Object Detection and Matching using Feature Classification

[1]Prinika Bhan, [2]Jagadeesh.B, [3]Geetha M.N
[1]BE, Student, Dept.of E&C
[2,3]Assistant professor, Dept.of E&C
[1, 2,3]Vidyavardhaka College of Engineering, Mysore, India
[1]prinikabhan1593@gmail.com
[2]jagadeesh.vvce@gmail.com
[3]geethashekar73@gmail.com

Abstract—**This paper considers the objective of efficient object detection and matching in images and videos. These objectives lead to the proposed classification scheme that classifies the extracted features in new images into object features and non-object features. It is shown out that this binary classification scheme has turned out to be an efficient tool that can be used for object detection and matching. This paper considers the objective of accurate matching and robustness. Due to this classification, the matching process becomes more robust and faster. In this case robust object registration also becomes fast. It shows the advantages of using classification stage for object matching and registration using the quantitative evaluation. This approach can be used for real-time object tracking and detection.**

Keywords—object detection; object matching; feature classification; object registration

I. INTRODUCTION

The proposed paper has the capability of detecting and registering objects in a video sequence captured by either a fixed camera or a moving camera, which is the corner stone in many computer vision applications. The camera can be a hand-held camera, a robotics camera, or an onboard camera. By the end of this paper, many challenging problems should be solved, namely object detection [1], [2], 3D object pose [3] , feature extraction and matching [4], and image registration [5]. The problem of object detection has been studied by many researchers. Supervised techniques have been used to detect objects whose class can be described statistically such as faces, facade windows, and vehicle rears based on learned appearances. Adaptive Boosting [2] and Active Appearance Models [6] are the two supervised techniques used in this paper. In many applications the objects of interest cannot be described by a generic model.

For example, tracking an arbitrary physical object cannot use the above techniques. Therefore, we use a reference model for this object. This model can be represented by a template or a set of relevant features (see Fig. 1). Input images are matched with the object template or features at run time in order to register the object with the current image. The kind of the registration depends on the shape of the object in consideration. If the object is planar, then the

registration aims to compute the homographic transform between a reference frame and the current frame or else if the object is 3D, then the registration aims to compute its 3D pose or projection matrix with respect to the camera [3], [7]. In all cases, a set of feature matches should be computed before carrying the registration process. A classical feature matching scheme can be used for establishing the matches. If the object at hand has a small size in the current captured images, then it may result in a difficult task. The problem is solved by detecting and matching the occurrence of a given object in a scene frame. If the number of feature points extracted in the scene frame is much higher than the number of feature points that lie in the reference set, then that again results in a problem. Thus, the computational load for establishing and pruning the matches will be very high.

In this paper, a classification of the image features is done based on object features and non-object features. To carry out this classification, a binary Support Vector Machine (SVM) classifier has been used. The training of the SVM classifier is carried out offline, while the classification of the extracted feature points in each video frame runs online and in real time. The proposed classification procedure scheme holds good for increasing the speed and reliability of feature matching. It is also helpful for reducing the computational load associated with the robust estimation of the registration parameters. Robust computation of the registration parameters (homography or 3D pose) is based on the robust technique RANSAC. If the percentage of false matches (outlier matches) is high, then its computational load will be too high. The rest of the paper is organized as follows. Section II states the problem we are focusing on. Section III presents the proposed approach. Section IV gives some experimental results.

II. PROBLEM STATEMENT

Given an object represented by one or more set of reference features, we would like to detect the occurrence of that object in a new image and if the detection is positive this object should be registered with the current new image via feature matching. Obviously, the associated feature matching is more challenging than stereo matching where the number of features in both images can be roughly the

same and where many simplifying geometric constraints can be used (epipolar geometry, small baseline, etc.). For example, the object is represented by about 18 feature points whereas the input image contains hundreds of extracted features. Thus feature matching between these two sets of features will be very difficult.
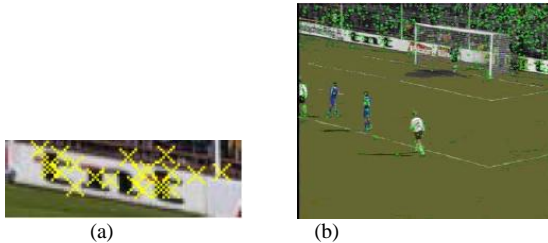

Fig. 1.   Matching the features in images **(a)** and **(b)** can be a difficult task (see text).
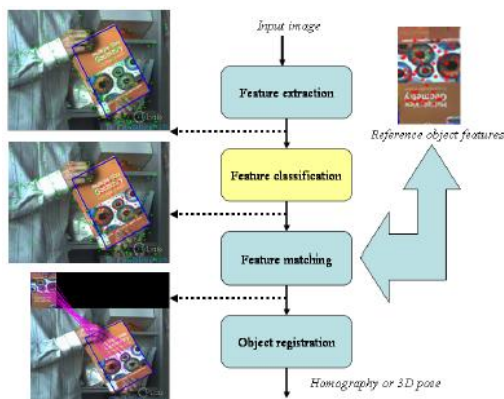
III.   PROPOSED APPROACH

Fig. 2.   The main stages of the proposed approach.

The proposed approach is summarized in Fig. 2 as follows: What differentiates this work from existing object matching and registration is the following: a pre-step is used for object feature selection which is carried out using a binary classification based on Support Vector Machines.

A.   Feature extraction

Feature points are extracted in the object reference template and in the input image. For this purpose, the SIFT (Scale Invariant Feature Transform) points [8] are used. SIFT points are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D rotation. Each SIFT point has a descriptor computed from a histogram of local oriented gradients around the key point and stored in a 128 dimensional vector.

B.   Feature classification

Before putting image SIFT points into correspondence with the object SIFT points, it will be advantageous to filter out the image SIFT points such that the majority of the retained image points belong to the object in question. Recall that the object features form a small subset of the input image features. This filtering procedure is very useful not only for increasing the reliability of feature matching

but also for reducing the computational load associated with subsequent processing. To this end, a binary classification scheme is used in which every extracted image feature is classified into one of these two classes: object class and non-object class. In this case, this classifier is built using Support Vector Machines with two classes of features [9]. The two-class SVM is built offline on two sets of training features. The positive training features (object features) are extracted in some given reference images of the object (real images and/or warped images). The negative training examples are composed of all feature points that are not belonging to the object (e.g., feature points belonging to the background). Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression. Input data are viewed as two sets of points in an n-dimensional space, an SVM will construct a separating hyper-plane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyper-planes are constructed, one on each side of the separating hyper-plane, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the neighboring data-points of both classes, since in general the larger the margin the lower the generalization error of the classifier [10]. SVMs were extended to classify data sets that are not linearly separable through the use of non-linear kernels. In this work, non-linear SVMs are used with polynomial kernels of degree two.

C.   Feature matching

At this stage, two sets of features are present. The first set is given by the reference features. The second set is given by the input image features that are classified as object features. In order to put the features in these two sets into correspondence, the first Nearest Neighbor classifier is used. To make the process fast, the KD-tree technique [11] is used. Moreover, a threshold is used in order to discard some matches. To register the object (homography or 3D pose computation) the RANSAC technique [12] is used. It is worth noting that the feature classification step introduced in this approach will make the matching process more robust and faster since most of the non-object extracted features are not handed over to the matching process. Moreover, the registration process based on the RANSAC technique will be faster since the outlier (false matches) percentage is decreased. Performance evaluations will be given in the next section.

IV. EXPERIMENTAL RESULTS

The proposed approach has been applied on several objects depicted in video sequences. Fig. 3 illustrates the application of the proposed approach on an input video frame. **(a)** The reference object features are shown in yellow crosses. **(b)** Illustrates the input video frames together with the extracted SIFT features which is shown in green. **(c)** Shows the features that are labeled as object points by the trained SVM which is shown in yellow crosses. **(d)** Shows the obtained matches after applying the KD-tree matching and the RANSAC technique. It has been observed that the classifier based on SVM provides few false positive features

(see Fig. 3**(c)**). The false positive features will be eliminated during the robust registration technique.

Fig. 4 illustrates the application of the classification stage on a 3D object. The proposed feature classification stage tells us if the current image contains or not an occurrence of the object at hand. To the end, ratio is threshold between the number of the detected positive features and the number of the reference object features. Fig. 5.**(a)** shows successful detections associated with a video sequence. Fig. 5.**(b)** illustrates an augmentation associated with one frame. Fig. 5.**(c)** shows the application of our proposed approach in the presence of significant illumination changes.

Tables I. (a) and I. (b) shows the performance of matching and registering of two different planar objects in 20 input frames. The right column of each table is used to show the performance without the use of the classification stage. The performance of second column is shown with the use of the classification stage. The introduced feature classification scheme has considerably reduced the matching CPU time as well as the percentage of false matches, which in turn has reduced the RANSAC CPU time.
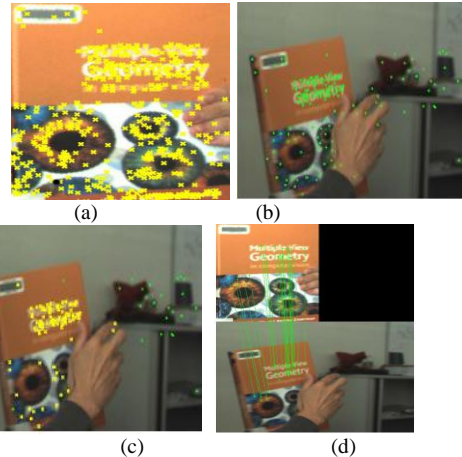

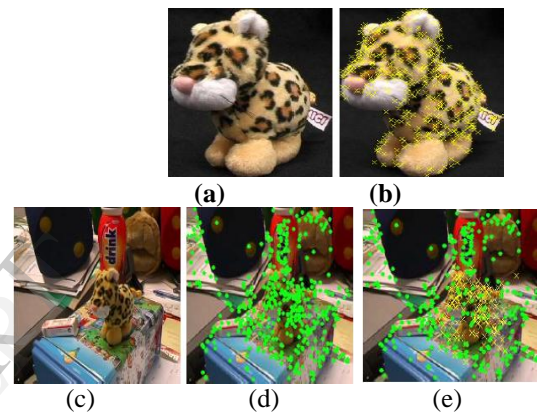Fig. 3.   Object matching and registration using the proposed approach (see text).

Fig. 4.   **(a)** and **(b)** depict the object and its reference feature points, respectively. **(c)** and **(d)** depict a given scene and the extracted features, respectively. **(e)** shows the features classified as object features (shown in yellow crosses).

	No class.	With class.
Reference object features	76	76
Extracted image features	509	509
Classified features (SMV)	-	172
Classification (CPU time)	-	1ms
False positive (SVM)	-	53
False negative (SVM)	-	7
Matched features (KD-tree)	66	32
KD-tree matching (CPU time)	141ms	35ms
Outliner percentage (RANSAC)	43%	12%

(A)   Book

	No class.	With class.
Reference object features	182	182
Extracted image features	686	686
Classified features (SMV)	-	209
Classification (CPU time)	-	1ms
False positive (SVM)	-	56
False negative (SVM)	-	3
Matched features (KD-tree)	109	64
KD-tree matching (CPU time)	173ms	40ms
Outliner percentage (RANSAC)	53%	33%

(B)   Board

TABLE.I: Comparing Object Matching With and Without the Pre-Step of Feature Classification. The Results Depict The Average Performance Over 20 Different Images of a Book (A) and of a Board (B).

V.   CONCLUSION

This paper presented an approach for detecting and matching an arbitrary object with an image. It has been shown that the use of feature classification makes the whole process of matching and registration faster and more robust.

REFERENCES

[1] M. Ozuysal, P. Fua, and V. Lepetit, "Fast keypoint recognition in ten lines of code," in *Computer Vision and Pattern Recognition*, 2007.

[2] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[3] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "Softposit: Simultaneous pose and correspondence determination," in *European Conference on Computer Vision*, 2002.

[4] E. Tola, P. Fua, and V. Lepetit, "A fast local descriptor for dense matching," in *Computer Vision and Pattern Recognition*, 2008.

[5] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[6] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–684, 2001.

[7] F. Dornaika, "Contributions `al'integration` vision/robotique: calibration, localisation et asservissment," Ph.D. dissertation, INRIA, 1995.

[8] D. Lowe, "Distinctive image features from scale invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–100, 2004.

[9] C.Cortes and V.Vapnick, "Support vectors networks," *Machine Learning*, vol. 20, 1995.

[10] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, pp. 169–186, 2003.

[11] M. de Berg, O. Cheong, M. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*. Springer, 2008.

[12] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communication ACM*, vol. 24, no. 6, pp. 381–395, 1981.