International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 8, Volume 2 (August 2015) www.ijirae.com

Salient Object Detection and Recognition

Madhu M Patil (MTech CSE)
Department of Computer Science and Engineering
KLS Gogte Institute of Technology,
Belagavi, Karnataka, India

Assoc Prof. S R Dhotre
Department of Computer Science and Engineering
KLS Gogte Institute of Technology,
Belagavi, Karnataka, India

Affiliated to Visvesvaraya Technological University, Belagavi

Abstract--- Object detection and recognition is the challenging task in the computer vision, because the property of the object changes at some moment. The challenging task is to detect the salient object, which is the prominent part in the image irrespective of background, size, angle, orientation and many more. Object recognition is the next part of object detection which aims to group the objects of the same kind and identify them. The project has training and test data set for experimentation. The objects are detected by extracting the features using the SIFT and Harris corner detection algorithms, and the image is transformed into the features vectors. These feature vectors form the dictionary of code words and they are clustered using K-means clustering and the centroids of similar code words are formed. Applying the feature quantization, the feature vectors are encoded and the distortion is reduced. All the SIFT features are categorized as the parts based on distance from cluster center to SIFT feature, so that each image can be represented with histogram, which shows the frequency of parts. The above mentioned approaches are applied for both training and test data sets. Using the KNN search, the objects for similar kind are searched in tests data from training data and recognize it. Thus the objects are recognized efficiently with less false images being detected. This method is implemented on Caltech101 and 17 flower category datasets and I found that the time and classification accuracy is better than the SVM classifier and spatial pyramid matching algorithms.

Keywords--- Object detection, Object recognition, SIFT, Harris corner detection, KNN search, K-means clustering.

I. INTRODUCTION

The salient object is the one that stands out relative to its neighbors. Object detection is the key mechanism that enables us to focus on the most prominent subset of available data. The objects are detected by extracting its features, such that the important information can be retrieved and is stored for object recognition purpose. The features are stored in the form of vectors and the data is trained such that useful information is stored.

The object recognition is a challenging task in computer vision. Humans recognize object from an image without any effort, where objects in an image vary in different viewpoints, orientation, scale, angle, illumination, translation may get changed. Humans can still recognize objects in images even if the objects are distorted. The task is very much challenging in computer vision. Thus efforts are being made to match the computer vision task and the human vision through feature quantization and histogram matching.

The feature vectors of object are used for object recognition. Each of the feature vectors contains information about each part of an image, so when these vectors are stored as code words in the training data, the recognition process becomes easier. For example, a story book contains many words, if any of the word is missed out, the information is lost and we will not be able to understand the story. In the similar way, all the code words contain information of an image which is stored as features.

The training data set and test data set are used for evaluation of the result, which contains image information. The training set is used to predict the response from many predictors. It is used to fit the parameters of classifier and is used to find the optimal weights. Both training set and test set, stores image information in the form of vectors. The training set is the raw data which stores large data. On the other hand, the test set is user given input for recognition from training set. The test data estimates how good the images are trained and to estimate image properties. The test set assesses performance of training set.

Thus the data is ready for object recognition, which needs classifier to do so. The classifier analyzes the mathematical properties of object features in an image, with data being organized into different categories. The classifier has two phases: First of all, through training phase, the unique description of an object is found. In second phase, through test set, the features are used to classify image features. All these methods help in successful detection and recognition of salient objects from large data set. The diagram below shows simple flow of object detection and recognition task.

II. RELATED WORK

The object detection and recognition in image processing is an important and challenging task. Many algorithms have been proposed to perform this task, as it is applied in almost all fields of image processing like in medical fields, agricultural fields, traffic controlling, and also in military and army and many more. Thus much work has to be carried out to efficiently detect and recognize salient objects with less number of false images being detected and perform the task in less time.

Earlier many of the object recognition approaches were proposed which includes as follows: in [11] we see that geometric representations dominated the development of analytic representations of algorithms. It is invariant to viewpoint and illumination, and also it has well developed theory. It can recognize objects with well defined shape. It can be expected as an affine patch feature woven into edge based prior art.

Further in [17], text retrieval approach for object matching was proposed. The text retrieval is based on implementation where descriptors matches were computed before quantizing the vector and file system inversions. The result of the retrieval is immediate, which returns the ranks of key frames similar to Google. But the textual vocabulary is static so it is not universal, thus cannot detect objects of all kind and is restricted to objects of well defined shape.

In [14], it finds correspondence between feature vectors. The algorithm uses correspondences as programming of quadratic integer, where function costs are based on geometric distortion similarity between corresponding feature points pairs. It handles outliers, but it assumes that data is clean and that there is no intra class variation.

The [12] is built upon index descriptor techniques obtained from local regions, and is robust to clutter and occlusion. The descriptors of local region are usually quantized hierarchically in vocabulary tree, which allows vocabulary to be used efficiently. It defines quantization directly, but it is not robust and takes more time.

In [13], the illumination change challenges are addressed by feedback strategy. The self adaption in object detection and recognition is possible in by variable illumination. Self adaption is achieved through feedback from detection to recognition phase, but it is not suitable when objects are present in an image and also the range of objects detected is not so efficient.

In [16], it has a method for invariant features extraction which are distinct from all images to perform matching between views of an object or image. The features are invariant to scale, rotation and has robust matching across range of affine distortion and change in illumination. Here a single feature matched correctly against database features which are large from many images with high probability. The objects among clutter and occlusion are robustly identified with near real time performance. Image matching deals with distance computations across all SIFT feature pairs and is costly.

In [7], when medians are used as thresholds of quantization, each feature is quantized to binary. Quantized features preserve both matching properties and distinctiveness, and after each quantization, each feature set that is tested maps to distinct binary patterns. Further, the match numbers between images with original and binary quantized SIFT features are almost similar. The binary SIFT features gives results comparable to original SIFT feature, but with more computation time.

In [2], it shows that Harris corner detection is better and not for color images, therefore [3] is proposed. In [3], use of color information is investigated in interest point detection. Corner points are extracted from both gray and color images and the extracted corners are analyzed for applications in image matching using cross correlation. The number of interest points increases with this method and they shift to more stable and distinct locations than luminance based methods for various transformations.

In [4], the difficulty of occlusion, background clutter and lighting changes are reduced and similar paths obtained after the SIFT operation are clustered in the same groups to constitute a cluster center of these clusters called visual words and the set of cluster centroids is treated as a vocabulary. The object recognition using this procedure is easy and efficient, but difficulty arises where the size of vocabulary is very large.

In [5], for each extracted visual word from query image, the weighted schemes use fixed weights, which may lose discriminative information. It has novel combining method for capturing query specific weights for visual words in every given image.

In [9], image local features form clusters that are described as visual word vocabulary. The feature extracted from an image is matched into visual word that is closest and the image is represented as histogram over vocabulary of visual words. It bridges the gap between low-level visual features and high-level concepts which are categorized. The proposed system is enhanced over classical BOW model.

[10] Gives the searching technique where two points want to operatively compute query point for K Nearest Neighbors without revealing to third party about private inputs. In [18], K-means clustering is explained where it is the unsupervised learning algorithm that classifies given data set through fixed number of clusters. The main idea is that for each cluster we should define K centroids

[8] Gives details of various detectors and descriptors for object detection and recognition purpose. It also compares different descriptors and suggests the best one which executes in less time. [6] Helps in recognizing object in an image from a cluttered environment and the selected views are then assumed as a model of object, which represent main properties of object, but getting the important views is the difficult task.

In [1], the SIFT extracts distinctive features from image which are useful for matching views of an object feature that is selected and design of an effective classifier helps in successful detection of an object and uses SVM classifier which takes more time. In [19], the training set and test sets have been explained which helps in storing features vectors efficiently. In [20] many feature detection and extraction methods are explained and [17] compares performance of descriptors computed for local interest region which concludes that SIFT is the better descriptor.

All the methods described above, have few shortcomings when implemented individually. For example when SIFT alone is used for feature extraction it may not consider the corner features. Thus here I consider combination of feature detector algorithms to get all features of an image.

In the similar way I consider the combined methods for clustering and classification approaches to design our algorithm efficiently. Thus I design my algorithm in such a way that few of the methods mentioned above can be combined and get the better results and get object detected and recognized efficiently and robustly.

III. SYSTEM ARCHITECTURE

To design the system I consider training data set and test data set for which the below mentioned procedures can be applied to get the desired results. The training set is the raw data which stores large data after the code word generation and the test data set is user given input for recognition from training set. The whole system can be designed in five steps:

1. Image acquisition.
2. Pre-processing.
3. Feature extraction and visual code word generation.
4. Feature quantization using code words.
5. Image representation and recognition.

(The steps from 1 to 4 are applied both for training and test data set. The 5th step is for the evaluation purpose which takes both the training and test data sets into account.)

A) IMAGE ACQUISITION.

There are different kinds of images with different contexts and also the sources for getting image is also different. The first task of any image processing techniques is to acquire the images from specific source Images can be acquired from the camera or from standard data sets available from the internet. In mu project we use the Caltect101 data sets, 17 category flower set.

B) PRE-PROCESSING.

Pre-processing is mainly done to remove the unwanted information and fix some values for the images, so that the value remains same throughout the project. In my project, pre-processing of the images is done by resizing to 256*256 and fixing the resolution. Two kinds of file formats are considered that is, .jpg and .bmp file formats. Any image of other file formats are found, they will not be considered for further processing. Thus the image data set is ready for further processing.

C) FEATURE EXTRACTION AND VISUAL CODE WORD GENERATION.

After pre-processing I extract the features from images using two of the approaches Harris corner detection and SIFT approach and generate the code words to store the useful information in the training data and test data set. The feature extraction helps in storing the useful information of images as vectors. The code words are basic elements to describe an image, where the interest points are detected using SIFT detector and image patches are represented by SIFT descriptors. Here features vectors obtained from SIFT descriptors are grouped which are similar, and the codebook is obtained. Every visual word is compared with code words and assigned to the nearest code word. Matching histograms are plotted and the number of bins in histogram, counts the number of words assigned to the code word. The code words are clustered and are stored in the training data set.

1) Harris corner detection: Harris corner detection is an edge detection algorithm which takes image corners into account. It is invariant to rotation, scale, illumination variation and image noise. It gives mathematical approach to determine whether the region is flat or edge or even corner. It is based on the function of auto correlation of an image and the function that measures changes of an image, where the patches are shifted by least amount in various directions.

2) SIFT approach: SIFT, the Scale Invariant Feature Transform is considered as both detector and descriptor. The image is transformed into a local feature vectors with translation, scale, rotation, and partially illumination invariant. In detection part, the key are selected from locations that are maxima or minima of a “Difference of Gaussian function”. Then it locates key-points at different regions, which makes these relevant points stable for characterizing an image. Features returned by SIFT are local and appearance based, according to key points given by the detector. It removes the effect of rotation and scale and creates descriptor using histogram of orientations.

D) FEATURE QUANTIZATION USING CODE WORDS.

After the features are extracted from images, feature vectors are obtained. Using these feature vectors, the nearest SIFT using the distance formula for each image is obtained. Then the histogram of similar features is obtained. The histogram of similar features is stored in the training data set for further classification. The main step here is to get the feature vectors. The same procedure is followed for test data set. Thus the training data set and test data set are ready for classification and recognition. The KNN classifier is used to get the 4 nearest neighbors for the test data set from the training data set. The optimal value is fixed for the matching of similar images from the data set. It is fixed to 2. If the match is greater than 2, image is considered to belong to given data set else it will be considered as false image.

E) IMAGE REPRESENTATION AND RECOGNITION.

The KNN classifier is used to get the 4 nearest neighbors for the test data set from the training data set. The optimal value is fixed for the matching of similar images from the data set. It is fixed to 2. If the match is greater than 2, image is considered to belong to given data set else it will be considered as false image.

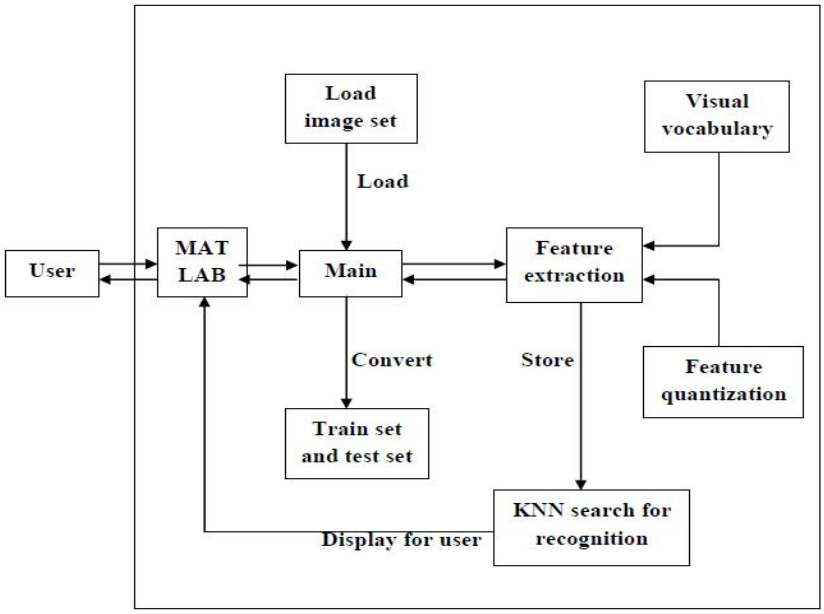

Fig 1. System Arctitecture

The system architecture involves the components as depicted in the above diagram. The MATLAB helps user to control the application developed. The main module loads the image dataset stored in database in .jpg format. Then for each image, features are extracted to store useful information of each part of an image using SIFT and Harris corner detection algorithms. Then these features are stored as visual code words which have all information. These visual code words are clustered in same group to constitute cluster. The cluster centers are called visual code words and the set of cluster centroids is vocabulary. Thus, dictionary of visual code words are obtained which contains description of each and every part that are categorized. But one should know which part belongs to which, to represent an image. The frequency of occurrence of each part represents an object, therefore I train the system. After the features have been detected using SIFT and Harris corner detection algorithms, I find nearest part that match these features. Now every feature is categorized into one of the nearest part based on the distance from features extracted to cluster center. Next, histogram for all cluster centers and frequency of parts are formed. Thus after calculating the distance between nearest parts of training dataset and test dataset, I can have the count of matches and objects recognized, and thus output is sent to MATLAB for display.

IV. IMPLEMENTATION

The implementation process is divided into four step process:

1. Load image.
2. Feature extraction and code word generation.
3. Feature quantization.
4. Representation of recognized images.

A) MODULES.

The implementation flows easily when all the four modules mentioned above works correctly. Therefore all four should be executed individually to reduce the termination of program. Each of the modules have their own significant part in program execution, if any one goes missing, it leads to improper results or even the termination of whole project.

1) Load Image.

The module 1 deals with loading of images from database. It can be static or dynamic. All image categories that I obtain from database should always be in the same path. In the program, first 500 images from each category are taken into account for training data set. And for test data set, I have considered 50 images from each category. Therefore the first module is successful only if the specified categories are present in database else it gives message to load different category of images.

2) Feature Extraction and Code Word Generation.

The module 2 deals with feature extraction and code word generation. It uses Harris corner detection to detect edge features to get proper shape of the images, which uses the gray scale image. The column co-ordinate feature vector of Harris corner detection procedure are used by SIFT descriptors to get feature vectors of every part of image. Then I store these feature vectors in the form of array, and generate code words by clustering these feature vectors using k-means clustering. I store these clustered vectors in training dataset.

3) Feature Quantization.

The module 3 deals with feature quantization. Using SIFT algorithm, it extracts features from image data set which is given as input. Then it finds the nearest SIFT feature from the earlier trained data set using distance measure. Then I store the histograms that have similar features in matrix of training data. I repeat the same procedure for test data set and keep both training data and test data for evaluation and recognition purpose.

4) Representation Of Recognized Images.

The module 4 deals with representation and recognition of images from the image data set. I have used KNN classifier to find four nearest neighbors. The test data set using KNN classifier finds four nearest neighbors from training data set. Thus images are classified and similar set of images are counted and recognized, leading to successful execution of the program.

B) ALGORITHMS USED

1) Feature extraction and code word generation: For any given image set, the features are extracted using SIFT and Harris corner detection. The gray scale image is used for process. Before the features are extracted, I fix the threshold, sigma and radius values. The Harris corner detection returns binary I mage, marking the corners, row co-ordinates and column co-ordinates of corner points. The columns co-ordinates are used by SIFT descriptor and store feature vectors in matrix format. Then these vectors are clustered and are stored in training data set.

Algorithm:
Load image sets
Count the image sets
For i= 1 to count of image set
For j=1 to count of images in image sets
Read image I
Convert to gray scale I
Fix threshold, sigma, and radius
(ci,r,c)=harrisdetection(I, threshold, sigma, radius)
Siftarray=sift (I, c, 1.5)
[code word ; siftarray]
End
End
Save the code-words as training data
Cluster the code-words
Save the clustered code words

2) Representation: After the feature extraction, I store the feature vectors in training data set and also the features are extracted from test data set. Then find the distance from feature extracted to cluster center and form histogram of similar features and store the histogram in training data matrix. Then using KNN classifier, from the test data I find four nearest neighbors. If the value is more than 2 (which is considered optimal), match can be found, and thus represent the images recognized.

Algorithm:
Load trained dataset
For i=1 to count of image set
For j=1 to count of images in image set
Read image I
Convert to gray scale I
(ci,r,c)=Harrisdetection()
Siftarray=sift (I, c, 1.5)
Calculate the nearest sift
Hist (similar)
Store (train)
End

End
Repeat the procedure for test dataset
Find the 4- nearest neighbors: knn(traindata,4)
(2 if considered optimal)
If array>2
Count the images recognized

V. EXPERIMENTS AND RESULTS

All the results are obtained by running the system on machine with 2GB RAM. The results are checked for different possibilities like successful check and unsuccessful check for given image sets. The data sets considered for experiment are: motorbike set, airplane set, lotus set and car test. The training data set contains 500 images for each category, therefore total of 2000 images are considered for training data set and the test data set contains 50 images for each category. The optimal value considered for KNN classifier is 2. The smaller the values more the images belonging to category. Thus experimental results by giving different input types are as below. Graphs for true images matched and recognized, for mixed matching and false match is as below

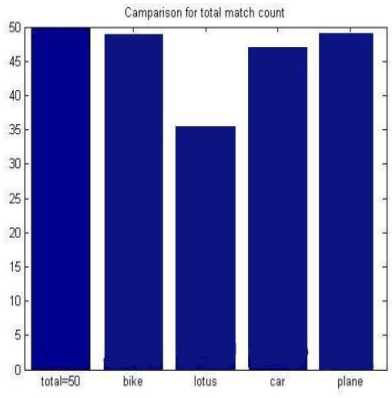

Fig 2. True match.

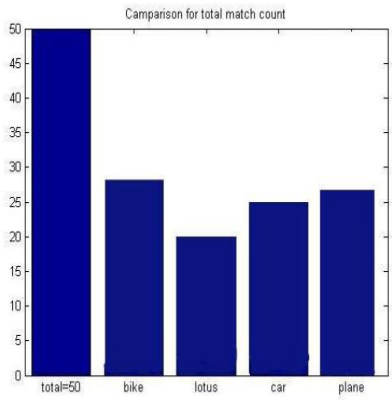

Fig 3. Mixed match.

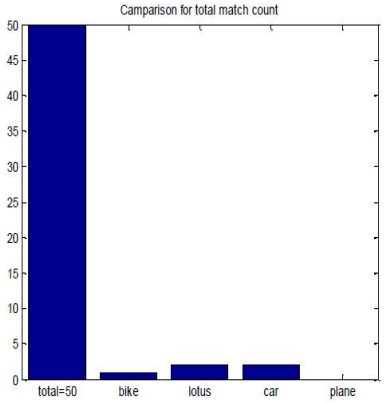

Fig 4. False match.

From the figure 2, the near accuracy of the results is obtained. The test data sets contain all the images belonging to same category. The image categories considered for recognition are matched exactly for each, which may have some false detection, as it depends on the features extracted for processing of these images.

From the figure 3, the near accuracy of the results for mixed match is obtained. The data set contained the mixed sets of all kinds and the images belonging to the given category are obtained efficiently.

From the figure 4, the near accuracy of the results for false match of Images is obtained. The test data sets contains image not belonging to the category. The image categories considered for recognition are not matched, thus it gives the better results. The result should actually be zero. But the result that I have got here is almost near to zero. Thus the false images are classified efficiently giving the better results.

VI. CONCLUSION AND FUTURE SCOPE

There are many method used for object detection and recognition purpose. My proposed system is one of them. The objects can be recognized by taking different features into consideration like shape, color, texture etc. In my project I have used shape and color features. So the work can be extended for object recognition using texture, which leads to more accurate results and also computational time increases. Thus the methods should be explored to detect and recognize the objects accurately within less time.

The proposed work shows near accurate results for salient object detection and recognition. Images loaded from database have 2500 image data set with each category containing around 500 images. The test data set contains 50 images from each category. The images that are classified using KNN classifier give near accurate results, which I can see from the graphs in the analysis part. The graphs shows that for 50 test images from each category, my model could recognize nearly all the images except 1 or 2, which may not belong to the category. Also some image belongs to the category, but is not detected, but this case is rare. Similarly for false image recognition, when none of the images belong to the category are checked, I could get the near results, but not accurately 0, my model may give the result as 1 or 0 and sometimes 2. When the test image contains multiple objects, our model can recognize it as belonging to the particular category. The results obtained are successful. Thus the images categorized into different types are obtained and are recognized correctly. The computational time to recognize images is less, so I can apply my proposed system to large data set belonging to specific category. The accuracy in recognizing the images of particular category is nearly 97%.

REFERENCES

[1]. Thao Nguyen, Eun-Ae Park, Jiho Han, Dong-Chul Park, Young Min titled "Object Detection Using Scale Invariant Feature Transform", in Genetic and Evolutionary Computing, Advances in Intelligent Systems and Computing in Springer International Publishing Switzerland, 2014.
[2]. Zhenxing Luo titled "Survey of Corner Detection Techniques in Image Processing", in International Journel of Recent Technology and Engineering (IJRTE) ISSN: 2277- 3878, Volume 2, Issue 2, May 2013.
[3]. B Sirisha, B Sandhya titled "Evaluation of Distinctinve Color Features from Harris Corner Key Points", in Advance Computing Conference (IACC), IEEE 3rd International Conference, 2013.
[4]. Mohammad Mehdi Farhangi, Mohsen Soryani, Mahmood Fathy titled "Informative Visual Words Construction to Improve Bag of Words Image Representation", in IET Image Processing, January 2013.
[5]. Lei Zhu, Hai Jin, 1ran Zheng, Xiaowen Feng titled "Weighting Scheme for Image Retrieval Based on Bag of Visual Words", in IET Image Processing, May 2013.
[6]. Alessandro Bruno, Luca Greco, Marco La Cascia titled "Object Recognition and Modelling Using SIFT Features", in Springer International Publishing Switzerland, ACIVES, pg 250-261, 2013.
[7]. Kadir A Peker titled "Binary SIFT: Fast Image Retrieval Using Binary Quantized SIFT Features", in Content Based Multimedia Indexing (CBMI), 9th International Workshop, pg 217-222, 2011.
[8]. [8] J Bernal, F Vilari, J Sanchez titled "Feature Detectors and Feature Descriptors: Where We Are Now", in Computer Vision Center 2010.
[9] Tinglin Liu, Jing Liu, Qinshan Liu, Hanqing Lu titled "Expanding Bag of Words Representation for Object Classification", in Image Processing (ICIP), 16th IEEE International Conference, 2009.
[10] Yinian Qi Mikhail J Atallah titled "Efficient Privacy Preserving K-Nearest Neighbor Search", in International Conference on Distributed Computing Systems IEEE, 2008.
[11] Joseph L Mundy titled "Object Recognition in Geometric Era: A Retrospective", in Springer J Ponce et al. (EDS): Toward Category Level Object Recognition, LNCS 4170, pg 3-28, 2006.
[12] David Nister, Henrik Stewenius titled "Scalable Recognition with a Vocabulary Tree", in CVPR 2006.
[13] Qiang Zhou, Limin Ma, David Chelberg titled "Adaptive Object Detection and Recognition Based on a Feedback Strategy", in Science Direct, Image and Vision Computing 24, pg 80-93, 2006.
[14] Alexander C Berg Tamara L Berg, Jitendra Malik titled "Shape Matching and Object Recognition Using Low Distortion Correspondences", in IEEE Proceedings, Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[15] Krystian Mikolajczyk, Cordelia Schmid titled "A Performance Evaluation of Local Descriptors", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, no, 10, October 2005.
[16] David G Lowe titled "Distinctive Image Features from Scale Invariant Keypoints", in International Journal of Computer Vision 60(2), pg 91-110, 2004.
[17] Joseph Sivic, Andrew Zisserman titled "Video Google: A Text Retrieval Approach to Object Matching in Videos", in IEEE Proceedings of Ninth International Conference on Computer Vision (ICCV), 2003.
[18] K-means: "home.deib.polimi/matteuce/clustering/tutorial_html/kmeans.html".
[19] Training data set and test data set: "http://stackoverflow.com/questions".
[20] "in.mathworks.com./help/vision/index.html".

