International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 11 – No. 3, August 2016 – www.ijais.org

An Efficient Approach towards Satellite Image Retrieval using Semantic Mining with Hashing

Linda Mary John
Research Scholar
Thakur College of Engineering and Technology
Mumbai, India

Kiran A. Bhandari
Associate Professor
Thakur College of Engineering and Technology
Mumbai, India

ABSTRACT
Satellite images have gained a wide popularity in the field of content-based image retrieval. A massive amount of these images are collected every year due to the high availability of satellites and computer technologies. However, extracting user-specific content from these images still remains a primary concern due to the presence of semantic gap. This limits the capabilities of CBIR. Therefore, an effective and efficient method is required for image retrieval. This paper puts forward a semantic-based image retrieval approach along with the advantages of hashing for better feature extraction and precise retrieval. Hashing accelerates the quality of similarity search among images by generating unique image-hash codes .This approach also aims to scale down the problems related to semantic gap for better retrieval results.

General Terms
Remote sensing, Satellites, Data sets, Content based image retrieval and Algorithms.

Keywords
Semantic mining, Hashing, Hash codes, Semantic gap, Feature extraction, Semantic-based image retrieval, Precision, Recall.

1. INTRODUCTION
The task for searching and retrieving relevant information from huge repositories of satellite images is a tedious process. Different approaches such as text-based retrieval and content-based retrieval systems were used previously for such retrieval purposes. However, these approaches lacked the ability to solve the problem of semantic gap .Therefore new approaches capable of capturing the semantic aspects of the images was a fundamental need .This led to the development of a new approach known as Semantic-based image retrieval (SBIR)System [1].

Semantic-based retrieval depends on the semantic meaning of the images. The main motive of this approach is the mapping of the low–level image features to the high-level human concept. It aims to extract the human cognitive concept in order to bridge the semantic gap. This approach presents images with semantic features for easy retrieval and efficient search [2].

The organization of paper is as follows: Section 2 shows the related works performed by various researchers in the field of image retrieval. Section 3 gives a detailed description on the methodology of the proposed system. Section 4 shows the Experimental results based on the proposed system. Section 5 shows the conclusion of the paper.

2. RELATED WORK
Nancy Goyal and Navdeep Singh (2014) in their work put forth a survey on various CBIR techniques based on high-level features. It listed three fundamental processes for semantic extraction i.e. 1) Based on knowledge 2) Based on manual intervention and 3) Based on external data. Different techniques for reducing semantic gap were also stated such as Object ontology, machine intelligence, relevance feedback and so on. The advantages and disadvantages of these techniques were also compared for better research purpose [3].

Nikita Upadhyaya and Manish Dixit (2016) presented a review depicting the importance between low-level features and high-level semantics in the field of CBIR .The research work explained the working of CBIR framework along with the categories of CBIR queries. A detailed explanation of various semantic retrieval techniques was also presented for better research study [4].

Ruksanamol T. A. and Femithamol A. M (2014) proposed an effective image retrieval approach using hash codes. The proposed method used hash codes for efficient image matching and retrieval. It presented two tasks: 1) Offline processing for generation of bitwise weight for hashcodes.2) Online processing to generate query adaptive weight for ranking of images. The future work included the use of filters for removal of noise for better retrieval quality [5].

Shijun Xiang et.al (2012) in their paper presented a novel approach based on block-based hashing algorithm. The paper proposed blocking-based strategy for NMF hashing. The experimental results depicted the robust nature of the hashing strategy. Different block-based image watermarking algorithms were considered for future study [6].

In 2004,Yadong Mu.et.al presented a novel hashing algorithm known as LAMP .This proposed method generated high-quality hash functions using kernel conjuring and weak supervision .The LAMP algorithm creates a random sampling plan for working set as well as carrying vectors, thus making it scalable for large datasets[7].

Dell Zhang.et.al (2004) presented a novel Self- Taught Hashing (STH) for semantic-based hashing. This method discovers the best l-bit binary codes for all documents using unsupervised learning, followed by training of l classifiers via supervised learning .This further calculates the l- bit code for any hidden document. However, this approach fails to enhance the speed and scalability of similarity search [8].

In the research paper titled “Semi-supervised hashing for scalable image retrieval” Jun Wang.et.al (2010) presents a semi-supervised learning method. This method aims to learn efficient hash codes capable of calculating semantic similarity or dissimilarity. The theoretical features of the SSH method is not studied in this research work [9].

Marie Liénou et.al (2010) in their work demonstrated annotation of satellite images with user-defined semantic concepts. The annotation task performed supervised classification on blocks of satellite images along with spatial information for better results. The learning of the data sets was based on LDA method and the experimental results tested on quickbird images provided great retrieval performance [10].

The research work given by P. Sumathi et.al (2016) deals with web mining based on semantic concepts. The mining process was enhanced using SVM's .The proposed method aimed to return results which reveals the semantic relationship among the images and text for an effective retrieval purpose [11].

Dr. M.V Siva Prasad et.al (2014) also demonstrated mining of user defined queries via Markov chains. It introduces a new approach known as Markovian Semantic Indexing which was applied on online image retrieval system similar to Google search. The aim of this research was to enhance user interaction by returning results that have greater probability to be selected by the user [12].

Vlado Kitanovski et.al (2007) presented a hybrid approach using hashing and watermarking concept for image validation. The proposed method generated an image hash which was later used as a key to create image watermark. This method enhanced the overall security and robustness of the retrieval system [13].

3. METHODOLOGY

The proposed methodology introduces a novel approach for an efficient image retrieval system. This novel approach is based on two approaches i.e. Semantic Mining and Hashing.

Semantic Mining is an important aspect used for reduction of semantic gap. It aids in mining and extracting high-level semantic features for effective retrieval .On the other hand Hashing converts distinct images vectors into hash codes which are later indexed into hash tables for quick image search and retrieval. Therefore the respective advantages of both these approaches serve the basis for our proposed method.

The proposed system is divided into two main phases:

1. Creation of Hash Database
2. Detection of Images.

3.1 Creation of Hash Database

In the first phase, the proposed method creates a database to store the dataset of trained satellite images .The images are trained via feature extraction and generation of hash codes. The training phase of the images takes place in the backend as follows:

3.1.1 Scale the image

In the first step, all the rgb images are converted to gray-scale images of identical pixel size of 256 * 256 for easy processing of the images.

3.1.2 Enhance the image

In the second step, all the resized images are enhanced using adaptive Histogram equalization. The Adaptive histogram equalization method returns images with improved contrast and enhanced edges for each area of the images.

3.1.3 Filter the image

The images obtained may be the dirty due to presence of some noise factor which ultimately leads to poor image processing capabilities. Therefore the enhanced images are again filtered using Gaussian filters. The Gaussian filters are used because it returns a weighted average of all the neighborhood pixels of each pixel, with the average weight inclined towards the value of the Centre-pixel.

3.1.4 Feature Extraction

In this step, images are semantically mined to extract high-level features for reducing the problem of semantic gap. The proposed methodology uses Scale-invariant feature transform (SIFT) method. This method uses a 4-stage process for extracting features:

3.1.4.1 Scale-Space Extrema Detection

This stage attempts to detect the scales and locations that can be identified from different views with respect to the same object. This can be achieved as follows:

$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$

Where * stands for convolution operator, $G(x, y, \sigma)$ stands for variable-scale Gaussian and $I(x, y)$ stands for the input image.

3.1.4.2 Keypoint Localistaion

This stage tries to eliminate those points with poor contrast or less defined edges from the list of key points.

3.1.4.3 Orientation Assignment

This stage assigns a consistent orientation to the above key points based on the local properties of the images, thus making it invariant to image location, image scaling and image rotation.

3.1.4.4 Keypoint Descriptor

In this step, a distinctive descriptor vector for each keypoint is calculated .This ensures partial invariance to other image variations.

3.1.5 Generate Hash codes

In this step, we determine the Eigen vectors from the feature matrix and a hash value is obtained from the mean of both vectors. A final column vector of the hash-values are formed which is then concatenated with the name of the image to form a unique hash code for each images. The hash codes are then stored into the database for the purpose of similarity matching with query image.

3.2 Detection of Images

In the second phase, images are given as query inputs to the system. These test images are then trained again using the steps given in section 3.1 in order to find out its features and hash codes. The system then compares the features of the test images with the trained images stored in the database and retrieves only those similar images with correlation coefficient greater than 0.85.The best match will be the image with highest correlation coefficient.

3.3 Analysis of Images

The performance of an**y** image retrieval system can be measured with respect to its retrieval speed and efficiency. Our method analyzes two factors i.e. precision and recall in determining the system performance.

4. EXPERIMENTAL RESULTS

4.1 Dataset Description

The dataset consist of 200 satellite images of high resolution. These satellite images belong to mainly three categories i.e. coastal, desert and forest images. All the images are trained for its features and hash codes and then later stored into hash database for retrieval purposes.

4.2 Results

The screenshots of the proposed methodology is as follows. It includes the output of the two phases and its relevant theory mentioned above:

4.2.1 Creation of Hash Database

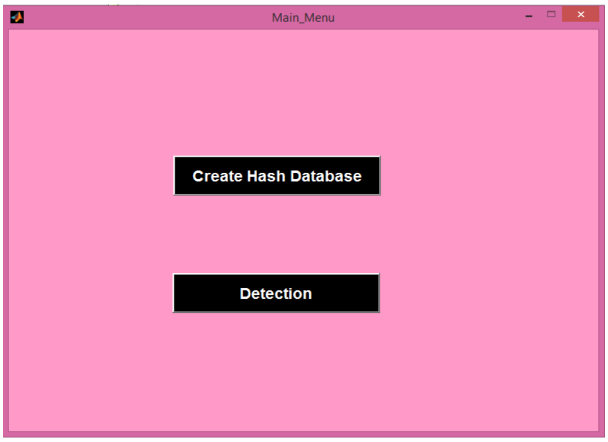

Fig 1: Main GUI of the proposed system

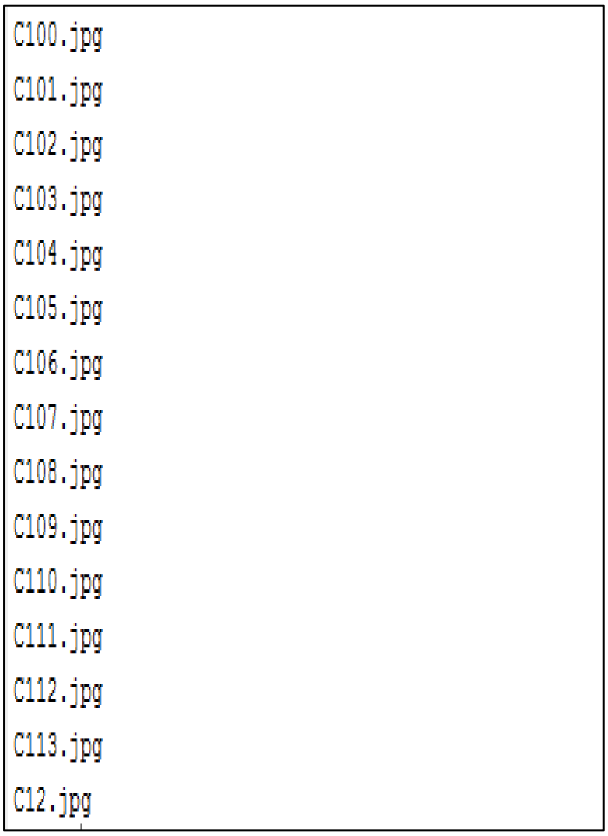

Fig 2: Training of images

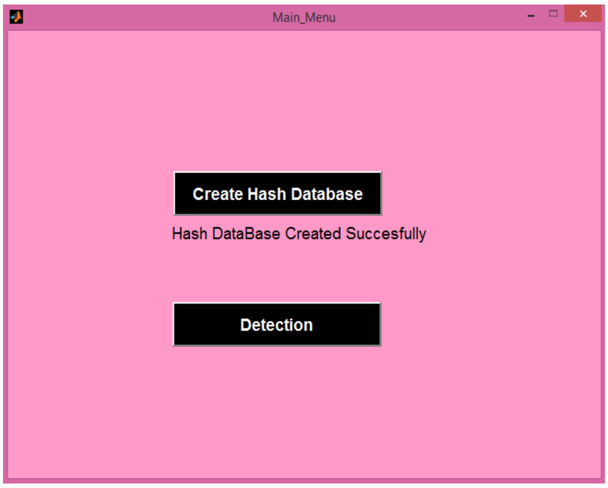

Fig 3: Creation of Hash Database

4.2.2 Detection of Images

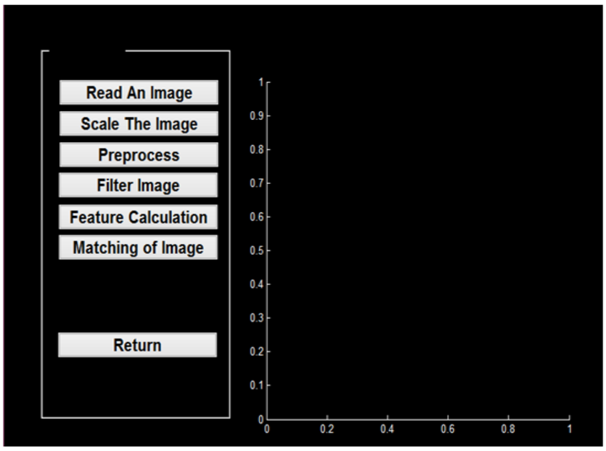

Fig 4: GUI for detection of Images

Fig 5: Reading a Test Image

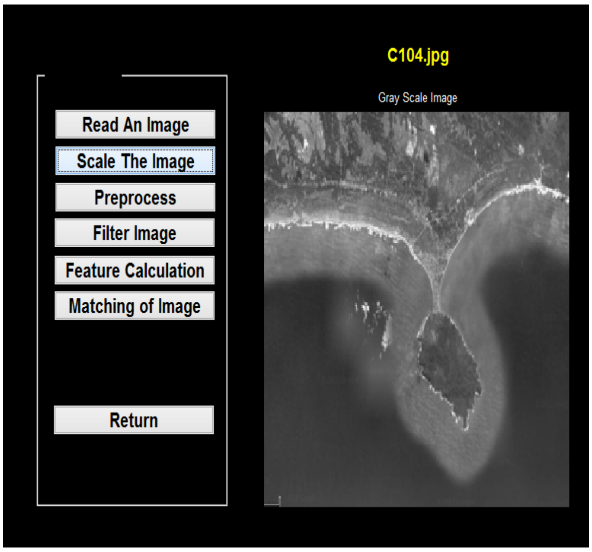

Fig 6: Scaling a Test Image

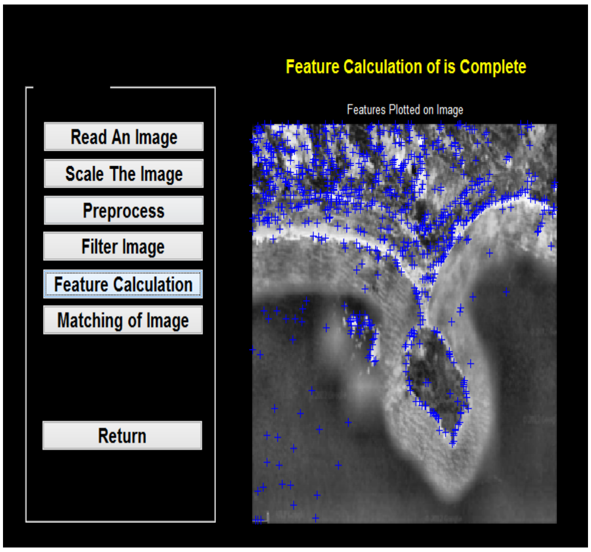

Fig 9: Feature Calculation a Test Image

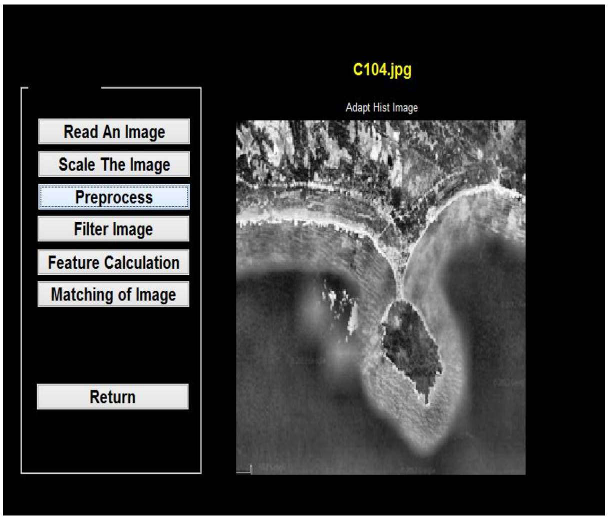

Fig 7: Enhancing a Test Image

Fig 10: Retrieved Image 1

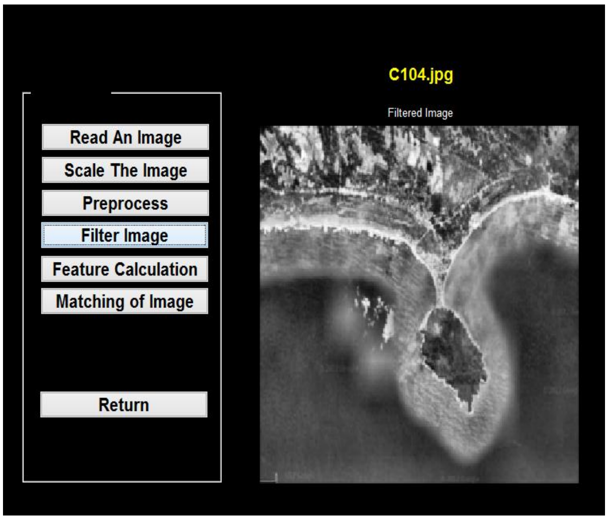

Fig 8: Filtering a Test Image

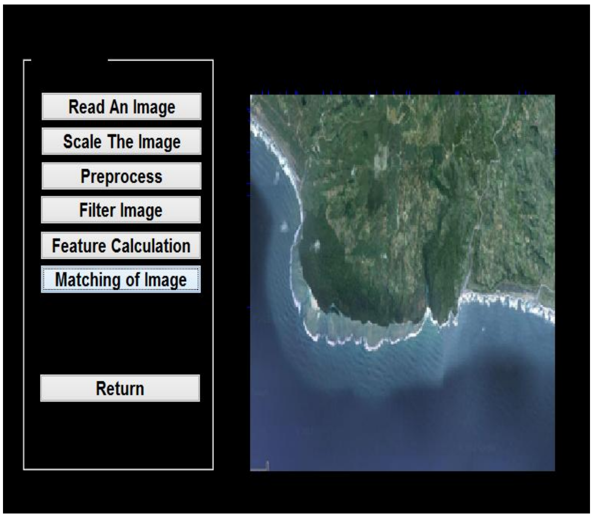

Fig 11: Retrieved Image 2

Fig 12: Retrieved Image 3

Table 1.Correlation Coefficient of retrieved images

RETRIEVED IMAGES	CORRELATION COEFFICIENT
Retrieved Image 1	0.86648
Retrieved Image 2	0.91437
Retrieved Image 3	1

5. CONCLUSION

Therefore, the image retrieval system using hashing and Semantic mining is an effective method which can be used on a great extent. The proposed system achieves precise retrieval results with relevant information and minimum semantic gap. The image with the highest correlation coefficient is retrieved as the best match. Moreover since the proposed method is completely software-based, the low cost of the method serves as an added advantage.

6. REFERENCES

[1] Mohammed Alkhawlani, Mohammed Elmogy, Hazem El Bakry: "Text-based, Content-based, and Semantic-based Image Retrievals: A Survey", International Journal of Computer and Information Technology Volume 04 – Issue 01, January 2015 .

[2] Hui Hui Wang, Dzulkifli Mohamad, N.A Ismail:" Image Retrieval: Techniques, Challenge, and Trend", International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:3, No:12, 2009.

[3] Nancy Goyal , Navdeep Singh:" A Review on Different Content Based Image Retrieval Techniques Using High Level Semantic Feature", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2014

[4] Nikita Upadhyaya and Manish Dixit:" A Review: Relating Low Level Features to High Level Semantics in CBIR ",International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.3 (2016).

[5] Ruksanamol T.A. and Femithamol A. M: "Class Based Image Search with Hash Codes", International Journal of Innovation Research & Development Vol 3 Issue 9, September 2014.

[6] Shijun Xiang and Jianquan Yang:" Block-based image hashing with restricted blocking strategy for rotational robustnes", EURASIP Journal on Advances in Signal Processing ,2012.

[7] Y. Mu, J. Shen, and S. Yan:" Weakly-supervised hashing in kernel space", in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010, pp. 3344–3351.

[8] D. Zhang, J. Wang, D. Cai, and J. Lu,"Self-taught hashing for fast similarity search," in Proc. Conf. Advances in Neural Information Processing Systems, 2004, pp. 1–8.

[9] J. Wang, S. Kumar, and S. Chang, "Semi-supervised hashing for scalable image retrieval," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010, pp. 3424– 3431

[10] Marie Liénou, Henri Maître and Mihai Datcu, "Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation," in IEEE Geoscience And Remote Sensing Letters, Vol. 7, No. 1, January 2010.

[11] P. Sumathi , R. Manickachezian:" Semantic-Based Web Mining For Image Retrieval Using Enhanced Support Vector Machine",International Journal of Applied Engineering Research ,Volume 11, Number 5 (2016) pp 3276-3281.

[12] Dr. M.V Siva Prasad, P. Sandeep Reddy, K. Krishna Reddy, "Mining User Queries with Markov Chain: Application to Content Based Image Retrieval System," International Journal of P2P Network Trends and Technology (IJPTT) – Volume 8 – May 2014.

[13] Vlado Kitanovski ,Dimitar Taskovski, Sofija Bogdanova, "Combined Hashing/Watermarking Method for Image Authentication", International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:1, No:6, 2007.