Determinant of homography-matrix-based multiple-object recognition

Nagachetan Bangalore¹, Madhu Kiran, Anil Suryaprakash

Visio Ingenii Limited
F2-F3 Maxet House
Liverpool Road
Luton, LU1 1RS
United Kingdom

ABSTRACT
Finding a given object in an image or a sequence of frames is one of the fundamental computer vision challenges. Humans can recognize a multitude of objects with little effort despite scale, lighting and perspective changes. A robust computer vision based object recognition system is achievable only if a considerable tolerance to change in scale, rotation and light is achieved. Partial occlusion tolerance is also of paramount importance in order to achieve robust object recognition in real-time applications. In this paper, we propose an effective method for recognizing a given object from a class of trained objects in the presence of partial occlusions and considerable variance in scale, rotation and lighting conditions. The proposed method can also identify the absence of a given object from the class of trained objects. Unlike the conventional methods for object recognition based on the key feature matches between the training image and a test image, the proposed algorithm utilizes a statistical measure from the homography transform based resultant matrix to determine an object match. The magnitude of determinant of the homography matrix obtained by the homography transform between the test image and the set of training images is used as a criterion to recognize the object contained in the test image. The magnitude of the determinant of homography matrix is found to be very near to zero (i.e. less than 0.005) and ranges between 0.05 and 1, for the out-of-class object and in-class objects respectively. Hence, an out-of-class object can also be identified by using low threshold criteria on the magnitude of the determinant obtained. The proposed method has been extensively tested on a huge database of objects containing about 100 similar and difficult objects to give positive results for both out-of-class and in-class object recognition scenarios. The overall system performance has been documented to be about 95% accurate for a varied range of testing scenarios.

Keywords: Feature matching, homography, scale invariance, occlusion tolerance, bag of visual words, RANSAC

INTRODUCTION
Computer assisted activities, involving object recognition have been widely used in applications such as Augmented Reality, retail industry for shopping, tourism, defense, education etc. All of these applications demand detection, recognition and localization from a large database of images. Time is also an important constraint in recognition based applications as the user expects the result in a short time. In the recent past, a number of feature detection methods have been developed for object recognition and localization.

When an image is sent out for recognition, one-on-one matching traversing through a large database of trained images takes place. This is a cumbersome and slow process, also vulnerable to additive noise contributed by various images in the database as well as at the user end while scanning for a test image.

¹ Send correspondence to Dr. Nagachetan Bangalore, Visio Ingenii Limited, United Kingdom: Email:nagachetan@visioingenii.com, Telephone: +44 (0) 1582 488763
Such conventional methods involve comparison of interest points in the training image with the query image by calculating the Euclidean distance between their descriptors [1, 2]. Nearest neighbor ratio matching strategy is applied in order to select the matching pairs [2, 3]. The image in the database that shows highest number of matches against the query image is classified as the recognized image. It has been observed from the experiments conducted by various conventional methods that the “number of matches” based recognition is vulnerable to noise and drastically reduces the percentage of true positives.

Hence to solve the above stated problems, a novel idea of determinant of homography based method has been presented in this paper. The rest of the paper has been divided into four sections. A literature review section describing existing feature detection and matching methods followed by the introduction is documented. Section II describes the proposed method. Section III presents the experimental results and analysis. Section IV is dedicated for conclusion and future work for the proposed method.

Section I : Related Research

Feature Detection and matching overview

Scale Invariant Feature Transform[1] and Speed Up Robust Features[2], both have proven to provide robust image matching across a substantial range of affine distortion, change in illumination and scale. These methods use local image features for object recognition. Each key point is given a stable location and orientation. In order to obtain robust descriptor for each key point, the scale of the key point is used to select the level of Gaussian blur. The gradient magnitude and orientation around the key points are then sampled. Thus an array of orientation histograms are created across to obtain 128 dimension.

Key point matching is the next important step in object recognition after key point localization and description. The best candidate match for each key point is found by identifying its nearest neighbors in the database. The nearest neighbor is found by finding the key point with minimum Euclidean distance for the invariant descriptor vector. In order to filter false matches, the distance between the closest nearest neighbor to second closest nearest neighbor is compared [1]. According to David Lowe's experiments the ratio of distance between the nearest neighbor and second nearest neighbor is 0.8 in order to clean up 90% of wrong matches.

For object recognition, the image under test is compared with each of the training image to find the one that gives maximum number of matches. The image that gives maximum number of key point matches is considered as the object under query.

Projective Relations and Homography Matrix

The fundamental matrix is the key concept in stereo vision. For a set of corresponding matches, in key point matching, the epipolar constraint is expressed as

\[ x^{T} F x = 0 \]  \hspace{1cm} (1)

Where F is the fundamental matrix, 3X3 singular matrix with 7 DOF (Degree of Freedom), x and x’ are the corresponding points of the matches.

A homography matrix is a projective transformation, represented as non-singular 3X3 matrix. If a point X in space is imaged in two views x in first and x’ in second, then

\[ x’ = Hx \]  \hspace{1cm} (2)

The matrix H has 8 DOF (Degree of Freedom) [4], therefore at least four point correspondences are needed in order to determine H(homography matrix) and 8 point correspondences are needed to determine homography using RANSAC
method. Homography estimation by RANSAC method is preferred as the image matches are contaminated with miss - matches and noise.

The basic RANSAC scheme
1) First the image matches are calculated.
2) Four random points are selected from the set of matches and homography is computed.
3) Select the pairs that agree with the homography. Pairs agreeing are the ones that satisfy
\[ d ( Hx ; x ') < t \]
\[ (3) \]
where \( t \) the threshold and \( d \) is the Euclidean distance.
4) Repeat until sufficient number of pairs are found
5) Re-compute homography with selected pairs.

Visual Bag of Words
It is an approach where, an image is treated as a collection of region, ignoring their appearance and spatial structure. This bag of key points method is based on vector quantization of affine invariant descriptors of image patches. It can be based on two alternative implementations using different classifiers: Naïve Bayes [7] and SVM (Support Vector Machines) [8]. Naïve Bayes is a simple classifier, usually used in text categorization. It is viewed as the maximum a posteriori for a generative model. The main advantages of the method are that it is simple, computationally efficient and intrinsically invariant. The results with SVM are clearly superior to those obtained with the simple Naïve Bayes classifier. With this method the largest ever classification has been achieved among images.

An SVM classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value and several attributes . The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. In SVM classification, the two - class data is separated with a hyper-plane with maximum margin [6]. For a given observation \( X \) and corresponding label \( Y \), the decision function is represented as
\[ f ( x ) = \text{sign} \left( \sum y_i a_i K ( x , x_i ) + b \right) \]
\[ (4) \]
Where \( x_i \) are the training features from the data-space \( X \) and \( y_i \) is the label of \( x_i \). Here the parameter \( a \) is typically zero for most \( i \).

The steps involved are [6]:
1) Selection of image patches and their descriptors.
2) A vector quantization algorithm is used to assign the patches of descriptors to predefined clusters.
3) Construction of a bag of key points, which counts the number of patches as-signed to each cluster
4) Applying a multi-class classifier, treating the bag of key points as the feature vector, and thus determine which category or categories to assign an image.

Section II : Determinant of Homography based Object recognition
In the classical method of object recognition, result is confirmed based on the number of matches
obtained during one to one image matching between query image and training image. But the disadvantage of this method is that, with increase in the number of training data set, the matching accuracy decreases. This is due to the fact that a few mismatched key points are still present.

The proposed methodology:
1) Find the homography matrix using the RANSAC method as mentioned in section I.
2) Find the determinant of homography and test how close it is to zero
3) If the determinant of homography is close to zero, the matrix is unstable and hence the matches can be ignored and total matches is assigned zero.
4) If determinant is greater than set threshold, the total number of matches can be considered.

Step 2 helps to determine the stability of the matrix [5], if the determinant of a homography is close to zero it corresponds to a degenerate case. Also, taking the advantage of parallel processing, it has been found experimentally that, the proposed method can recognize an image in 0.75 seconds with an accuracy of 98% among a database of 100 images.

The proposed method can be used in conjunction with any one of feature matching methods like SIFT, ORB [8] or SURF. ORB is reasonably robust and much faster compared to SIFT, hence ORB has been selected for feature matching.

If the training data set is of very large size such as greater 1000, Bag of key points method mentioned in Section I is used. This involves two steps.
1) Use bag of key points method to assign the category/categories to an image.
2) While querying, once categorized, do a one on one matching within the category and use determinant of homography method to arrive at the exact image.

The advantage of this method is that, both speed and accuracy of object recognition is tackled. The section that follows below gives the set of experimental results obtained.

**Section III Experimental Results**

The methodology of object recognition mentioned in section II is evaluated using a data-set of logos. The data-set consists of 100 images of 100 different logos. The feature detection method used for evaluation purpose is ORB. The query images used for testing are data-set of real world planar images captured from different viewpoints. The emphasis of the proposed method of object recognition is the accuracy of object recognition and scalability of data-set along with the latency in recognition.

Figure1. Query data-set examples
Figure 1 shows a set of sample query data set used. The images have undergone deformation and affine transformation to emphasize on the accuracy under different testing conditions. The proposed method of object recognition is compared with the classical method [2], which classifies the object based on total number of matches during one to one matching.

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Maximum number of key points set for detector</th>
<th>Size of Data-Set</th>
<th>Accuracy of object recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Method of feature match count</td>
<td>500</td>
<td>100</td>
<td>85.00%</td>
</tr>
<tr>
<td>Determinant of homography based method</td>
<td>500</td>
<td>100</td>
<td>98.00%</td>
</tr>
</tbody>
</table>

Table 1. Comparison of accuracy of object recognition between classical and Proposed methods.

Figure 2. Graph showing the increase in latency of object recognition with increase in size of training data set.

The algorithm has been implemented on a i5 processor based computer, with a memory of 4 GB RAM. It can observed from Figure 2 that the rate of increase in latency with increase in size of training data is fairly low.

**Conclusion**

The determinant of homography method is helpful for object recognition as it enables accurate object recognition from a large database of images. Such accuracy has been achieved by filtering false matches. Quality of homography matrix estimated from the matched pairs aids in the estimation of quality of matches. Parallel processing has considerably reduced the latency in recognition. By using the visual bag of words method, it has been possible to improve the scalability of object recognition in addition to maintaining the accuracy of the multiple object recognition system. Future work will aim at improvements in higher scalability and accuracy of the recognition system.
References


