Vision Guided Pick and Place Robotic Arm System Based on SIFT

Girish G. Patil

Abstract— The work presents a complete vision guided robotic arm system for picking and placing of objects. For object recognition and localization purpose, the approach exploits Scale Invariant Feature Transform keypoint extraction to segment the correspondences between the object model and the image onto different potential object instances with real-time performance. Form the obtained correspondences, the best picking point is estimated. The coordinates of this point is then transferred to robotic arm coordinate system, which allows the arm to pick and place the object to the desired locations. The use of SIFT based clustering allows the system to be used for applications under extreme conditions of occlusion, where standard appearance-based approaches are likely to be ineffective. This system overcome most of the challenges occurred in actual workspace and in real time applications by presenting sufficient speed of operation. The system can handle complex ambient illumination conditions, challenging specular backgrounds, condition of occlusion, diffuse, and transparent as well as specular objects efficiently. The system can work with a single view of object, reducing the time and complexity to create database. The work presented is simple and fast for practical implementation as well as low cost. Many measures of efficacy and efficiency are provided on random disposals of objects, with a specific focus on real-time processing. Experimental results on different and challenging kinds of objects are reported using the custom designed robotic arm to demonstrate the effectiveness of the technique.
.

Index Terms— Object Recognition, Scale Invariant Feature Transform, Image Processing, Feature Extraction, Robotics.

1 INTRODUCTION

COMPUTER vision and pattern recognition techniques have been widely used in the past for industrial applications and especially for robot vision. Imaging a 2-years-old baby taking his favorite toy out of a box full of toys of different shapes, sizes, and colors; everyone can agree with this. This is an easy task for the baby. At first glance, everyone takes this for granted. Everyone is able to learn this in the first years of his life. Unsupervised recognition and localization of an unknown object in a cluttered environment is still an unsolved problem in the field of robot vision. Today robots get more and more involved in industrial processes, because they are superior to man regarding requirements on strength, speed and endurance. Robotic automation processes became very successful in the recent years, and offers a wide range for research. This process occurs in nearly every industrial sector and many of the household applications. Robotics has dealt with such tasks a very long time, but there are only few solutions suitable for special applications. In many fields of industry and household applications, indeed, there is the need to automate the pick-and-place process of picking up objects, possibly performing some tasks, and then placing down them on a different location..

Most of the pick and place systems consider the case of well separated objects, well aligned on the belt and allowing a synchronized grasping of the objects. In this case, simple photocells proved to be sufficient to initiate the picking phase However, there are several applications in which this approach will be insufficient, since forcing the objects to stay well separated and aligned on the working area will waste space and time of the process. Moreover, there can be objects which needs to and are convenient to be kept in bins, for saving time and/or for hygienic and other reasons. In this case, high resolution cameras should be used, together with specific machine vision algorithms [1].

In the case the objects are positioned at random over the working area. The objects are of different dimensions and appearances. A vision-based pick and place system presents several challenges:

1. It should be capable to work with every type of object of different dimension
2. Objects could be very complex, with many faces, reflective surfaces or they could be packaged in transparent flow packs
3. These applications often require very high working frequency,
4. In bin picking applications, objects are much cluttered and the occlusions make the object only partially visible

The proposed approach is meant to tackle all these points by proposing a feature-based segmentation technique capable to segment multiple occluded objects. When objects are complex, reflective, low-textured and heavily occluded, very few distinctive feature points can be extracted from the image. Having few features to be matched with the original sample, segmentation of the object is not straightforward. Thus, this work proposes the use of highly sophisticated software in order to overcome the challenges stated earlier. In the proposed approach the Scale Invariant Feature Trans-

- *Girish Patil is recently completed his masters degree program in Electronics System and Communication from Government College of Engineering, Amravati.E-mail: patilgirish213@gmail.com*

form [2] is used for computation of features required for object recognition purpose and also provides the localization with sufficient speed, efficiency and robustness. The robotic arm manipulator is designed such that it will be precise in its movements to pick and place the target object as fast and correctly as possible inspite the properties of the object.

2 Pick and Place System

The complete system is divided in two phases. i.e. Object Recognition and Robotics. Fig.1 shows the working of the proposed project. There are four important phases involved in the object recognition. In the first phase the user initiates the system by through user interface to facilitate the user to choose the target object by selecting the object image from database. In the Object Detection phase, the snapshot of the current scene is taken in real time, and then the objects in the current scene image are detected and localized with the help of SIFT based feature extraction and matching. Once the objects are detected, and then the target object is selected among all the other objects present in the current scene based on feature mapping in the third phase of object selection. The fourth phase i.e. Object Localization exploits camera calibration to obtain 2D coordinates of the grasping points of the selected object. In particular, standard camera calibration is employed to transform the 2D image coordinates obtained in the first phase into 2D world coordinates i.e. pixel to cm conversion. The coordinates obtained from the object recognition process are provided to the microcontroller which directs robotic arm manipulator for precise movements of all its arms in order to pick up the selected object and place down to the user intended space in Object Pick and Place phase [3].

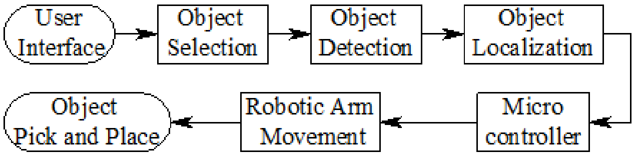

Fig. 1: Diagram of Pick and Place System

3 Object recognition and localization

- Object recognition : Here, the number of significant SIFT features are extracted from the target object image as well as from the scene image; then by applying a proper similarity measure, these features are matched and the best matched pairs of features are kept for next phase;
- Object localization: From the best matched pairs of features, by applying appropriate mathematical formulation, the location of center point of the detected object in the current image is calculated.

3.1 Object Recognition

In offline process, the image of the object is taken from a single top view. Then the SIFT technique [4] is applied on this image for the extraction of possible features and then it is followed by the calculation of the unique descriptor containing the location of feature, scale and orientation. These descriptors for that particular object are saved in database. The process is shown in Fig. 2. In online process, the image of the scene is captured from where the object is to be recognized and picked up. The same procedure is followed for the scene image and descriptors for the scene image are calculated. Now, these descriptors are matched with the descriptors of the called object. Among many methods for local feature extraction and model matching, here the SIFT and Two Nearest Neighbor (2NN) is selected as SIFT has proved to be very robust to noise and invariant to scaling, rotation, translation and (at some extent) illumination changes as well as compatible for real time applications and 2NN algorithm is used to form the clusters of the matching features from object image and the scene image.

The SIFT can be used to obtain the set of keypoints K_M for model M and keypoints K_I for current scene image I,

$$K_M = \{k_i \triangleq [x_i^M, y_i^M, \theta_i^M, D_i^M], \quad i = 1,2, \dots p\}$$
$$K_I = \{k_j \triangleq [x_j^I, y_j^I, \theta_j^I, D_j^I], \quad j = 1,2, \dots q\}$$

Where,

x ,y - The 2D image coordinates, .
θ - Main orientation computed
D - 128 -value SIFT descriptor.

From the two sets K_M and K_I, the standard 2NN algorithm is applied for the computation of Euclidean distance between D_i^M and D_j^I to determine the corresponding model to image matches $M = (m_1, m_2, \dots, m_N)$, where each match m_q contains the (x, y) coordinates on the two reference systems and the main orientation on the current image $m_q = \{(x_j^I, y_j^I, \theta_j^I), (x_i^M, y_i^M)\}$.

For the derived set M, the suitable and simplest approach for evaluating the registration transform between model*(M)*and image *(I)* is to estimate the planar homoFig.y using a least squares approach.

3.1 Localization of Center Point

In the implementation of given approach by SIFT, after the matching process of descriptors, the (X, Y) coordinates of all the matched descriptor in scene image is obtained. The total no. of descriptors may vary but they will be definitely situated on the object, around its center point. Here, the above said formula is used individually on X-coordinates and Y-coordinates to give the centroid among the known positions of the available descriptors. Let, $X = (X_1, X_{2,\dots\dots\dots,} X_n)$ and $Y = (Y_1, Y_{2,\dots\dots\dots,} Y_n)$, be the (X, Y) coordinates of the matched descriptors, therefore the (X, Y) coordinate of the center point is estimated as

$$X_{center} = (max(X) + min(X))/2$$
$$Y_{center} = (max(Y) + min(Y))/2$$

The coordinates of center point of the object in pixel is given as

$$(X, Y)_{center} = (X_{center}, Y_{center})$$

The accuracy of this approach is tested and verified for no. of objects and various scenes through rigorous experimentation.

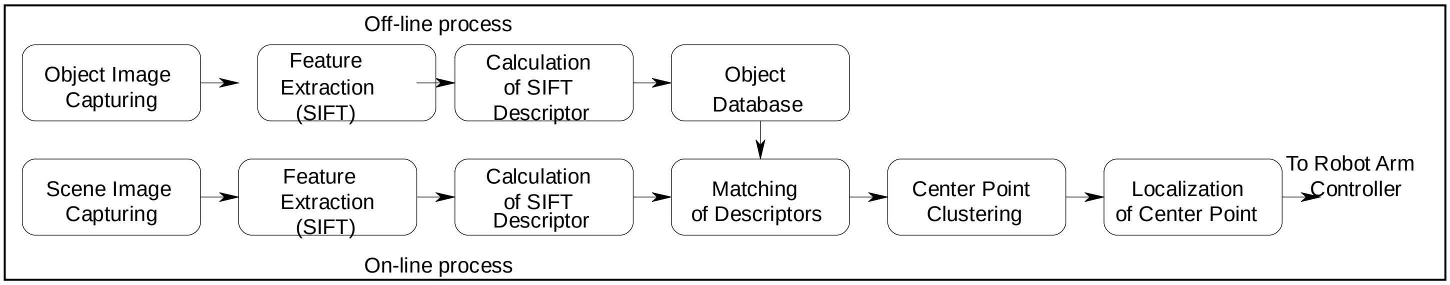

Fig. 2: Object Recognition and Localization

4 Geometry of Robotics Arm Used

OWI's second-generation robotic arm kit, the Robotic Arm Edge is used to demonstrate the pick and place system which utilizes the basic mechanics of robot arm construction and control. The proposed robotic arm manipulator consists of mainly base, shoulder, elbow, wrist and gripper. All the basic parts and the motor assembly construction with dimension is as shown in Fig. 3. This arm is analogous to human arm. Here, the base provides the horizontal rotation movement while shoulder, elbow, and wrist provide the vertical movements to reach to target position. This arm has four joints and three links and one gripper. This arm has five degree of freedom.

.

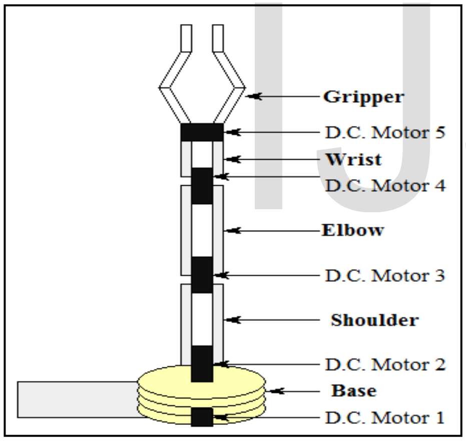

Fig. 3: Geometry of Robot Arm Manipulator

The Table 1 shows the description of the basic parts movements in order of their construction. Here, the D.C. motors with gearbox arrangement are used as actuators for each joint. The base provides the clockwise/Anti- clockwise rotation of the arm. Then shoulder, elbow and wrist provides the required up/down movement of the arm in the vertical direction while the end effector *i.e.* gripper provides the open/close movements for correct gripping of the object. Using five motors with gearboxes, the arm has five degrees of freedom: The base allows the rotation of 270°, a 180° shoulder motion, a 300° elbow motion, a 120° wrist motion, and a 0-1.77" (0-4.5cm) gripping motion. Table 1 helps to understand this degree of freedom. It shows rotation movements of the robotic arm with respective degrees.

The material used for the construction of the robotic arm is the polycarbonate. This material is used because it is very sturdy, robust and designing of the required arm structure has become a little simpler task due to flexibility of the material [5].

TABLE 1
Basic Arm Movements

Part	*Movement*	*Degree of Freedom*
Gripper	Open/Close	-
Wrist	Up/Down	120^{o}
Elbow	Up/Down	300^{o}
Shoulder	Up/Down	180^{o}
Base	Clockwise/Anti-clockwise	270^{o}

5 Performance Analysis

The object recognition and localization for pick and place operations involves two phases *i.e.* object recognition and actual pick and place operation. In this session overall system performance is analyzed to measure the feasibility and effectiveness of the proposed approach.

5.1 Database Creation

In the given approach the database of the object images is created in a very simple manner. In this process, the snapshot of the model object is to be taken from the different views for the correct match with scene image. In this approach a single snapshot from only one view suffices the condition of correct matching. Therefore, the single snapshots of the objects are taken. Here, five objects are considered *i.e.* battery, soldering paste box, match box, card reader and balm box. The reason to choose such objects is that these objects are very common household objects having different sizes and shape with reflective and transparent surface portion. Moreover, as these objects found so easily so a non-technical person can also check the feasibility of the project with such objects. The single images of these objects are captured from camera against the black background by placing them on the same working area. The images are stored in one folder in PC with the desired notation having uniform size of 320 X 240 pixels. The images of different size can also be used but it will take more time for recognition. The images of the different objects are presented in Fig. 4.

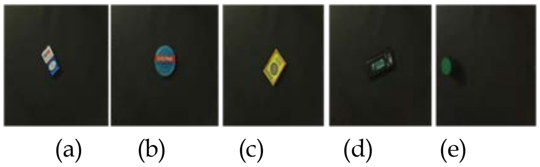

Fig. 4: Snapshots of The Database Objects (a) Battery (b) Soldering paste box (c)Match Box (d) Card reader (e) Balm box

5.2 Current Scene Images Creation

In ordere to test the ability of the object recognition and localization software, the experiments are carried out by taking the scene images in diffetent situation such as illuminition variation; occlusion, rotation of objects with different degrees and orientations. Here, different types of objects are chosen so as to test the feasibilty of program to its extent. Here, the first three scenes are taken to show the robustness against the change in illumination.

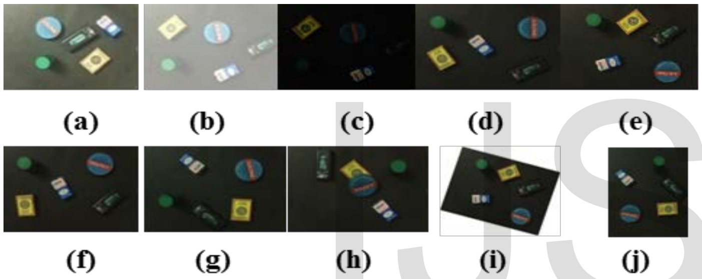

Fig. 5: Snapshots of The Current Scenes Used in Experiments

Fig. 5 (a) shows the more brightness than that of the object image brightness whereas the image in Fig. 5 (b) shows the higher brightness than the first one, in opposite to that Fig. 5 (c) shows the darker version of the same image. The Fig. 5 (d), (e), (f) and (g) shows the variation in the orientation of the object with reflective and transperent surface portion explored. Fig. 5 (h) shows the occlusion of the objects. In Fig. 5 (g) the complete scene image is rotated by 35^{o} while in Fig. 5 (h) the scene image is rotated by 90o with the vertical flipping of the image.

5.3 Experimental Result for Object- Battery

As the overall operation consists of two broad areas *i.e.* object recognition and actual pick and place, the Table 2 provides all the information about time taken for execution various steps involved in complete operation. The details of timing information regarding time taken for descriptor calculation, matching of images and plotting the results included in time required for object recognition where, it can be observed that most of the time is taken for descriptor calculation. The pick and place timing includes the picking and placing timing individually where the picking phase has taken slightly more time than the placing phase.

Table 2 indicates the time taken by robotic arm for actual picking operation and placing operation for the object battery. This table shows that, the robotic arm will take more time for operation if the object is placed away from the origin (top left corner). The Fig. 6(a) shows the same information with the help of bar chart where it can be clearly observed that in scene 8 the battery is placed at more distance from the origin, hence taking more time for the operation because of the occlusion in scene.

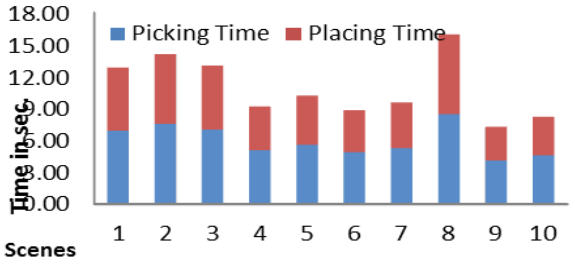

Fig. 6 (a): Timings for Picking and Placing operation for Object- Battery

Now, the overall operation *i.e.* objects recognition and actual pick and place operation timings are shown in Table 2. Here, the overall operation takes place around quarter a minute. So, it can be concluded that the proposed approach can be suitable for real time household as well as industrial application.

The Fig. 6(b) presents the timing analysis of the overall approach in logical manner. Where it can be observed that the maximum time from overall operation is consumed by the picking and placing phase time than object recognition process. The maximum time observed is 19.77 sec. The peak at scene 8 depicts this information.

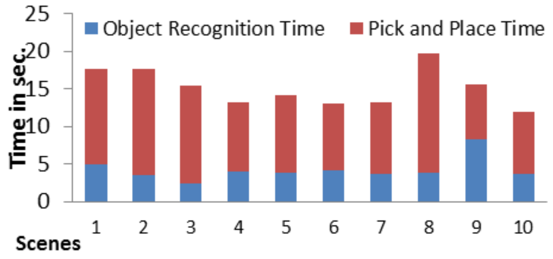

Fig.6 (b): Time Required for Overall Operation for Object- Battery

5.4 Performance of Robotic Arm

For the accurate and effective performance of the complete system along with the accuracy of object recognition phase, the accuracy of pick and place hardware plays important role. In order to evaluate the performance of the robotic arm, the hardware is ordered to move to one specific location again and again. In this procedure, the deviation between the desired location and achieved location is measured in terms of Euclidian distance between these location.

TABLE 2
TIMING INFORMATION FOR OBJECT- BATTERY

	Object recognition				*Pick and Place*			
Scene	**Descriptor calculation**	**Matching**	**Plotting**	**Object recognition**	**Picking**	**Placing**	**Pick and Place**	**Complete System Operation**
	(sec)	**(sec)**	**(sec)**	**(sec)**	**(sec)**	**(sec)**	**(sec)**	**(sec)**
Scene1	4.6984	0.0228	0.1305	4.8537	6.902	5.952	12.854	17.708
Scene2	3.3921	0.0052	0.0983	3.4969	7.53	6.58	14.110	17.607
Scene3	2.1864	0.0019	0.1244	2.316	6.99	6.04	13.030	15.346
Scene4	3.8824	0.0013	0.1161	3.9999	5.06	4.11	9.170	13.170
Scene5	3.7299	0.001	0.1601	3.891	5.58	4.63	10.210	14.101
Scene6	4.0547	0.0012	0.0988	4.1548	4.89	3.94	8.830	12.985
Scene7	3.5432	0.001	0.129	3.6733	5.25	4.3	9.550	13.223
Scene8	3.6607	0.0009	0.1392	3.8008	8.46	7.51	15.970	19.771
Scene9	8.1344	0.0018	0.1607	8.297	4.11	3.16	7.270	15.567
Scene10	3.5319	0.0011	0.1415	3.6746	4.58	3.63	8.210	11.885

Table 3 shows the details of the experiment. Here, total 12 iteration are taken for one specific location, for this case the center point of workspace (13cm, 10cm) is taken as test location and the achieved location by the arm is noted for each iteration. The average Euclidian distance is obtained as 1.07 cm. This shows that the arm performs quite good to reach the desired location with a maximum error of 1.07cm.

TABLE 3
PERFORMANCE OF ROBOTIC ARM HARDWARE

	Sr. No.	*X-loc. (cm)*	*Y-loc. (cm)*	*Euclidian Distance (cm)*
Test Location		**13**	**10**	
Achieved Locations	1.	13.5	8.5	1.26
	2.	13	10	0.00
	3.	12.5	8	1.44
	4.	11.5	11	1.34
	5.	11.5	9	1.34
	6.	13	12	1.41
	7.	13.5	10.5	0.84
	8.	14	8.5	1.34
	9.	13.5	8.5	1.26
	10.	12.5	12	1.44
	11.	14	9	1.19
	12.	13	10	0.00
			Average	**1.07**

Fig. 7 shows the various locations achieved by the arm for the test location (13cm, 10cm). It can be observed that the achieved locations are within the area of objects. Therefore, there is maximum probability of reaching the gripper to the objects. For two iterations, the arm has reached to the exact test location**.**

.

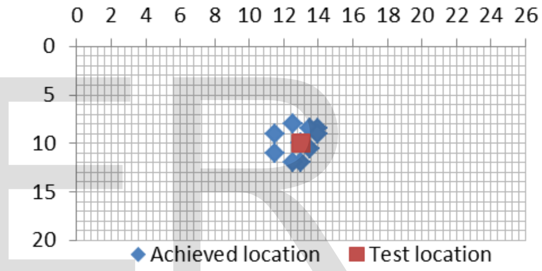

Fig. 7: Locations Achieved by Robotic Arm

5.5 Comparison of Matching Ratio

The next Table 4 shows the information about no. of features that need to be matched correctly without false matching and actual no. of features that are matched correctly. Ideally, all features of the object should be matched to features of that particular object placed in scene image. In this approach only those features are kept which gives the correct matches only. The ratio should be tending to 1. Here, the ratio achieves the desired value for scene 2 and 7.

Table 4 depicts the accuracy of the given approach. Where the pixel position of actual center point of the object in the given scene is compared against the pixel position of the center point estimated by the proposed approach and the difference between them is shown as error.

Fig. 8 shows the variation of this ratio for various scenes. From the peaks, it can be directly observed that for scene 2 and 7, the no. of features actually matched between the scene and object battery are exactly the same as that of no. of features to be actually matched.

TABLE 4
FEATURE MATCHING DETAILS FOR OBJECT –BATTERY

Scene	*No. of Features to be matched correctly*	*No. of actual features matched correctly*	*Matching Ratio*
Scene1	5	3	0.60
Scene2	2	2	1.00
Scene3	7	6	0.86
Scene4	28	17	0.61
Scene5	14	8	0.57
Scene6	17	11	0.65
Scene7	3	3	1.00
Scene8	18	11	0.61
Scene9	4	3	0.75
Scene10	10	7	0.70

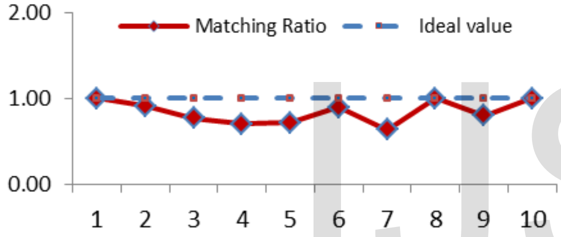

Fig. 8 Variation of Matching Ratio with All Scene for Object Battery

5.6 Comparison of Centre Distance

Table 5 depicts the accuracy of the given approach. Where the pixel position of actual center point of the object in the given scene is compared against the pixel position of the center point estimated by the proposed approach and the difference between them is shown as error.

TABLE 5
DIFFERENCE BETWEEN ACTUAL AND ESTIMATED POSITION OF CENTER POINT FOR OBJECT BATTERY

Scene	*Estimated Position of Center point (in Pixel)*		*Actual Position of Center point (in Pixel)*		*Euclidean Distance (in Pixel)*
	X1	**Y1**	**X2**	**Y2**	**D**
Scene1	253	63	245	60	2
Scene2	199	193	201	190	3
Scene3	199	193	201	190	3
Scene4	146	115	144	115	2
Scene5	201	192	200	193	2
Scene6	139	120	139	116	3
Scene7	111	50	107	47	3
Scene8	224	180	221	185	1
Scene9	128	191	128	190	4
Scene10	50	105	49	107	2

Table 5 depicts the accuracy of the given approach for object match box. The maximum distance is observed for scene 9. This is because one feature is wrongly matched with the feature of other object giving the large deviation in estimated center point.

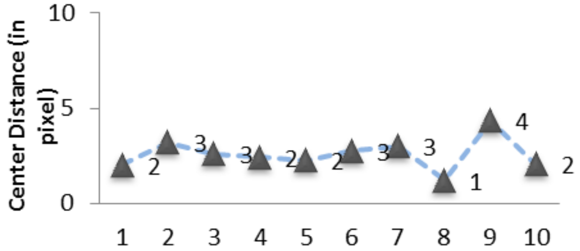

Fig.9: Scenewise Variation of Centre Distance (in pixel) for Object Battery

Fig. 9 shows the Fig. of Euclidian distance between actual center of object and the estimated center of the object. It shows the proposed approach gives very less distance between the actual and estimated center point except for scene 9. Yet, these are the satisfactory results.

5.7 Justification of Errors

Ideally, the value of matching ratio should be 1 but in experiments, it is found that this ratio achieved the desired value in some cases and in most of the cases it has achieved the value very close to 1. This is because of the threshold value, the amount of scene degradation and change in orientation of the target object. From the various experimentation, the threshold value is selected as 0.80. The matching ratio varies with the changes in orientation of the object in the scene. More is the change in object orientation, less the matching ratio. Moreover, the illumination and occlusion also affects the matching ratio. But, upto certain extent the illumination does not affect much on the recognition. Inspite of these problems, the no. of features matched proved sufficient for object recognition and localization.

To provide the accurate localization of the target object (*i.e.* centre point location) is the main aim of the system. The maximum difference obtained between the actual centre point and calculated centre point is about 7 pixels and mean difference is only 3 pixels. This is the result of locations of the features are situated in various parts of object, not necessary to be symmetric around the centre point; hence formula for centre point estimation causes the difference. As compared to size of image (320 X 240) pixels, this difference is negligible. And if converted to centimetres, this difference seems less than half cm, therefore the probability of estimated centre point to be on the object becomes large.

As the DC motors are used as actuators for the arm movements, the movements of arm depends on the amount of current applied to the DC motor windings. The minor deflection in current causes deviation in arm movements resulting in degradation in accuracy of the arm to reach for the target point. The mean distance between the test location and achieved location is found as 1.07cm, which again not a significant distance as compared to total size of workspace.

6 Conclusions

In this work, pick and place robotic arm using a computer vision based on Scale Invariant Feature Transform based clustering for picking and placing of objects is developed. This system utilizes a simple sensor consisting of a camera for object sensing. SIFT features are used for the correct segmentation of the scenes for recognition of object. It is shown that a single image from single view is sufficient to solve the purpose. The SIFT technique met the various challenges in real time applications such as clutter, scale changes, occlusion, illumination change, operating speed, etc. and it is proved the best one overcoming most of the challenges. Thus, the SIFT method used in the proposed approach as a powerful tool for the enhancing the capability of machine vision to develop an autonomous and robust pick and place robotic system.

From the various experiments it can be concluded that, the approach proved efficient to extent in terms of robustness, ability to handle objects of different shapes/size and reflectance properties including specular, diffuse and textured objects, as demonstrated by several real examples using the complete setup of the system. This system can also handle harsh environmental conditions such as non-uniform backgrounds and complex ambient illumination. The proposed technique is simple, low cost, fast and generic enough to accommodate variations in household and industrial automation applications.

Future Scope

As future directions, there is need to evaluate different detection strategies for texture-less objects as well as studying algorithms to detect whether an object is occluded especially in its picking point or not and implementing strategies for ordering the detected and pickable objects to minimize the traveling time of the robot. The emphasis will be given on minimizing the time required for the object recognition and localization by implementing the simultaneous working for processes of object recognition and robotic arm movements. Further, on the manufacturing of more robust, more sturdy and fast operating robotic arm. To make the system that can work really very fast and literally called as real time operating pick and place system.

References

[1] P. Piccinini, A. Prati, R. Cucchiara, “*Real-time object detection and localization with sift-based clustering*”, in proc. of Image and Vision Computing, 2012, pp. 1–15.

[2] D. Lowe, “ *Distinctive image features from scale-invariant keypoints*”, Int. J. Comput. Vis., 2004, pp. 91–110.

[3] G. Patil, D. Chaudhari, “*SIFT Based Approach: Object Recognition and Localization for Pick-and-Place System*”, Int. J. Adv. Research in Comp. Sci. and Software Engg., 2013, pp. 196–201.

[4] P. Piccinini, A. Prati, R. Cucchiara, “*Multiple object segmentation for pick-and-place applications*”, In: Proceedings of IAPR Conference on Machine Vision Applications (IAPRMVA 2009), Yokohama , Japan , 2009, pp. 361–366.

[5] http://www.owirobot.com/robotic-arm-edge-1

Author Profile

Girish Patil obtained diploma in Electronics and Communication Engineering from Government Polytechnic, Amravati in 2007, graduation in Electronics and Telecommunication Engineering from S.G.B. Amravati University, Amravati in 2010. Recently completed post-graduation in Electronic System and Communication from Government College of Engineering, Amravati.