

# A System for Retrieval Mobile Document Images Efficiently from Digital Library

Omesh Kalambe, Shubhangi Giripunje

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, G. H. Raisoni College of Engineering Nagpur (MS), India  
[omeshkalambe33@gmail.com](mailto:omeshkalambe33@gmail.com)

<sup>2</sup> Professor, Department of Electronics Engineering, G. H. Raisoni College of Engineering Nagpur (MS), India  
[shubhangi.giripunje@raisoni.net](mailto:shubhangi.giripunje@raisoni.net)

**Abstract**— Recently mobile devices are used to improve the user experience of digital library browsing and search. Various applications such as education, product retrieval and location search. Searching the document and transferring the query are important issues in mobile document search. In this paper, provides a model for automatically generating a caption for images with a mobile document image retrieval framework. The proposed system, consist of three stages: Content Selection, Surface realization and document retrieval. Here author considering both extractive and abstractive model and discuss. A Hamming distance KD-Tree is used for scalable searching. A JBIG2 based query image compression to fulfill the low bit rate query.

**Keywords**— Caption generation, image annotation, Digital library, Hamming space, JBIG2 compression, K-D tree, mobile visual search

## INTRODUCTION

Recent years the amount of digital information available on the Internet is growing. This digital information consists of scanned documents stored in the digital image format. Many search engines retrieve images, simply by matching user queries against collocated textual information. A query is formulated as a photo that captures the visual objects of user interest, for example, a book cover, a document page, a picture. The query is sent to the server end, where the similar documents are matched and returned.

To improve the matching efficiency, develop methods that generate description of words for an image automatically. An image can be annotated to generate a list of keywords such as planes, bombs, airport, where as the caption “planes carrying bombs are landing at the airport” would make the relationship between the words. The description is sufficient to understand the image, after that retrieve the document related to the image from digital library. A system that generates these descriptions automatically can improve the image retrieval.

### 1. Automatic Caption Generation

Although image understanding is a popular topic within computer vision, relatively little work has focused on caption generation. A handful of approaches create image descriptions automatically following two stage architecture. The picture is first analyzed using image processing techniques into an abstract representation, which is then rendered into a natural language description with a text generation engine. Much work within computer vision has focused on image annotation, [3] a task related to but distinct from image description generation. The goal is to automatically label an image with keywords relating to its content without however attempting to arrange these into a meaningful sentence or text.

Keyword based indexing techniques are popular and the method of choice for image retrieval engines. Furthermore, image descriptions tend to be concise, focusing on the most important depicted objects or events. A method that generates such descriptions automatically could therefore improve image retrieval by supporting longer and more targeted queries, by functioning as a short summary of the image’s content, and by enabling the use of question-answer interfaces. The image description generation adopts a two stage; consist of content selection and surface realization. The former stage analyzes the content of the image and identifies “what to say”, whereas the surface realization determines “how to say it”.

In this paper, researches had focus on various problems related to captions generation for mobile images, normally on web every single image co-occur with some related document. A Captioned images embedded with article and learn both techniques of content selection as well as surface realization. Here the extractive and abstractive models for generating meaningful, short and precise captions for the image are used. The benefit of this model, it does not require manual annotation of images. The probabilistic image annotation model suggests keywords for an image.

## 2. Document Retrieval

The problem of document image retrieval has been widely studied in 1990s, due to a wide variety of applications in digital library. The main target is to find the similar documents by querying a scanned document image over a large document corpus. Previous works typically depend on Optical Character Recognition (OCR) techniques [22]. More recently, visual matching is becoming a promising alternative to solve the limitation of poor OCR performance in scanned documents retrieval.

However, none of existing techniques is dedicated to sort out the challenging issues in the emerging task of mobile document image retrieval. An image query from a mobile camera is negatively affected by photograph distortion of the embedded camera on a mobile device is different from the scanned image, such as rotations, lighting changes, and de-focusing. Such kinds of distortion have posed a big challenge in OCR techniques. Existing visual descriptors in terms of words and paragraph layouts, they doesn't describe the textual regions in document images properly and also query delivery latency in mobile search scenarios.

The distortion of mobile captured images would seriously degenerate the performance of visual search. So a robust visual descriptor is required to characterize these line drawing regions. To retrieve document images, the relative layout of the shape, rather than the absolute scale and position, plays an important role. In addition to line drawings, a significant proportion of document images are textual regions, which, in many cases, presents visually similar or even identical paragraph layouts and fonts. When wireless network connections are subject to bandwidth limit, a mobile query of small size is preferred.

The Hamming Distance KD-Tree to seek the tradeoff between document retrieval performance and memory complexity of indexing structure in searching textual regions. JBIG2 based query image compression, to reduce the query delivery latency in search scenario.



Figure1: Left: Document image capture by mobile device.  
Right: Document retrieve from digital library [2]

## 3. Digital library

The digital library is a collection and scanned documents stored in the digital format, which enables users to browse and search locally as well as remotely. The World Digital Library (WDL) project has collected huge amounts of books from handwritten, maps, printed documents, newspapers etc.

### RELATED WORK

An image comprehension is considered as most popular factor in computer vision. The image description should generate from the images. To generate description from the images two steps must be followed. The first one is to analyze the image with the help of image processing techniques and extracts foremost factors from the images by means of some extraction methods which is then transcribe into natural language text by taking into account text generation engine.

In addition to above mentioned technique, image description can be generating by other fashion also. For example, Hiroshi Miki, Atsuhiko Kojima, and Koichi Kise [4], evaluate various usages of objects from the images and also identify different functions of an object, and in turn classify them accordingly by means of probability networks. This paper represents the method of creation of model of object recognition that can be done by examine relationship between human actions and function of object.

P. He'de and all [5], author has described usually to represent images of objects in some natural language or in a human readable form image annotation system is utilize in image base management. Automatic image annotation is a way to assigns metadata in the form of captioning or keywords to a digital image. Their system describe that manual database creations, they required color and texture to generate a caption in a natural language. In order to generate a description of Images step wise procedure must followed. In first step, perform segmentation technique on images with respect to available objects in the image. In second step, retrieve the attributes from the database, and finally generate a description for image using templates.

B. Yao and all in [6], discuss the methods of image parsing and text generation. It generates image text description in detail using parsing technique. Parsing technique shows correspondence between different sharing visual patterns within an image and partition the image into various parts namely scenes-object-parts. Specifically inputted images get decompose by means of image parsing engine. At the end, the task of text description is to generate meaningful and human understandable text. Paper also consist of image parsing engine that parse the input image into various parse graph and an And-or Graph that shows syntactic and semantic relation between visual features of images scenes, objects, parts. And-or Graph not only parse the image into top down approach but also provides mechanism that transfigure parse graph into semantic representation. Semantic web furnish interconnection between various semantic elements that are capture from previous method.

Ali Farhadi and all [7], demonstrate the formation of sentences from the required images by simply compare given images and sentences of documents and obtained a score in the paper related to every picture tells a story. This score is used to solely attach given sentence to the images. This method simply retrieves a given sentence rather than composing a new one. Here also an images split into three parts scenes, objects, and action applied over the scenes and objects. It contains two main factors Illustration and annotation. From the collection of pictures find picture suggested by keywords by an illustration method. V. Ordonez and all [8], Advances in Neural Information Processing Systems, describe the caption generation by using large collection of images based on word based model by taking into account about 1 millions photographs which is really a huge and enormous database. But in word based model some drawbacks arises. As the image annotation model does not take function words into account. Image annotation mechanism also auspiciously used relevance models, mainly implemented for information retrieval. In image annotation model keywords are used to convert image into human readable form. Relevance models, originally developed for information retrieval, have been also successfully used for image annotation (e.g., [9], [10]). A key idea behind these models is to find the images most similar to the test image and then use their shared keywords for annotation,

Simone Marinai and all [11], deal with a general framework for document image retrieval has been proposed. The system allows users to retrieve documents on the basis of both global features of the page and features based on blocks extracted by layout analysis tools. Global features include texture orientation, gray level difference histogram, and color features. The block-based features use a weighted area overlap measure between segmented regions and Authors describe the integration of the word and layout indexing and retrieval in a unique framework that can be used in Digital Library (DL) applications. We first review most common paradigms exploited by Internet DLs for document retrieval. Hong Liu and all [12] deal with document image retrieval. It retrieves the document image from the database, when they find most similar document images. It makes use of density distribution features and key block features, it is weak for those images which are very similar to each other in distribution and block features. The technique of density distribution of document image retrieval based on key block features of document images, for improving the retrieval performance Key block features are applied to confirm the reliability of the raw candidate images.

Joost van Beusekom, Daniel Keysers, Faisal Shafait, Thomas M. Breuel in [13], they use the layout information for document image retrieval. This method performs in two steps. In first step, the distance for every pair of blocks from the two layouts are computed using a block distance and in second step a matching is done to minimize the total distance between the two layouts and thus assign blocks to each other.

Shijian Lu in their paper [14] discusses the document image retrieval technique. The query keyword or query document image techniques are used. This paper presents a new word image annotation technique. The image can be annotated by using a set of topological shape features. The document images can be retrieved by using query keywords. There are various advantages of this technique; the first one is much faster because it does not require the time-consuming. The second is character shape features in use are more tolerant to the document skew and the variations in text fonts and text styles and finally most importantly, its collision rate is much lower.

Shijian Lu and all in [15] discuss a novel document image retrieval algorithm based on local feature sequence and common substring matching. The local feature sequences can be extracted without print-core detection. In this method can locate the match parts words by words. The discuss method achieve good performances on some document image databases. However, the local features such as character, character shape and word shape are depend on the quality of images.

The problem of logo recognition is very important in document image analysis because it enables immediate identification of the source of documents based on the originating organization [16]. In this method, segmentation free and layout independent and author address logo retrieval in an unconstrained setting of 2-D feature point matching. At the end, quantitatively evaluate the effectiveness of technique using large collections of real-world complex document images.

Tomohiro Nakai in their paper [17] described a real-time retrieval method for document images. In this technique the query consist of image of document captured by a camera. In real time document image for the query image is retrieved from the document image database. Since an object is linked with its relevant information in the database. From a captured image of the object, the relevant information is retrieved and provided to users. When a digital camera is used as an input device, object recognition of camera captured images is a difficult problem.

RezaAzmi, Hossain Akbari, Mohammad Akbari, Hossain Shirazi in [18], a system for retrieving document images from digital library based on visual similarity is described. The goal of document image retrieval is finding, all documents related to a query document, based on their visual characteristics. The method first creates a graph layout based on document extracted blocks. This graph represents document layout structure and second they create a vector based on path patterns to represent a graph. In this method the query document directly retrieve from the database without comprised in database files.

Kazutaka Takeda in paper [19] retrieves a real time document image from a 10 million pages database. They use the technique of Locally Likely Arrangement Hashing (LLAH), but this technique has some drawback. They required a large number of memories and another is the retrieval accuracy is going to decreases when a database is increases.

## PROPOSED METHOD

The proposed method is divided in two stages, one is client side and other is server side. The client can take single image by using the camera or select the image from gallery as input. The caption can be generated for the image by using the article stored at the server side, by selected keyword. The mobile query is send to the server side, as a wireless connection the mobile query must be small in size is preferred. JBIG2 compression technique is used for compress the image at the client side. The same technique is used for the decompressed the image

at the server side. Text segmentation is used for segment the text. Hamming distance KD-tree is used for the searching the document. Finally the rank fusion algorithm is applied for retrieving the document. The top rank document is shown at the client side.

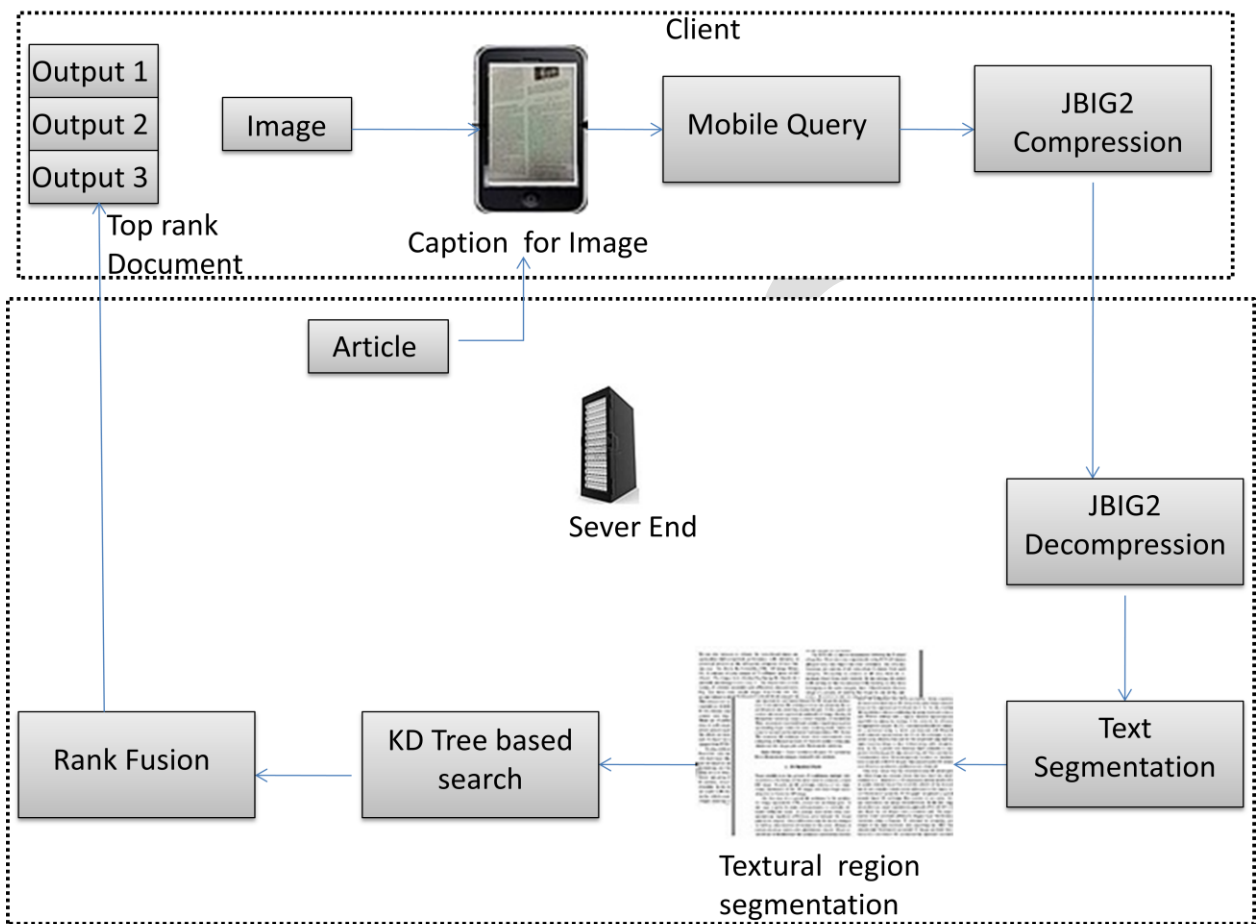


Figure 2: The architecture of project

The proposed method is divided into three tasks: Content selection identifies what the image, whereas surface realization determines how to verbalize the chosen content and third stage is efficient retrieval of the document. Before describing our model in detail, we summarize our assumptions regarding the caption generation task and the nature of the data on which it is being modeled. The caption describes the content of the image directly or indirectly. In traditional image annotation, keywords describe salient objects; captions supply more detailed information, not only about objects and their attributes, but also events. The accompanying document describes the content of the image. Since our images are implicitly rather than explicitly labeled, so do not assume that researchers can enumerate all objects present in the image. Instead, hope to model event-related information such as “what happened,” “who did it,” and “where” with the help of the document.

### A. Image Content Selection

Researchers define probabilistic image annotation model based on the assumption that images and their surrounding text are generated by a shared set of latent variables or topics. Specifically, we describe documents and images by a common multimodal vocabulary consisting of textual words and visual terms. Due to polysemy and synonymy, many words in this vocabulary will refer to the same underlying concept. Using Latent Dirichlet Allocation (LDA [20]), a probabilistic model of text generation, we represent visual and textual meaning jointly as a probability distribution over a set of topics. Our annotation model takes these topic distributions into account while finding the most likely keywords for an image and its associated document

Words and images are distinct modalities, but both modalities are on same level as they describe same objects. The first step is the segmentation of the picture into regions, using image segmentation algorithm. Regions are then described by a standard set of features, including color, texture and shape. The visual features receive a discrete representation and each image is treated as a bag of words. To achieve this, Scale Invariant Feature Transform (SIFT) algorithm is used [21], [22].

They represent any image-caption-document tuple together as a mixed document. Here assume that the two modalities express the same meaning. The LDA is a three- level Bayesian model. It consists of the various documents. The document is model by mixture with topic. The words for the document are generated by using the topic distribution.

## B. Image Annotation

Image annotation is the application of the computer vision. Image annotation is the method to assign a keywords or description of words to the image. It is used a machine learning technique to automatically apply the keyword for the image. First task is to learn the image features training annotation of image. After that the technique can be constructed by using the machine translation. Previously the document image can be retrieve by using the content based retrieval system. The advantage over content based retrieval system is less time required to search the document and also it is less expensive.

Researchers use the Latent Dirichlet Allocation that generates a text for the image. It represents visual and textual meaning together as a probability distribution over a set of topics. It uses the topic distributions for detecting the most likely keywords for an image and their related document. The mix LDA is compared with the other types of LDA. the Mix LDA significantly ( $p < 0:01$ ) is preferable than other models.

## C. Extractive Caption Generation

The idea behind Extractive caption generation is to create a summary simply by identifying and subsequently concatenating the most important sentences in a document, independently of text type, style and subject matter. For the task of caption generation, only the extraction of a single sentence is required. This sentence must be maximally similar to the description keywords generated by the annotation model [23], [24]. There are different ways to represent extractive caption generation

### Word Overlap-Based:

In this sentence selection method measure a similarity of the image keyword and document sentence. Where  $W_I$  is the set of keywords and  $S_d$  a sentence in the document.

$$\text{Overlap}(W_I, S_d) = \frac{W_I \cap S_d}{W_I \cup S_d}$$

### Vector Space-Based:

The disadvantage of word overlap method is removing by representing a keyword and sentences in vectors space. The vectors represented by keywords  $\vec{W}_I$  and document sentence  $\vec{S}_d$

$$\text{sim}(\vec{W}_I, \vec{S}_d) = \frac{\vec{W}_I \cdot \vec{S}_d}{\|\vec{W}_I\| \|\vec{S}_d\|}$$

### Topic-Based:

In this method, the similarity can be measured by a probability distribution over a set of latent topic between the image and sentences by using the same topic distributions

## D. Abstractive Caption Generation

There is often no single sentence in the document that uniquely describes the image's content. In most cases the keywords are found in the document but interspersed across multiple sentences. The selected sentences make for long captions, which are not concise. For these reasons, we turn to abstractive caption generation technique.

### Word-based Caption Generation:

To content selection is modeled as the probability of a word appearing in the headline given that the same word appears in the corresponding document and is independent of other words in the headline. The likelihood of different surface realizations is estimated using a bigram model. Since the individual words cannot frame a meaningful caption.

### Phrase-based Caption Generation:

In Word-based Caption Generation, there is no guarantee that these will be compatible with their surrounding context or that the captions will be globally coherent beyond the trigram horizon. To avoid these, shift our attention to phrases. The phrases are linked together to create headlines using a set of handwritten rules,

## E. Scalable Search:

To accomplish effective and efficient retrieval, we resort to the K-D Tree based approximate nearest neighborhood search, with a Hamming embedding scheme [25] to reduce memory cost. The motivation is to introduce a compact binary code to reduce the memory cost from storing original local descriptors for backtracking. In the proposed Hamming Distance (HD) KD-Tree, we replace the Euclidean feature space with a Hamming space. The Hamming distance KD-Tree enables very fast similarity matching, while maintaining matching accuracy,

## F. Low Bit Rate Query

Here discuss, the query compression scheme towards low bit rate mobile visual search in the field of document images. Rather than extracting compact descriptors as in [26], [27], [28], we propose to compress query images in JBIG2 at low complexity.

JBIG2 first tries to segment a document image into three types of regions, namely, text regions, halftone regions, and regular regions. For text regions, the symbol compression is used. For halftone regions, the grid coding is applied and the arithmetic coding is applied to regular regions.

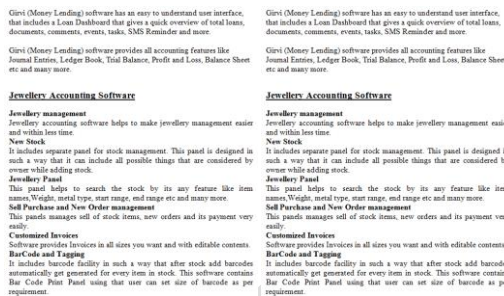


Figure 3: Left: Original Image of size 120 kb Right: compressed Image of size 24 kb

## EXPERIMENTAL RESULT

At the client side, single image can be taking as input. Client showing the original as well as compressed image is shown in a diagram. The original image of size 362 kb is compressed to 12 kb for low bit query rate. The article is shown in the below, is stored at the server side. The article is used for generating a caption for the image. Image annotation model is used to extract the keyword from the article. The phase based method, generate a caption for the image (Shown in figure).

The Hamming distance KD tree is used to search the document in the library. Server side contains a large number of documents. When the client searches for the document they can be retrieving on the basis of rank assign to the document. The document retrieve at the client side is shown in the figure.

### Database:

The article for images that used to generate a caption is stored in database. The database contains large number of document stored in a txt file format. For example the butterfly image shown in the client side, the article used for generating a caption is as follows.

### Article:

Barcode-An optical machine-readable code, A barcode is an optical machine-readable consist of data relating to the object to which it is attached. It consists of multiple numbers of parallel lines. It can also available in into rectangles, dots, hexagons and other geometric patterns. They are generally referred to as barcodes as well. for reading the Barcodes, it is scanned by special optical scanners called barcode readers, by using the barcode reader recognize the product. In current year software became available on devices like Smartphone, computers to recognize the barcode. [30].

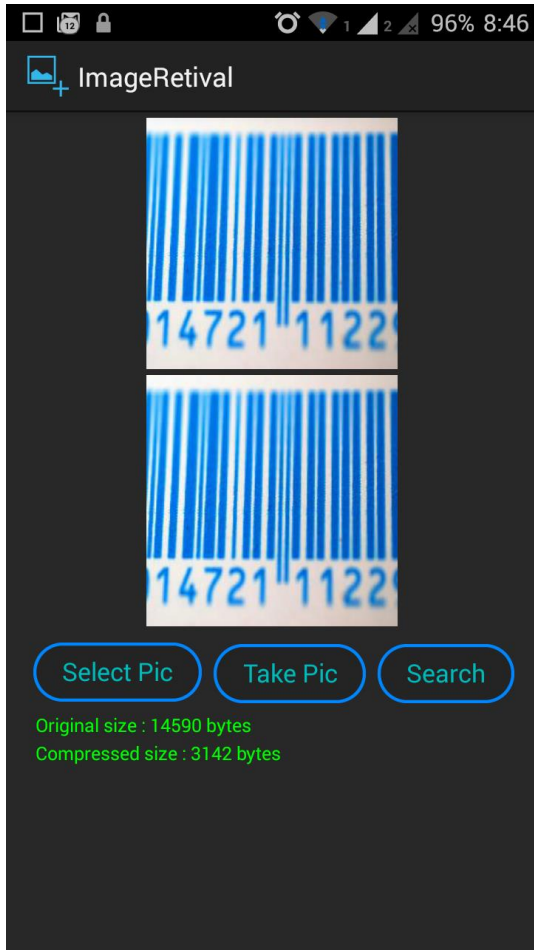


Figure 4: Top - Original image  
Bottom - Compressed image

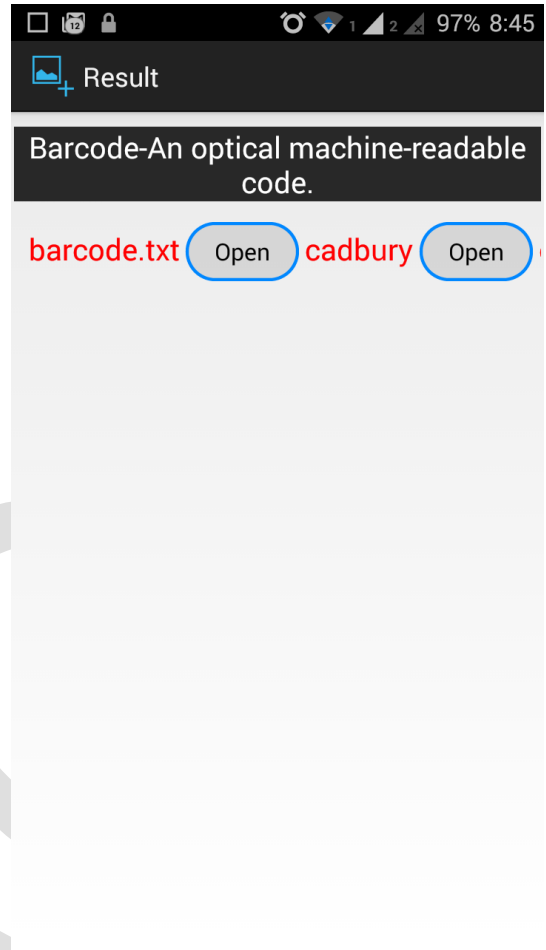


Figure 5: Caption for the Image\

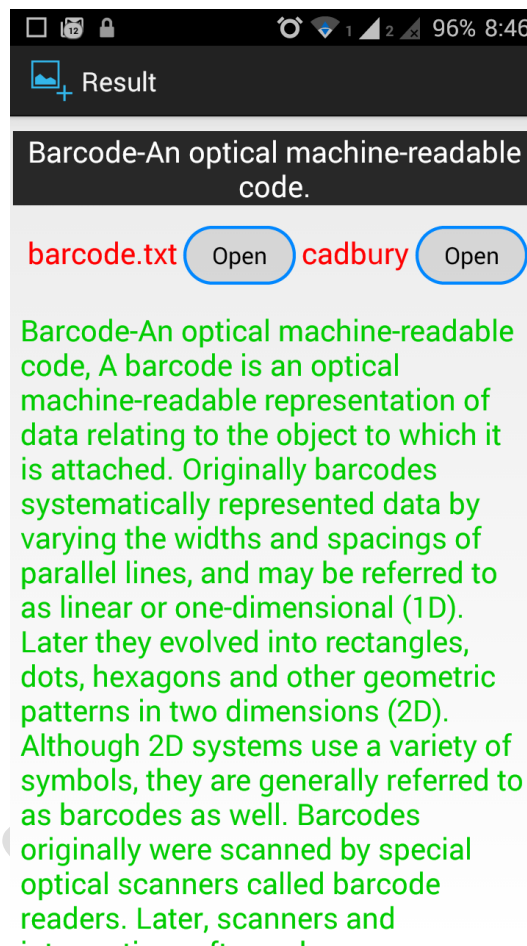


Figure 6: Document for Image.

## CONCLUSION

The proposed method generate caption for mobile images automatically and also retrieve the document. An efficient method of automatic text generation from the images and searching methods that search the images with respect to content of images, without giving more importance to surrounding text. So that gets more targeted images. Also shows relationship between content selection and surface realization to achieve caption generation using phrase based model for image annotation method.

The Hamming Distance KD-Tree is addressing the issue of high memory cost in building up the indexing structure towards scalable search. JBIG2 based compression scheme of low complexity is introduced to reduce the query delivery latency while maintaining comparable search accuracy.

## REFERENCES:

- Yansong Feng and Mirella Lapata, "Automatic Caption Generation for News Images" IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 4, april 2013
- Ling-Yu Duan, Rongrong Ji, Zhang Chen, Tiejun Huang and Wen Gao, Fellow, "Towards Mobile Document Image Retrieval for Digital Library" IEEE transactions on multimedia, vol. 16, no. 2, february 2014
- A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for Image Annotation," Int'l J. Computer Vision, vol. 90, no. 1, pp. 88-105, 2010.
- A. Kojima, M. Takaya, S. Aoki, T. Miyamoto, and K. Fukunaga, "Recognition and Textual Description of Human Activities by Mobile Robot," Proc. Third Int'l Conf. Innovative Computing Information and Control, pp. 53-56, 2008.
- P. He'de, P.A. Moe'lic, J. Bourgeois, M. Joint, and C. Thomas, "Automatic Generation of Natural Language Descriptions for Images," Proc. Recherche d'Information Assist'e par Ordinateur, 2004.
- B. Yao, X. Yang, L. Lin, M.W. Lee, and S. Chun Zhu, "I2T: Image Parsing to Text Description," Proc. IEEE, vol. 98, no. 8, pp. 1485- 1508, 2009.
- A. Farhadi, M. Hejrati, A. Sadeghi, P. Yong, C. Rashtchian, J. Hockenmaier, and D. Forsyth, "Every Picture Tells a Story: Generating Sentences from Images," Proc. 11th European Conf. Computer Vision, pp. 15-29, 2010.
- V. Ordonez, G. Kulkarni, and T.L. Berg, "Im2Text: Describing Images Using 1 Million Captioned Photographs," Advances in Neural Information Processing Systems, vol. 24, pp. 1143-1151, 2011.



- V. Lavrenko, R. Manmatha, and J. Jeon, "A Model for Learning the Semantics of Pictures," Proc. 16th Conf. Advances in Neural Information Processing Systems, 2003.
- S. Feng, V. Lavrenko, and R. Manmatha, "Multiple Bernoulli Relevance Models for Image and Video Annotation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1002-1009, 2004.
- Simone Marinai, Emanuele Marino, Francesca Cesarini, and Giovanni Soda "A General System for the Retrieval of Document Images from Digital Libraries" IEEE Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04) 2004.
- Hong Liu, Suoqian Feng, Hongbin Zha, Xueping Liu "Document Image Retrieval Based on Density Distribution Feature and Key Block Feature" IEEE Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05) 2005.
- Joost van Beusekom, Daniel Keysers, Faisal Shafait, Thomas M. Breuel, "Distance Measures for Layout-Based Document Image Retrieval" IEEE, Proceedings of the Second International Conference on Document Image Analysis for Libraries (DIAL'06) 2006.
- Shijian Lu, Member, IEEE, Linlin Li, and Chew Lim Tan, Senior Member, IEEE, "Document Image Retrieval through Word Shape Coding" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 30, NO. 11, NOVEMBER 2008.
- Jilin Li , Zhi - GangFan , Yadong Wuand Ning Le, "Document Image Retrieval with Local Feature Sequences" 10th International Conference on Document Analysis and Recognition, 2009.
- Guangyu Zhu and David Doermann, "Logo Matching for Document Image Retrieval" 10th International Conference on Document Analysis and Recognition 2009.
- Tomohiro Nakai , Koichi Kise and Masakazu Iwamura, "Real - Time Retrieval for Images of Documents in Various Languages using a Web Camera" 10th International Conference on Document Analysis and Recognition 2009.
- RezaAzmi, HossainAkbari, MohammadAkbari, Hossain Shirazi, "LGL-DIR: Layout Graph for Layout based Document Image Retrieval" 2nd International Confonence on Education Technolog and Computer (ICETC) 2010.
- Kazutaka Takeda, Koichi Kise and Masakazu Iwamura, "Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH" International Conference on Document Analysis and Recognition 2011.
- D. Blei, "Probabilistic Models of Text and Images," PhD dissertation, Univ. of Massachusetts, Amherst, Sept. 2004.
- Lowe, "Object Recognition from Local ScaleInvariant Features," Proc. IEEE Int'l Conf. Computer Vision, pp. 1150-1157, 1999.
- Lowe, "Distinctive Image Features from ScaleInvariant Keypoints," Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- K. Sparck Jones, "Automatic Summarizing: Factors and Directions," Advances in Automatic Text Summarization, I. Mani and M.T. Maybury, eds., pp. 1-33, MIT Press, 1999.
- I. Mani, Automatic Summarization. John Benjamins Publishing Co., 2001.
- H. Jegou, M. Douze, and C. Schmid, "Hamming distance and weak geometric consistency for large scale image search," in Proc. ECCV, 2008, pp. 304-317.
- D. Chen, S. Tsai, V. Chandrasekhar, G. Takacs, J. Singh, and B. Girod, "Tree histogram coding for mobile image matching," in Proc. DCC, 2009, pp. 143-152.
- V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed histogram of gradients a low bit-rate feature descriptor," in Proc. CVPR, 2009, pp. 2504-2511.
- R. Ji, L.-Y. Duan, J. Chen, H. Yao, J. Yuan, Y. Rui, and W. Gao, "Location discriminative vocabulary coding for mobile landmark search," Int. J. Comput. Vision, vol. 96, no. 3, pp. 290-314, 2012.
- S. Mao and A. Rosenfeld, "Document structure analysis algorithms: A literature survey," in Proc. SPIE Electronic, 2003, pp. 197-207.

[www.wikipedia.com](http://www.wikipedia.com)