Bag of Features Based Remote Sensing Image Classification Using RANSAC And SVM

Bharathi S[1], Karthik Kumar S[2], P Deepa Shenoy[3], Venugopal K R[3], L M Patnaik[4]

Abstract—In this paper Bag of features framework for remote sensing image is proposed. Different families of bag-of-features are designed to capture the invariance of object translation, rotation, and scaling. The proposed framework, works well with feature extraction part. The most representative features are selected using Gabor convolution, Scale-invariant Feature Transform (SIFT) keypoints and Random Sample Consensus (RANSAC) method. PCA is used to increase data retrieval speed by reducing dimensionality. Vocabulary building is very challenging and difficult task. The feature vector forms the classifier under K-means and Support Vector Machines (SVM) for semantic annotation. In the testing stage, the keypoints are extracted from every image, fed into the visual dictionary to map them with one (Bag-of-feature) vector, which is finally fed into the multi-class SVM training classifier model to recognize the category of an image. The time complexity of the classification is not very complex; it takes 3mins for our dataset. Bag-of-feature is one of the best methods for content based image classification.

Keywords—visual dictionary, SIFT, RANSAC, SVM, Bag of Features.

I. Introduction

Research in computer vision involves finding distinctive features, which are efficient in computation time and memory consumption. The classifiers, on the other hand, are required to have good generalization ability, yet achieve high performance rates. Remote sensing image classification has become very important research area. The bag-of-features approach which is derived from text mining methods has shown impressive levels of performance on image classification tasks. Image classification is an important application in computer vision. Our research goal is to improve methods for Image classification, more specifically for objects from remote sensing image. An image is classified based on its semantic category of a scene like forest, road, building etc using Bag-of-features.

In recent years the Bag-of-features (BOF) model has been extremely popular in image categorization. Bag-of-features approaches have been successfully applied to visual recognition or classification. In such approaches, images are represented as histograms built from a sparse set of visual features. The method treats an image as a collection of unordered appearance descriptors extracted from local patches. Then the patches or descriptors are quantized into discrete visual words of a codebook dictionary, and then the image histograms are compared and classified according to the dictionary. In this paper, Gabor convolution, SIFT keypoints and RANSAC methods are used for feature detection. For a given collection of image patches, vocabulary is formed by performing K-means clustering algorithm. Data set is trained using SVM. Finally, visual vocabulary is ready for BOF. Then remote sensing images are classified using test data.

Gabor features are widely used in many computer vision applications such as image segmentation and pattern recognition. To extract Gabor features, a set of Gabor filters tuned to several different frequencies and orientations are utilized. The computational complexity of these features, due to their non-orthogonality, prevents their use in many real-time or near real-time tasks.

SIFT keypoints are used for feature extraction. It has the following features. It has rich information and is suitable for fast and accurate features in huge data set and will produce a large number of feature vectors even though there are a few objects. It achieves real requirements through optimization of SIFT matching algorithm. It is convenient to combine with other forms of feature detectors. RANSAC has been used for robust fitting of a model from a set of observed data which contains outliers valuable for the optimization of image matching. In many cases, RANSAC is an effective robust estimator which can get the correct matches even when more than 50% mismatches exist in the sample. Support vector machine (SVM) is state-of-the-art machine-learning algorithm based on statistical learning theory. It is a set of related learning algorithms used for classification and regression. SVM is also a non-parametric classifier. It tries to find the optimal classification hyper plane in high dimensional feature space to handle complicated classification and regression problems by solving optimization problems.

Visual vocabulary construction is critical part of the Bag of features (BOF) model. The visual dictionary is a way of constructing a feature vector for classification that relates "new" descriptors in query images to descriptors seen in training. Visual dictionary consists of (1) collection of large sample of feature images, and (2) quantizing the feature space according to their statistics. Then a novel images features can be translated into words by determining which visual word they are nearest-to in the feature space. In this paper BOF is used for remote sensing object classification.

A. Organization

This paper is organized as follows – Section II deals with related topics. Section III describes the study area. Section IV presents the architecture model and methodology. Section V is about the problem definition. Section VI gives the implementation of the proposed algorithm and performance analysis. Section VII contains the conclusion.

[1]Research Scholar, Department of MCA, Bangalore University, Bangalore, India, [2]Department of MCA, Dr.AIT, Bangalore, India, [3]Department of computer Science and Engineering, UVCE, Bangalore, India, [4]Honorary Professor, IISC, Bangalore, India
Email ID: bharathishivu_s@yahoo.co.in

II. Related Work

BOF approaches have been applied successfully for nude detection [1] using Hue-SIFT descriptor. This approach has as its main advantage in the fact that it does not depend on any skin or shape models to identify nudity. This method works only for smaller data set. This has to be enhanced by adding structural and scale information to the basic BOF representation, experimenting with more sophisticated classifiers. Bag-of -Words is combined with texton-based classifier for differentiating anaplastic and non-anaplastic medulloblastoma on digitized histopathology[2], the experimental results are superior. Bag-of-features representations have recently become popular for content-based image classification owing to their simplicity and good performance. The basic idea is to treat images as loose collections of independent patches, sampling a representative set of patches from the image, evaluating a visual descriptor vector for each patch independently, and using the resulting distribution of samples in descriptor space as a characterization of the image[3]. Characterizing the influence of different clustering strategies and the interactions between sampling methods and classification are not considered. The time complexity is also very important for classification. After obtaining the features using SIFT, PCA is used decrease the projection time. In case of real time classification[4] using the above technique, 26 images per second can be classified.

A novel approach for[5] object recognition using tactile observations obtained from the touch-sensitive fingers of a manipulation robot is presented. Class of bag-of-features techniques are used to create a feature vocabulary for the tactile observations. By means of unsupervised clustering on training data, our approach learns a vocabulary from tactile observations which is used to generate a histogram codebook. The histogram codebook models distributions over the vocabulary and is the core identification mechanism. As the objects are larger than the sensor, the robot typically needs multiple grasp actions at different positions to uniquely identify an object. To reduce the number of required grasp actions, decision-theoretic framework is applied, that minimizes the entropy of the probabilistic belief about the type of the object. The Time complexity is very high to recognize whether and how objects are deformable. A bag-of-hierarchical-co-occurrence features method [6] is proposed in incorporating hierarchical structures for image classification. The spatial alignments of objects components are effectively characterized by the local autocorrelation function of visual words for local co-occurrences, and the method is shift invariant. Both narrow and broad descriptors are extracted and the visual words are hierarchically assigned to the descriptors to capture various levels of characteristics.

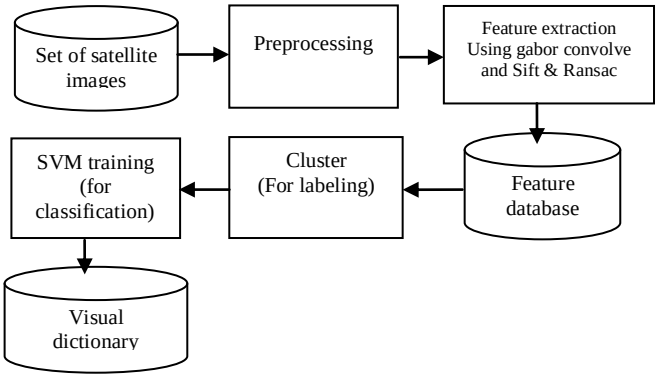
A method for obtaining discriminative and compact features by a Regularized Linear Discriminant Analysis – RLDA – subspace algorithm for visual categories is explored [7]. This approach significantly improves the result of visual vocabulary.

A multi-scale bag of feature approach is used for efficient large size map retrieval[8] using inverted file. In future this method has to extend for large-scale-mapping, self-localization, map-matching where map retrieval techniques should play an important role. BOF is also used for medical image retrieval and classification[9] efficiently. This framework works only for specific image, it has to generalize for all type of medical image retrieval methods to accommodate alternative assignment and weights. BoF for object categorization and semantic video retrieval[11] shows surprisingly strong performance regardless of order less and colorless representation. This work proposed a novel soft-weighting method to assess the significance of a visual word to an image. Local patches used are not invariant to scale and the SIFT descriptor is easily suffered from the quantization effect when dividing the regions around the keypoints into fixed grids.

III. Study Area

The area taken for study is Bangalore of different resolution. This area is highly heterogeneous, which has the combination of different features. The study area contains the buildings, water bodies, roads, barren land and green area. It is between the longitude 77 30’ 18.30’’ E to and latitude 12 56’ 56.49’’ N. This is an IRS p6 LISS III data acquired on 2006 with a spatial resolution 15m and four spectral bands (red, green, blue, swri2) were used. Data obtained by GPS (Global positioning system) were used as ground truth information for classification of the images in 2006.

IV. Methodology



a)Training stage: Process for BoF image representation

b) Testing stage: process for image classification

Fig. 1. Architecture for Bag of feature based classification

A. Preprocessing

Once the data is acquired, it needs preprocessing. This includes geometric corrections, radiometric correction and image clipping. Geometric corrections and radiometric corrections have already been done. Image is in RGB format, it is converted into gray-scale image and unwanted parts are clipped-off from the image. Image is passed through the digital filters to remove noise and inconsistency. The image is normalized by scaling its values so that they fall within a small-specified range, such as 0.0 to 1.0. Range

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[$newmin_a$, $newmax_a$] Variance is scaled to fit in the range one. This preserves the relationship with the original image. Range [$newmin_a$, $newmax_a$] is computed by

$$v' = ((vmin_a/(max_a - min_a)) * (newmax_a - newmin_a) + newmin_a) \quad (1)$$

Variance is scaled to fit in the range one and it is computed by,

$$v' = ((v - \overline{A})/\sigma_A) \quad (2)$$

where $\overline{A}$ and σA are the mean and the standard deviation respectively of image A. Both are linear operations.

B. Feature Extraction

This section briefly describes the feature extraction method used in this paper. Feature detection is the process of deciding where and at what scale to sample an image. The output of feature detection is a set of keypoints that specify locations in the image with corresponding scales and orientations. Gabor convolution and Sift keypoints and RANSAC feature extraction methods are used.

Gabor filters allow local frequency information to be extracted from an image. Gabor filters estimate the strength of certain frequency bands and orientations at each location in the image. Frequency and orientation representations of Gabor filters are similar to those of the human visual system. A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. Specific convolution kernels for convolution filters are used to extract features from remote sensing images.

Local features of an image are first projected to different directions or points to generate a series of ordered bag-of-features, based on which different families of spatial bag-of-features are designed to capture the invariance of object translation, rotation, and scaling. Keypoints are salient image patches that include rich local information of an image. Keypoints are detected by robust feature detection methods like SIFT - is a 128-dimensional feature vector that captures the spatial structure and the local orientation distribution of a region surrounding keypoints. The SIFT feature is one kind of local feature. It keeps invariant for rotation, zoom scale and light variation. It also has some stability for perspective transformation, affine transformation and noise.

RANSAC is an algorithm for robust fitting of models in the presence of many data outliers. RANSAC[12] is a re-sampling technique that generates candidate solutions by using the minimum number observations(data points) required to estimate the underlying model parameters, unlike conventional sampling techniques that use as much of the data as possible to obtain an initial solution and then proceed to prune outliers, RANSAC uses the smallest set possible and proceeds to enlarge this set with consistent data points. The RANSAC[10] algorithm proceeds as follows. Repeatedly, subsets of the input data (e.g. a set of tentative correspondences) are randomly selected and model parameters fitting the samples are computed. In a second step, the quality of the parameters is evaluated on the input data. Different cost functions have been proposed the standard being the number of inliers, i.e. the number of data points consistent with the model. The process is terminated when the probability of finding a better model becomes lower than a user controlled probability η. The $1 - \eta$ confidence in the solution holds for all levels of contamination of the input data, i.e. for any number of outliers within the input data.

C. Clustering (K-Means)

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. This algorithm looks for a fixed number of clusters which are defined in terms of proximity of data points to each other. This algorithm is to used classify or to group objects based on attributes into *K* number of group. *K* is a positive integer. The grouping is done by minimizing the sum of squares of distance between data and corresponding cluster centroid.

The K-means algorithm assigns each point to the cluster whose centroid is the nearest. The center is an average of all the points in the cluster, that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets. K-Means cannot handle non-globular data of different sizes and densities. K-means clustering is used to cluster the dataset which is generated from the image.

D. SVM for Training and Testing

Support Vector Machine (SVM) is a supervised machine learning method used for classification. An SVM kernel based algorithm builds a model for transforming a low dimension feature space into high dimension feature space to find the maximum margin between the classes. In the field of geospatial data processing, there is a high degree of interest to find an optimal image classifier technique. SVM is now considered to be one of the powerful kernel based classifier that can be adopted for resolving classification problems. For two-class case, the decision function for a test sample *x* has the following form:

$$g(x) = \sum_i \alpha_i y_i k(x_i, x) - b \quad (3)$$

where $k(x_i, x)$ is the response of a kernel function for the training sample x_i and the test sample x; y_i is the class label of x_i; α_i is the learnt weight of the training sample x_i, and b is a learnt threshold parameter. These feature vectors are known as support vectors. It can be shown that the support vectors are those feature vectors lying nearest to the separating hyperplane.

E. Visual Dictionary

The visual dictionary is a way of constructing a feature vector for classification that relates "new" descriptors in query images to descriptors. A visual dictionary is constructed to represent the dictionary by clustering features extracted from a set of training images. The image features represent local areas of the image, just as words are local features of a document. Clustering is required so that a discrete vocabulary can be generated from millions (or billions) of local features sampled from the training data. Each feature cluster is a visual word that is labeled and kept ready for classification. Given a novel image, features are

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

detected and assigned to their nearest matching terms and converts the BoF representation into a visual word vector.

F. Testing Stage

Each training image describes Bag-of-Features. The feature points are extracted from the testing image. Finally, the generated vector will be fed into the multi-class SVM training classifier model that was built in the training stage to classify and recognize the objects in the remote sensing image.

V. Algorithm

a. Problem Definition: For a given high resolution image, the main objective of this work is to:

i) To develop an efficient visual vocabulary
ii) To build an efficient, high performance image classification model using bag-of features

Algorithm: The main objective is to create an efficient visual dictionary of remote sensing images for different features and design a high performance classification model for object identification. The algorithm for classification is given in Table1.

Input: m: Set of high resolution satellite images
n: image for classification

Output:

a: Set *n* number of classified images with better accuracy

```
Bof-cls-model ()
begin
   Preprocess the image m.
   creating-dictionary (m)
   feature-extraction (m)
   k-means (data, n)
   SVM-training (samples, label)
   classify-image (n)
   classification-result (n)
end

creating-dictionary (m)
feature-extraction (m)
      Gabor- convolution ()
      calculate the radial filter component.
      for each row of the input image
            apply frequency domain convolution
      end
sift-key (m)
      Ransac (m)
      generate feature vector
      apply PCA
K-means (data, n)
Repeat
      Set initial centers of clusters, c1, c2…ck, to the arbitrarily selected k
      vectors
      Classify each vector x1 = [x11, x12…x1d] into the closest center ci
      Recalculate the cluster center ci = [ci1, ci2…cid] until centroid no
      longer moves
      Assign the label.
SVM-training (samples, labels)
      build SVM-kernal classifier
      (visual dictionary)
classify-image (n)
      input image for classification
      match (n)
            //extract local descriptor for testing image
      sift-key (n)
      assign the descriptor to visual modeling
      compare images using sift-key and
      ransac ([x1; x2], fitting, distfn, degenfn, s, t)
```

```
classification-result (n)
      //classify image using SVM
   SVMClass ((TestFeature'), AlphaY, SVs, Bias, Parameters, nSV,
   nLabel)
      //Annotation result
   Display classification result.
```

Fig. 2. Algorithm for Bag of feature based classification

VI. Experimental Results

A. Simulation Software

Simulation is performed using Matlab7.5. MATLAB is a high-level scientific and engineering programming environment which provides many useful capabilities for plotting and visualizing data and has an extensive library of built-in functions for data manipulation. Data preprocessing, image segmentation, classification and computation is done using the tools supported by MATLAB.

B. Performance Analysis

The Bag-of-features image representation is analogous. Bag-of-features approach first extracts patches from the images and use local feature descriptors to describe them, then codes these descriptors by quantization against a previously learnt “visual dictionary” (the vocabulary)[6]. This process tends to give the same label to similar local features.

A visual vocabulary is constructed to represent the dictionary by clustering features extracted from a set of training images. The image features represent local areas of the image, just as words are local features of a document. Clustering is required so that a discrete vocabulary can be generated from millions (or billions) of local features sampled from the training data. Each feature cluster is a visual word. Given a novel image, features are detected and assigned to their nearest matching terms (cluster centers) from the visual vocabulary. In this paper K-means clustering algorithm is used to cluster the detected features and label the classes. The SIFT descriptor[5] computes a gradient orientation histogram within the support region. For each of 8 orientation planes, the gradient image is sampled over a 4*4 grid of locations, thus resulting in a 128-dimensional feature vector for each region. A Gaussian window function is used to assign a weight to the magnitude of each sample point. This makes the descriptor less sensitive to small changes in the position of support region and puts more emphasis on the gradients that are near the center of the region. To obtain robustness to illumination changes, the descriptors are made invariant to illumination transformations of the form aI(x) +b by scaling the norm of each descriptor to unity.

Retrieval performance is evaluated using PCA, which reduces the dimensionality of dataset by transforming new set of variables to summarize the features of the data.

Support Vector Machines train a multiclass classifier on the labeled data. SVM can be a potential classification technique for high performance satellite image classification. SVM is an optimal classifier with a maximal margin in feature space. The SVM model is based on the principles of structural risk minimization. The visual dataset contains 3 categories viz. Building, Forest and Roads.

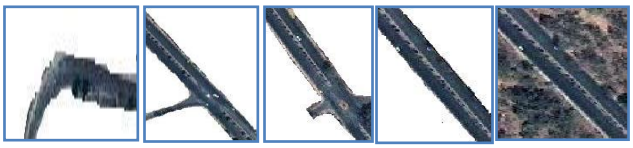
ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Once descriptors have been assigned to clusters to form feature vectors, the problem of generic visual categorization is reduced to that of multi-class supervised learning, with as many classes as defined visual categories. The categorizer performs two separate steps in order to predict the classes of unlabeled images: training and testing. During training, labeled data is sent to the classifier and used to adapt a statistical decision procedure for distinguishing categories. After keypoints are represented as bag-of-features, it can be classified as keywords in semantic annotation task. Then build supervised classifiers based on visual-word features from labeled images and apply them to predict the labels of other images. Each training image can be described as a "bag of words" vector by mapping the keypoints to a visual words vector. The above steps are used to maximize classification accuracy while minimizing computational effort.

In this work, focus is on object detection, recognition and classification. This simulation contains 50 samples for creating visual dictionary contains 30 buildings, 10 forest and 10 road samples. The most difficult part is in creating our own dataset and this is quite challenging. Upon selecting *Create database* option in the Menu, features will be automatically extracted and visual dictionary gets created. Dictionary is trained using SVM classifier. The classification results for different features are show in the Fig.3 for an image chosen from test data set. If object is not found in the test data set, it cannot classify and display the result can't classify. The proposed method is tested and verified for 30 images in the test data set. The results are shown in Table. 1 gives the accuracy of the classification algorithm and accuracy assessment is tabulated in the Table.2. The time complexity of the algorithm is also computed, it takes 3 mins to classify the entire dataset.

Another important application of Bag of features image classification is image retrieval, finding the most similar images in a gallery to a given query image. This is also implemented in our work along with the classification.

a) Building

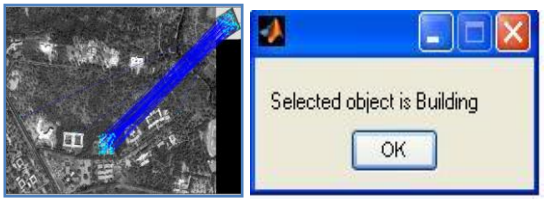


b) Road

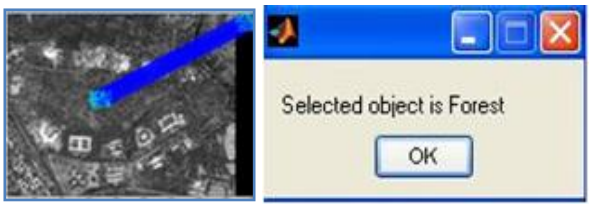
c) Forest

Fig. 3. Examples for few objects in the visual dictionary

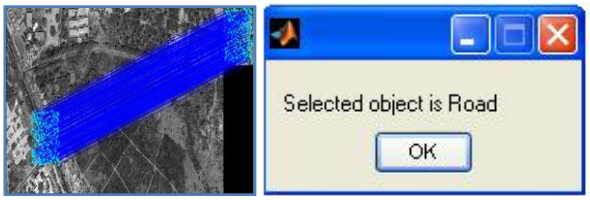
a)Menu and Testing images



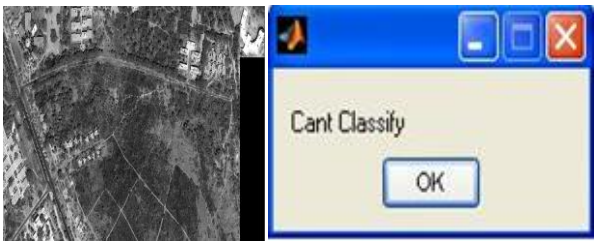
b) Scaling (rotation)



c) Illumination changes



d) Affine



e) Feature not found in the test image

Fig. 4. Classification result using BOF

Table I Confusion matrix for bag-of-feature based object detection

	Building	***Road***	***Forest***	***Classification accuracy in %***
Building	17	2	0	94
Road	1	5	0	90
Forest	0	1	5	90

Table II Accuracy assessment for the classification of objects

Accuracy	***Classification error***	***Kappa***	***Absolute error***	***Relative error***	***Normalized absolute error***	***Root mean squared error***
91	9	0.89	0.9 ±0.263	09.00 ±263	0.263	0.245 ±0.000

VII. Conclusion

The information in remotely sensed images plays an important role in environmental monitoring, disaster forecasting, geological survey, and other applications. With the steadily expanding demand for remotely sensed images,

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

many satellites have been launched, and thousands of images are acquired every day. This leads to an exponential increase in the quantity of remotely sensed images in database. Therefore, how to retrieve useful images quickly and accurately from a huge and unstructured image database becomes a challenge. Bag of features is one of the solutions for the above problem.

Bag of feature is basically a simplified representation of an image. Rather than simple feature matching, we create a global representation of an object. Thus, we take a group of features, create a representation of the image in a simpler form and classify it. The small scale important targets and regions of remote sensing image arrest more attention than the entire remote sensing image. These image slice features important targets and regions extracted using Gabor Convolution, SIFT and RANSAC. Then the features are clustered using K-Means clustering algorithm to construct vocabulary and trained using SVM classifier. Classification is done on the images in the test dataset.

Good results were obtained on the example images used in the experiments. Only three classes are considered in this work viz., building, forest and road. This technique can be used even to classify water bodies, vegetation land etc. Focus of this paper is on building a query and browsing the system. In future works, we propose to implement this algorithm for segmented images.

REFERENCES

[1] Ana P. B. Lopes, Sandra E. F. de Avila, Anderson N. A. Peixoto Rodrigo S. Oliveira and Arnaldo de A. Ara´ujo, "A Bag-Of-Features Approach Based On Hue-Sift Descriptor For Nude Detection",17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland, August, 24-28, 2009.

[2] Joseph Galaroa, Alexander R. Judkinsb, David Ellisonc, Jennifer Baccond, Anant Madabhushia," An Integrated Texton and Bag of Words Classifier for Identifying Anaplastic Medulloblastomas", 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 September 3, 2011.

[3] Eric Nowak, Fr´ed´eric Jurie, and Bill Triggs,"Sampling Strategies for Bag-of-Features Image Classification", Springer-Verlag Berlin, Heidelberg, pp. 490–503, 2007.

[4] J.R.R. Uijlings, A.W.M. Smeulders, R.J.H. Scha, "Real-time Bag of Words, Approximately", CIVR, Santorini, July 8-10, 2009.

[5] Alexander Schneider, J¨urgen Sturm, Cyrill Stachniss, "Object Identification with Tactile Sensors using Bag-of-Features", IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, USA. October 11-15, 2009.

[6] Takumi Kobayashi, Nobuyuki Otsu," Bag of Hierarchical Co-occurrence Features for Image Classification", International Conference on Pattern Recognition, IEEE, 2010.

[7] Xian-Hua Han1, Yen-Wei Chen,Xiang Ruan, "Image Recognition By Learned Linear Subspace Of Combined Bag-Of-Features And Low-Level Features", Proceedings of 2010 IEEE 17th International Conference on Image Processing, Hong Kong, September 26-29, 2010.

[8] Kondo Kensuke and Tanaka Kanji, "Multi-Scale Bag-Of-Features for Large-Size Map Retrieval", Proceedings of the 2010 IEEE International Conference on Robotics and Biomimetics, Tianjin, China, December 14-18, 2010.

[9] Jingyan Wang, Yongping Li, Ying Zhang, Chao Wang, Honglan Xie,Guoling Chen, and Xin Gao, "Bag-of-Features Based Medical Image Retrieval via Multiple Assignment and Visual Words Weighting", IEEE Transactions on Medical Imaging, Vol. 30, No. 11, November 2011.

[10] Liang Cheng, Hao Hu, Yecheng Wang, Manchun Li, "A New Method for Remote Sensing Image Matching by Integrating Affine Invariant Feature Extraction and RANSAC", 3rd International Congress on Image and Signal Processing (CISP2010),pp 1605-1609, 2010.

[11] Yu-Gang Jiang, Chong-Wah Ngo, Jun Yang, "Towards Optimal Bag-of-Features for ObjectCategorization and Semantic Video Retrieval", CIVR, Amsterdam, July 9–11, 2007.

[12] Shao-Wen Yang, Chieh-Chih Wang and Chun-Hua Chang," RANSAC Matching: Simultaneous Registration and Segmentation",Funded by Taiwan National Science Councilunder Grant 96-2628-E-002-251-MY3, 2010.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)