

Retrieval of Geographic Images Using Multi Support Vector Machine and Local Invariant Features

S.Radha¹, A.S. Rehna²P.G. Student, Dept of Electronics and Communication Engineering, MACE, Trivandrum, Kerala, India¹P.G. Student, Dept of Electronics and Communication Engineering, MACE, Trivandrum, Kerala, India²

ABSTRACT: This paper deals with retrieval of geographic images. For the retrieval of geographic images it uses local invariant features. SIFT algorithm is used. Multi SVM algorithm is used to perform image retrieval and using this algorithm the false image retrieval is considerably reduced. An SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. Local features allows a greater range of objects and spatial patterns to be observed. SIFT detector is used to identify the key-points in the image. Feature extraction is used to extract features from the image. Using the proposed algorithm the image retrieval is shown to give accurate results. It not only performs image retrieval but also detection and classification of geographic images.

KEYWORDS: Multi SVM, SIFT, land use, land cover, remote sensing.

I.INTRODUCTION

Nowadays, we observe a huge amount of images stored in electronic format particularly in the case of biological and medical applications. Therefore, efficient image retrieval technique become a fundamental requirement for searching and retrieving images from a large digital image database. An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Manual image retrieval is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image retrieval. This paper investigates one such class of methods wherein local image regions are characterized by features designed to be invariant to differences in appearance resulting from geometric transformations such as rotation or scaling as well as from photometric transformations such as changes in illumination. The image regions themselves are also detected in an invariant manner. These so-called local invariant features have been successfully applied to a range of standard (non-geographic) computer vision problems, and there has been increasing interest in using them for overhead image analysis.[1]The fundamental contribution of this paper is an investigation into local invariant features for overhead image retrieval.

A. Background

Detecting buildings from very high resolution (VHR) aerial and satellite images is extremely useful in map making, urban planning, and land use analysis. Although it is possible to manually locate buildings from these VHR images, this operation may not be robust and fast. Therefore, automated systems to detect buildings from VHR aerial and satellite images are needed.[3]. We compare the scene classification performance of 13 features, including structure, texture and colour features. First, image classification are performed using a single feature and the performance of different features are compared. Both the k-nearest-neighbour (knn) classifier and the support vector machine classifier (SVM) are employed.[4].

In this paper, we present a novel indexing structure that was developed to efficiently and accurately perform content-based shape retrieval of objects from a large-scale satellite imagery database.[5]. The method for image classification is based on representational power of graphs with the efficiency of the bag-of-words model. A graph is constructed from local patches of interest regions and their spatial arrangements. Each graph is represented with a histogram of sub-graphs selected using a frequent sub-graph mining algorithm[10].A method which transforms the scene content and the associated spatial information of that scene into graph data. A new image content representation is described using frequent sub-graph histograms for classifying complex scenes such as dense and sparse urban areas. A method for constructing an image graph which encapsulates the spatial information of the scene. Recurring spatial structures of a scene class are encoded in a histogram of frequent sub-graphs. The method of image classification using sub-graph histogram model improves the performance of the bag-of-words model using the spatial information encoded in the sub-graph histogram representation. The bag-of-words representation based on appearance features is used in image and text classification. But its drawback is that shape patterns of the image are neglected. A method for image classification is by using bag-of-words representation of textons considering spatial information. Textons were first introduced by Julesz. It is defined as the atomic visual elements of a visual scene. In the method of classifying remote sensing images using mixture distributions, the image is represented as a number of individual patches[11].Probabilistic framework is used for classifying highly textured images. The probabilistic generative model of data is defined using the following assumptions: the data is described by appearance and shape/pattern characteristics; pattern and appearance features are independently sampled and exchangeable; the probability distribution that generates the data takes the form of a mixture model parameterised by the set $\{\theta, \pi\}$ there is a one to one correspondence between mixture components and classes (i.e. a component is associated to one class only).A new rotation invariant pattern descriptor is described. The distribution of textons in a patch is modelled by a mixture distribution, based on the assumption of textons independence and exchangeability. The optimal number of classes to describe a database can be computed exactly by minimizing the stochastic complexity of the model.

B. Proposed Approach

The proposed approach is based on Multi SVM algorithm. It makes use of Scale Invariant Feature Transform (SIFT).The block diagram is shown below. First the input image or target image that has to be detected is identified. The next step is to detect the key-points in the image. After detecting the key-points then feature extraction is done. It is done by using simple statistics, homogenous texture and color histogram. Then by using Euclidean distance measure the most similar images are retrieved.

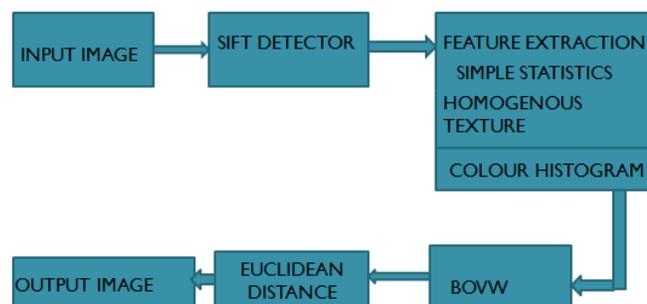


Figure 1.1 Block Diagram of Proposed Approach

II .SIFT**A. SIFT**

Scale-invariant feature transform (or SIFT) is an algorithm to detect and describe local features in images. The algorithm was published by David Lowe in 1999. SIFT descriptors are extracted from an image in two steps. First, a detection step locates points that are identifiable from different views. This process ideally locates the same regions in an object or scene regardless of viewpoint or illumination. Second, these locations are described by a descriptor that is distinctive yet invariant to viewpoint and illumination. For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a door were used as features, they would work regardless of the door's position; but if points in the frame were also used, the recognition would fail if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors.

SIFT^[18] can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes.^[17] This section summarizes Lowe's object recognition method and mentions a few competing techniques available for object recognition under clutter and partial occlusion.

SIFT-based analysis exploits image patches that can be found and characterized under different image acquisition conditions

Following are the major stages of computation used to generate the set of image features:.

1. Scale-space extrema detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation. Candidate locations are initially selected from local extrema in difference of Gaussian (DoG) filtered images in scale space. The DoG images are derived by subtracting two Gaussian blurred images with different σ

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

where $L(x, y, \sigma)$ is the image convolved with a Gaussian kernel with standard deviation σ , and k represents the different sampling intervals in scale space. Each point in the 3-D DoG scale space is compared with its eight spatial neighbours at the same scale, and with its 18 neighbours at adjacent higher and lower scales. The local maximum or minimum are further screened for minimum contrast and poor localization along elongated edges. The last step of the detection process uses a histogram of gradient directions sampled around the interest point to estimate its orientation. This orientation is used to align the descriptor to make it rotation invariant (RI).

2. Key point localization: Key points are selected based on measures of their stability.

3. Orientation assignment: One or more orientations are assigned to each key point location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

4. Key point descriptor: The local image gradients are measured at the selected scale in the region around each key point. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

A SIFT descriptor is extracted from the image patch centered at each interest point. The size of this patch is determined by the scale of the corresponding extreme in the DoG scale space. (For our evaluation data set below, most patches range in diameter from 6 to 50 pixels with a few that are larger.) This makes the descriptor scale invariant. The feature descriptor consists of histograms of gradient directions computed over a 4×4 spatial grid. The interest point orientation estimate is used to align the gradient directions to make the descriptor RI. The gradient directions are quantized into eight bins so the final feature vector has dimension 128 ($4 \times 4 \times 8$). This histogram-of-gradients descriptor can be roughly thought of a summary of the edge information in a scale and orientation normalized image patch centred at the interest point. We also consider extracting SIFT descriptors from a fixed grid instead of from the salient interest points. We refer to this as grid-based feature extraction. It is often all called dense sampling as it typically results in a larger number of descriptors since interest points are not detected in non-salient regions (of uniform intensity for example). We refer to the standard approach as saliency-based feature extraction.

This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features.

First, the SIFT detector is translation, rotation, and scale invariant which is the level of invariance needed for our application. Second, an extensive comparison with other local descriptors found that the SIFT descriptor performed the best in an image matching task.

III. IMAGE RETRIEVAL

Figure. 3.1 shows the basic image retrieval process. It involves extracting features from the query image. Thus query image features are obtained. From the image collection database we extract features. Then we form a database of those features. The next step is similarity matching. The comparison of query image features and those features from database to obtain retrieved image. The image having similar features is retrieved.

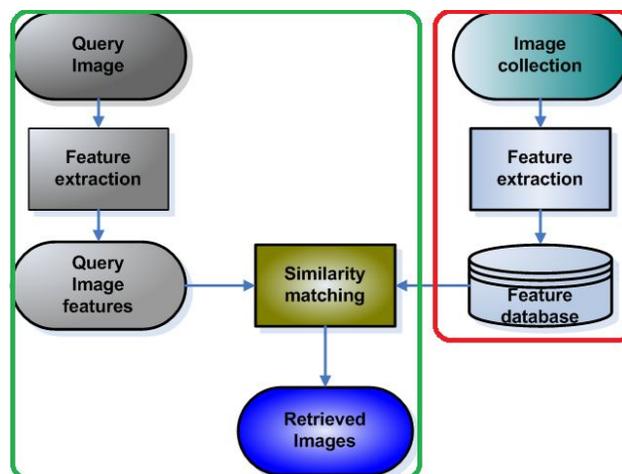


Figure 3.1. Basic Image Retrieval

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

- A. **Feature Extraction:** In this step local features are extracted from the image. Local features are designed to find local image structures in a repeatable fashion and to represent them in robust ways that are invariant to typical image transformations, such as translation, rotation, scaling, and affine deformation. Local features constitute the basis of approaches developed to automatically recognize specific objects[8]. The most popular local feature extraction method is the Scale Invariant Feature Transform (SIFT), introduced by Lowe. A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. Gabor filters have been widely used in pattern analysis applications.^[16] Three standard image features are considered: simple statistics, homogeneous texture, and color histogram features.

1.Simple Statistics: A 2-D feature vector is computed for each ground truth image consisting of the mean and standard deviation of the grayscale values,

$$f_{SS} = (\mu, \sigma).$$

This is referred to as the simple statistics feature and serves as a baseline for the experiments.

2.Homogeneous Texture: Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single colour or intensity. It is an innate property of virtually all surfaces, including clouds, trees, bricks, hair, fabric etc.[9] Homogeneous Texture Descriptors compliant with the MPEG-7 Multimedia Content Description Interface [56] are extracted using banks of Gabor filters tuned to five scales and six orientations. A 60-dimensional feature vector is formed from the mean and standard deviation of the 30 filters

$$f_{texture} = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \dots, \mu_{1S}, \sigma_{1S}, \dots, \mu_{RS}, \sigma_{RS}]$$

where μ_{rs} and σ_{rs} are the mean and standard deviation of the output of the filter tuned to orientation r and scale s . To account for differences in range, normalized versions of the features are also produced in which each of the $2RS$ components is scaled to have a mean of zero and a standard deviation of one over the ground truth data set. Simple cells in the visual cortex of mammalian brains can be modelled by Gabor functions.^{[12] [13]} Thus, image analysis with Gabor filters is thought to be similar to perception in the human visual system. Its impulse response is defined by a sinusoidal wave (a plane wave for 2D Gabor filters) multiplied by a Gaussian function.^[14] Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions.^[15] The two components may be formed into a complex number or used individually.

3. Color Histogram: Colour histogram features are computed in three colour spaces: RGB, hue lightness saturation (HLS), and CIE Lab. Each dimension is quantized into eight bins for a total histogram feature length of 512. The histograms are normalized to sum to one (L1 norm equal to one). These results in three different colour histogram features: f_{RGB} , f_{HLS} , and f_{Lab} .

B. BOVW Representation

Bag of visual words (BoVW) is a popular technique for image classification inspired by models used in natural language processing. BoVW downplays word arrangement (spatial information in the image) and classifies based on a histogram of the frequency of visual words. To obtain a visual vocabulary, a large set of local features extracted from a training image corpus are clustered. In this way the local feature space is divided into informative regions (the visual words) and the collection of the obtained visual words is the visual vocabulary.

C. MULTI SVM

SVM are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two

categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Multiclass SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. SVMs can be used to solve various real world problems. SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings. Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly. Hand-written characters can be recognized using SVM.

Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support vector machines, a data point is viewed as a p -dimensional vector (a list of p numbers), and we want to know whether we can separate such points with a $(p - 1)$ -dimensional hyper-plane. This is called a linear classifier. There are many hyper-planes that might classify the data. One reasonable choice as the best hyper-plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper-plane so that the distance from it to the nearest data point on each side is maximized.

IV. SIMULATION RESULTS

Figure. 4.1 shows the selection of input image. The target image that has to be selected. It is this image that has to be tested and identified what type of geographic image it is.



Figure. 4.1.Input image

Figure 4.2 shows the important locations in the image referred to as key-points. The key-points are detected using SIFT detector. The key-points are assigned different orientations. It is indicated in figure 3. Each key-point is assigned a particular magnitude and orientation.

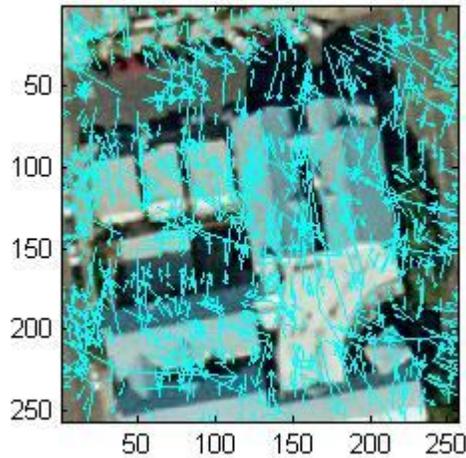


Figure. 4.2. Detecting the key-points in input image

Figure. 4.3 shows homogenous feature extraction. Feature extraction is done to extract features from an image. It is done by using Gabor filter. Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

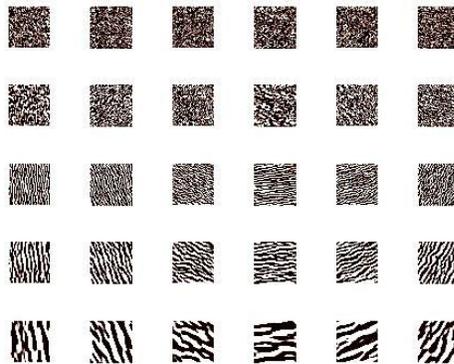


Figure. 4.3 Homogenous Texture Extraction

Figure. 4.4 shows feature extraction using colour histogram. The three different colour spaces used are RGB, HLS and CIE lab. It is indicated in figure 4.4.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

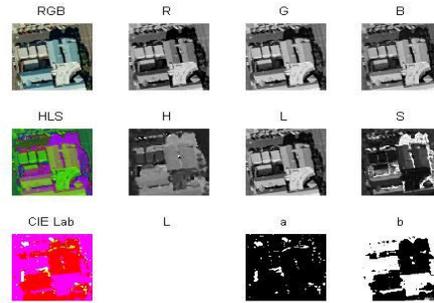


Figure. 4.4. Colour Histogram

Figure 4.5 shows the most matching results. The most similar images are shown below.

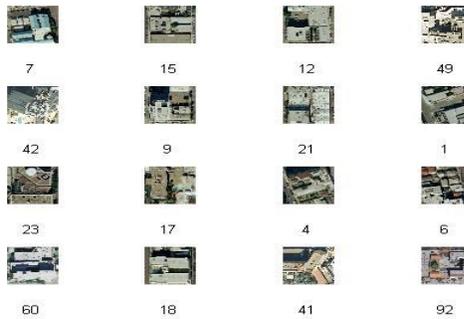


Figure. 4.5. Most Matching Results

Figure 4.6. shows precision vs recall curve. The comparison is done using local features, texture, colour histogram and simple statistics. Local features are better than other features.

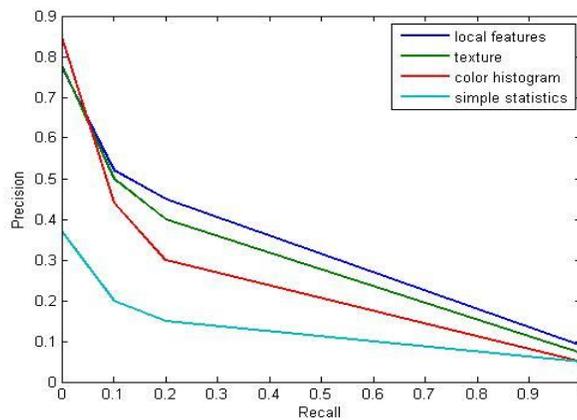


Figure. 4.6. Precision vs Recall curve

V. CONCLUSION

We presented an investigation into local invariant features for overhead image retrieval, the first such study of its kind. We demonstrated that local invariant features are more effective than standard features such as color and texture for image retrieval of LULC classes in high-resolution aerial imagery. We performed image retrieval using Multi Support Vector Machine and it was seen that the false image retrieval has been considerably reduced.

REFERENCES

- [1]. Yi Yang and Shawn Newsam, "Geographic Image Retrieval Using Local Invariant Features" IEEE Trans. Geosci. Remote Sens. vol.51,no.2, Feb 2013.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no. 2, May 2008.
- [3]. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images", IEEE Trans. Geosci. Remote Sens., vol. 49, no. 1, pp. 211 to 221, Jan.2011
- [4]. Lijun Chen, Wen Yang, Kan Xu and Tao Xu, "Evaluation of Local Features for Scene Classification Using VHR Satellite Images" IEEE JURSE 2011 - Joint Urban Remote Sensing Event Munich, Germany, April 11-13, 2011
- [5]. Grant J. Scott, Matthew N. Klaric, Curt H. Davis and Chi-Ren Shyu "Entropy-Balanced Bitmap Tree for Shape-Based Object Retrieval From Large-Scale Satellite Imagery Databases" IEEE Trans. Geosci. Remote Sens. , vol. 49, no. 5, May 2011
- [6] Beril Srmaek and Cem nsalan "Urban Area Detection Using Local Feature Points and Spatial Voting", IEEE Geosci. Remote Sens. Letters , vol. 7, no. 1, Jan 2010.
- [7] Qiao Sheng Xu, Tao Fang, Deren Li, and Shiwei Wang "Object Classification of Aerial Images With Bag-of-Visual Words", IEEE Geosci. Remote Sens. Letters , vol. 7, no. 2, Apr. 2010.
- [8]. Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of CVPR, pages 2169–2178, Washington, DC.
- [9]. John R. Smith and Shih Fu Chang ".Automated binary texture feature sets for image retrieval". In Proc.IEEE Int. Conf. Acoust.,Speech and Signal Proc.,May 1996
- [10]. Bahadr O zdemir, Selim Aksoy "Image Classification Using Subgraph Histogram Representation ". ICPR 2010 20th International Conference on Pattern Recognition, pages 1112 – 1115, Aug. 2010
- [11] Jean-Baptiste Bordes, Veronique Prinnet "Mixture Distributions for Weakly Supervised Classification in Remote Sensing Images ".BMVC 2008
- [12]. S. Marčelja. "Mathematical description of the responses of simple cortical cells." 'Journal of the Optical Society of America', 70(11):1297–1300, 1980.
- [13]. J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, July 1985.
- [14]. Fogel, I.; Sagi, D. (1989). "Gabor filters as texture discriminator". *Biological Cybernetics* **61** (2)
- [15]. 3D surface tracking and approximation using Gabor filters, Jesper Juul Henriksen, South Denmark University, March 28, 2007
- [16]. Haghghat, M.; Zonouz, S.; Abdel-Mottaleb, M. (2013). "Identification Using Encrypted Biometrics". *Computer Analysis of Images and Patterns*. Lecture Notes in Computer Science **8048**. p. 440
- [17] Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision* **2**. pp. 1150–1157.
- [18] David Lowe "Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image", David Lowe's patent for the SIFT algorithm, March 23, 2004