difficulties faced by the manual annotation approach became more and more acute. To overcome these difficulties, content-based image retrieval was proposed. That is, instead of being manually annotated by text-based key words, images would be indexed by their own visual content, such as color and texture. Since then, many techniques in this research direction have been developed and many image retrieval systems, both research and commercial, have been built. The advances in this research direction are mainly contributed by the computer vision community.

II. Related Works

A. Content–Based Image Retrieval

Due to increasing demands of managing pictorial data such as art galleries, medical image archiving, trademark signs, etc., the development of efficient image retrieval systems becomes extremely important. Recently, many content-based image retrieval (CBIR) systems have emerged to satisfy some of the needs [4], [5] . Although these modern image databases are queried by image content, the query image still needs to be formulated into some internal form for efficient execution.

In general, the purpose of CBIR is to present an image conceptually, with a set of low-level visual features such as color, texture, and shape. These conventional approaches for image retrieval are based on the computation of the similarity between the user's query and images via a query by example (QBE) system [3]. Despite the power of the search strategies, it is very difficult to optimize the retrieval quality of CBIR within only one query process. The hidden problem is that the extracted visual features are too diverse to capture the concept of the user's query. To solve such problems, in the QBE system, the users can pick up some preferred images to refine the image explorations iteratively. Early image retrieval systems manually annotate images with metadata, and use text-based retrieval to search on tags. However, manual annotation is extremely time consuming, especially for very large scale image databases.

Also, it is difficult to describe images accurately with a controlled set of keywords. Discrepancies between the query terms and tagging terms lead to poor retrieval results. They still incur some common problems, namely redundant browsing and exploration convergence. To tackle such issues, Content Based Image Retrieval systems were developed to use visual features such as color and texture to index and retrieve images. The primary challenge in CBIR is the semantic gap between the high-level image content and the low-level visual features. CBIR techniques also tend to be computationally intensive due to the high dimensionality of

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

visual features.

Feature Extraction: Feature (content) extraction is the basis of content-based image retrieval. In a broad sense, features may include both text-based features (key words, annotations) and visual features (color, texture, shape, faces). However, since there already exists rich literature on text-based feature extraction in the DBMS and information retrieval research communities, we will confine ourselves to the techniques of visual feature extraction.

Within the visual feature scope, the features can be further classified as general features and domain specific features. The former include color, texture, and shape features while the latter is application-dependent and may include, for example, human faces and finger prints. The domain-specific features are better covered in pattern recognition literature and may involve much domain knowledge which we will not have enough space to cover in this paper.

B. Text-Based Image Retrieval

Text-based image retrieval. Current web image search engines like Google Image Search and Bing primarily rely on textual metadata. They take textual queries and match them with the metadata associated with the images, such as image file name, URL, and other surrounding text in the webpage containing the image[7]. Since the textual information surrounding an image may not necessarily describe the image content, and it is difficult to describe visual content using text, the retrieval performance of the metadata-based searches can still be poor.

III. Proposed System

We propose a novel method, Integrating Text and Visual Features to achieve the high precision and recall of CBIR in coping with the large-scale image data.

A. Data Collection

Textual and Visual data is collected for generating a repository for image search. Data is collected with text and images. The text based image data is stored as name of the image, its category and its full description. For visual features of images, edge histogram is considered from image texture. They are four directional edges and a non-directional edge. Four directional edges include vertical, horizontal, 45 degree, and 135 degree diagonal edges. These directional edges are extracted from the image-blocks. If the image-block contains an arbitrary edge without any directionality, then it is classified as a non-directional edge. Normalized edge histogram descriptor is collected for image retrieval process. For textual data, we record the name, category, full textual description, and so on for each product. Visual product images are collected and it tags and annotations are also extracted. Both textual data description and visual feature and tag description are utilized for vertical search.

B. Text Processing

Query entered by the user is obtained and pre-processed to get the important terms from the query. Pre-processing includes Stopwords Removal – Stop words, i.e. words thought not to convey any meaning, are removed from the text. This process removes all words that are not nouns, verbs or adjectives. For example, stop words removal process will remove all the words like: he, all, his, from, is, an, of, and so on. Stemming - The stem is the common root-form of the words with the same meaning appear in various morphological forms (e.g. player, played, plays from stem play). Stemming will find the stems of the output terms to enhance term frequency counting process because terms like "learners" and "learning" come down from the same stem "learn". This process will output all the stems of extracted terms. Similar words for the terms obtained is identified and used for image search. The actual term and the similar terms are combined and retrieve results from repository. This is the initial result set that is obtained from text processing.

C. Visual processing

From the image collected, We capture the spatial distribution of edges with five edge strengths generated from an edge histogram descriptor. The five edges in the edge histogram descriptor are horizontal, vertical, 40 degree diagonal, 135 degree diagonal and non-directional. After the edge extraction from image-blocks, we count the total number of edges for each edge type in each sub-image. Since there are five different edges, we can define five histogram bins for each sub-image. Then, since there are 4x4=16 sub-images, we have total 16x5=80 bins for the edge histogram. These 80 histogram bins are the only standardized semantics for the MPEG-7 edge histogram descriptor. However, with the local histogram bins only, it is not sufficient to represent global features of the edge distribution.

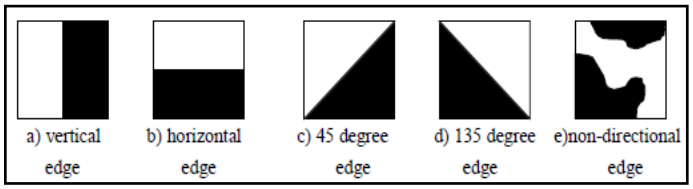

Fig 1: Five types of edges

Note that to improve the retrieval performance, we need global edge distribution as well. In this paper, we generate the semi-global and global edge histograms from the local histogram bins. Then, the global, semi-global, and local histogram bins are used to evaluate the similarity between images. The normalized value of these edges are obtained for visual processing. Based on the normalized value, the images are further processed for image comparison and retrieval.

D. Weighting Visual Features

Base representation of the term in the visual feature space is computed. There are N items with term T in their descriptions, and each item is represented by a M-dimensional visual feature vector. The mean vector of the N feature vectors is used as a base representation of term T in the visual feature space. Mean vector is computed for the number of items containing the term. When N is large enough, mean vector will preserve the common characteristics in the image features and smooth over the various sections. In such a manner, the mean vector is rendered as a good representation of the keyword. The mean vector will preserve the common characteristics in the image features and smooth over the various sections. Two distributions such as positive and negative distributions are computed and weight vectors are computed for comparison.

E. Ranking

A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second. In mathematics, this is known as a weak order or total pre-order of objects. It is not necessarily

a total order of objects because two different objects can have the same ranking. The rankings themselves are totally ordered. For example, materials are totally preordered by hardness, while degrees of hardness are totally ordered. Based on the terms and weight vector, the set of images are ranked. The weight vector is computed for different terms mentioned in the user query. For each term, value of weight vector is compared and the term which with high valued weight vector is ranked first. Based on the ranking, the images are retrieved and displayed to user. Hence we achieve high accurate image retrieval by combining visual and textual features.

F. Integrate Textual and Visual Features

Our goal is to integrate textual and visual features in vertical search by learn the visual representations for textual terms and obtain the semantic interpretations of visual features. We present a vertical search engine which uses integrated model to improve image retrieval performance.

The system is comprised of three major components: the Crawler, the Pre-Processor, and the Search and UI component. Data is collected with text and images. The text based image data is stored as name of the image, its category and its full description.

For visual features of images, edge histogram is considered from image texture. They are four directional edges and a non-directional edge. These directional edges are extracted from the image-blocks. If the image-block contains an arbitrary edge without any directionality, then it is classified as a non-directional edge. Normalized edge histogram descriptor is collected for image retrieval process. User intention gap and semantic gap is bridged by integrating textual and visual features. After visual and text based image data collection, user query is collected and pre -processed by applying stop words and stemming to get important terms from user query.

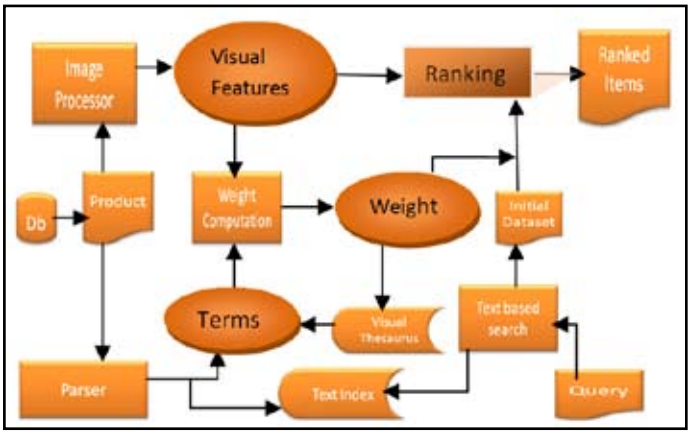

Fig. 2: Integrate Text and Visual Features

On getting the terms, text-based image retrieval is done by comparing the terms and the descriptions of images. Visual features are reweighted according to their significance to the query terms and the retrieved images using text based retrieval. Based on the visual features, the images are ranked and displayed to user. Hence, precision and recall of content based image retrieval is improved by integrating text and visual features of images. Weighting scheme infers user intention from query terms, and enhances the visual features that are significant toward such intention.

Semantics of Local Edge Histogram: After the edge extraction from image-blocks, we count the total number of edges for each edge type in each sub-image. Since there are five different edges, we can define five histogram bins for each sub-image. Then, since there are 4x4=16 sub-images, we have total 16x5=80 bins for the edge histogram.

Normalization and Quantization of the Bins: After generating local edge histograms for all 16 sub-images, we need to normalize each bin in the histogram by dividing it with the total number of image-blocks with an edge in the corresponding sub-image. Then, each histogram bin has a value ranging from 0 to 1. To represent the normalized bin values in binary form, we need to quantize them. Since the normalized bin values are normally distributed in a small range (say, from 0 to 0.3), bin values are non-linearly quantized. The quantization tables are obtained by adopting the Lloyd-Max algorithm. Then, assigning 3 bits per bin we have total 3x80=240 bits to represent the local histogram .

Three major components:

1. The Crawler fetches product pages from retailer websites.
2. A customized parser extracts item descriptions and generates the term dictionary and inverted index.
3. Simultaneously, the image processor extracts visual features from item images.
4. Next, we integrate textual and visual features in a reweighting scheme, and further construct a visual thesaurus for each text term.
5. Finally, the UI component provides query interface and browsing views of search results.

IV. Experimental Data

A text query Q, It first retrieves an initial result set {Ii}. The results are displayed in columns, with Ii in the title row. We also use traditional CBIR as baseline. It skips query expansion, and uses the original feature vector qi (extracted from Ii) as the query, to find visually similar items from the database.

A. Experimental Results and Comparison

To compare integrating visual & Text with baseline. First, 100 distinct Q in the form of "adjective noun" (e.g., night and black) are randomly selected. Five items from the initial set of each Q are randomly selected as seed images to generate 500 queries in total. For each query, the top 10 results from integrating visual & Text and baseline (they could overlap), together with 20 randomly selected items from the same category as Ii, are prepared for user evaluation. For each participants have experiences with search engines.

For each query, the participant is provided with Q and prepared items (permuted), and asked to mark items that he/she determines to be relevant with Q. We received results from 76 Qs and 201 queries. Comparing with baseline. In the table, a true positive (TP) is a retrieved item (CBIR) that is confirmed (marked as relevant) by the evaluator; a false positive is a retrieved item that is marked as irrelevant.

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Fig 3: Text query

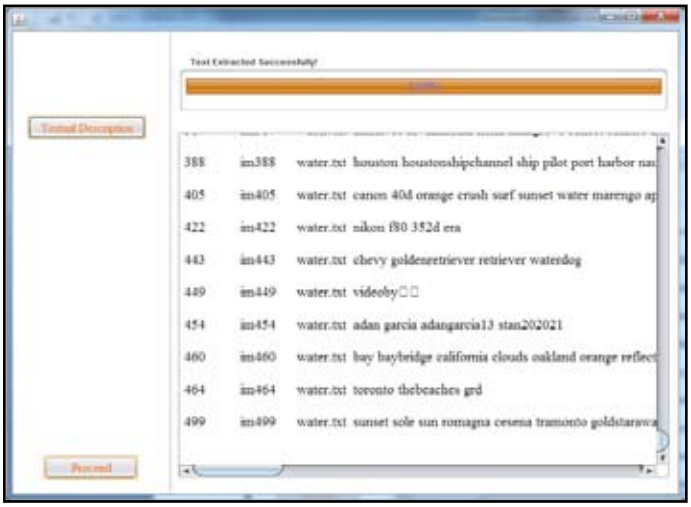

Fig. 4: Textual Description

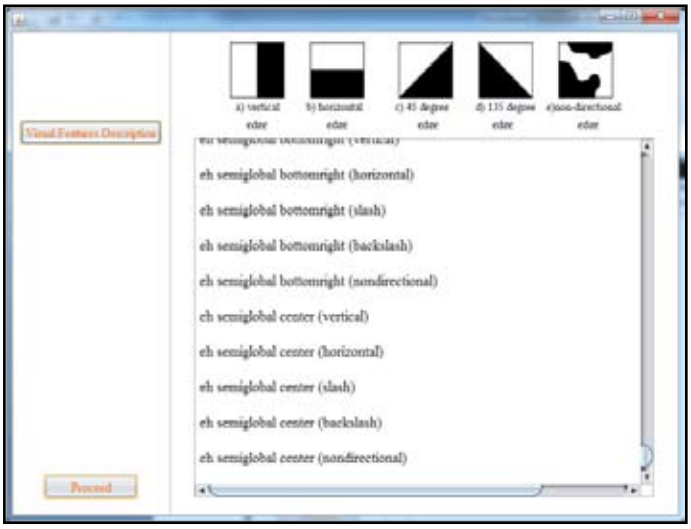

Fig. 5: Visual Feature Description

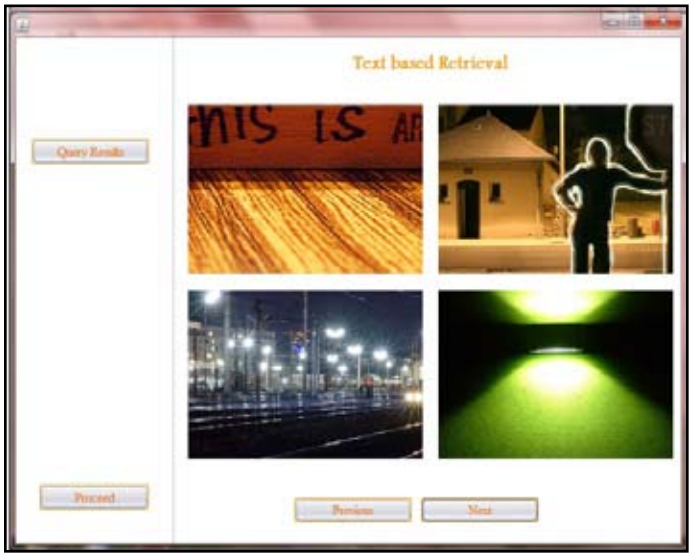

Fig. 6 :Text Based Image Retrieval

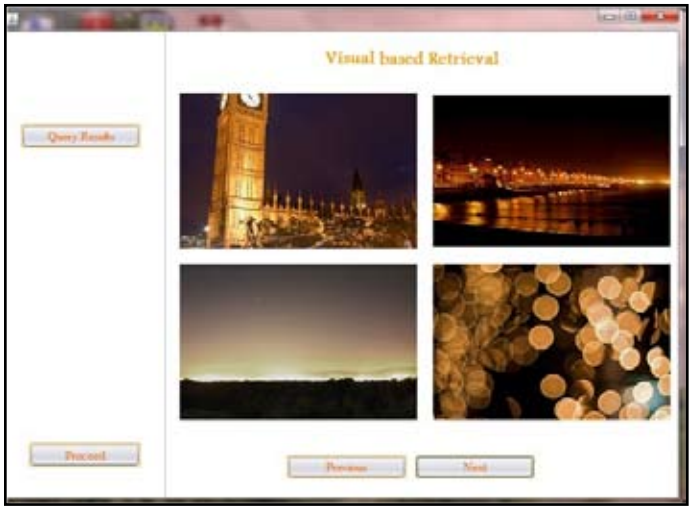

Fig. 7: Visual Based Image Retrieval

Integration of textual and visual features results in better search performance. Text-guided weighting scheme for visual features helps to improve accuracy. Weighting scheme infers user intention from query terms. In contrast to the other compared approaches, our proposed approach can be viewed as the best solution to precision.

V. Conclusion

To deal with the excessive computation for high dimensional data problem of CBIR and is not easy to accurately identify problem of TBIR, it have presented new approach by integrating Text and Visual Features. In summary, the main feature of integrating Text and Visual feature is to improves precision optimize the retrieval quality of better search performance interactive CBIR. The proposed algorithm integrate textual and visual feature represented text terms in the visual feature space, and developed a text-guided weighting scheme for visual features. Such weighting scheme infers user intention from query terms, and enhances the visual features that are significant toward such intention.

As a result, traditional problems such as precision and recall are solved. The effective results reveal that the proposed approach integrates Textual and Visual Feature is very effective in terms of precision and coverage. Moreover, the new search algorithm integrate Textual and Visual Feature search can bring out more

accurate results than other well- known approaches.

References

[1] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.
[2] J. Cui, F. Wen, and X. Tang, "Intentsearch: Interactive On-Line Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia,2008.
[3] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, "Intentsearch: Capturing User Intention for One-Click Internet Image Search," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 34, no. 7,pp. 1342-1353, July 2012.
[4] A.W.M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis Machine Intelligence,vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
[5] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," ACM Trans. Multimedia Computing, Comm., and Applications, vol. 2, no. 1, pp. 1-19, 2006.
[6] R. Datta, D. Joshi, J. Li, James, and Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 39, article 5, 2006.
[7] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Systems Man and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.
[8] G. Carneiro, A.B. Chan, P.J. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 29, no. 3, pp. 394-410, Mar. 2007.
[9] J. Li and J.Z. Wang, "Real-Time Computerized Annotation of Pictures," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 30, no. 6, pp. 985-1002, June 2008.
[10] A. Vailaya, A. Member, M.A.T. Figueiredo, A.K. Jain, H.-J. Zhang, and S. Member, "Image Classification for Content-Based Indexing," IEEE Trans. Image Processing, vol. 10, no. 1, pp. 117-130, Jan.2001.

Author's Profile and Image

Najimun Nisha .S received the BE degree from the Department of Computer Science and Engineerig at S.Veerasamy Chettiar College of Engg & Tech, Puliangudi. Now doing ME from the Dept.of Computer Science Engineering at S.Veerasamy Chettiar College Of Engg and Tech, Puliangudi.

Mehar Ban.K.A received B.E. and M.E. degree from the department of computer science and engineering at MK University, Madurai and Anna University, Chennai, Tamilnadu, India, respectively. She is currently working as a Assistant Professor in S.Veerasamy Chettair College of Engineering and Technology, Puliangudi, Tirunelveli Dist,Tamilnadu. Her research interests include data mining and Image processing. She is member of CSI, India.