Saliency Based Fast Object Localization and Recognition for Mechanical Assembly

Sudipta Bhuyan[1] Debashis Sen[2] and Sankha Deb[3]

Abstract— **This paper presents a fast object localization and recognition approach for mechanical assembly based on detection of multiple salient regions. Our novel multiple salient object detection technique employs locally adaptive center-surround operations and proto-object partitioning. The proposed approach is implemented in a vision guided industrial robot workspace to perform assembly involving different types of components. Effectiveness and utility of the proposed approach in robotic assembly is demonstrated through different measures obtained from multiple experiments. Detection and recognition rates, angular and positional errors, and computation time of our implementation are found to be better than when multiple salient object detection is not considered.**

I. INTRODUCTION

In today′s era, manufacturing automation has attracted a lot of focus due to rapidly changing customer requirements. Use of robot has brought a revolution in industrial automation. In manufacturing industry, robots are being deployed to perform a variety of tasks like pick and place, assembly, inspection, and material sorting, with high precision and accurate repeatability. Traditionally, all these activities were performed by humans. Although in many cases involving active tasks like detection and recognition, humans can perform better than machines, manual efficiency decreases with time.

Traditional mass production systems do not require detection and recognition because robots operate in a very structured environment. However, manufacturing automation is rapidly heading towards facilitating customizable mass production systems that allow flexibility in product design and diversity in material handling [1]. Therefore, it is very much desirable to incorporate an industrial robot with the ability to operate in an unstructured environment where it may encounter diversified products at random positions. It is challenging to program a robot to detect and recognize each element present in such an environment filled with uncertainties. Along with the accuracy, the speed of detection and recognition is also a significant factor. For a robot to operate in unstructured manufacturing environments, the first and foremost requirement is to detect each object present quickly and consistently /reliably. To do this, the machine should have sensory capabilities as humans, among which, vision is vital. The robotic system with vision capabilities should incorporate the human-like fast detection approach, where instead of searching each location in a scene /image, regions are attended to in a selective manner [2], speeding up the detection and localization process significantly [3].

To accomplish this, highly advanced and sophisticated machine vision techniques are required in many cases. Literature on vision guided robot manipulation for different tasks in manufacturing environment has predominantly considered object recognition through various matching approaches. The work of [4] and [5] are based on fitting /matching CAD models of objects onto those in an image. [5] used Chamfer matching along with CAD model for object position and orientation estimation. In [6], the matching between object templates and regions in input image for recognition is based on edge contour computation. [7] has done neural network based recognition based on Fourier descriptors and geometric moments as object features, which are used for matching. [8] used neural network for learning and recognition, where the Feature vector used for matching is obtained from object′s contour, type of curvature or topographical surface information, and depth information. Template matching based Pose estimation of objects having circular features is done in [9]. Object recognition by matching SIFT (Scale-Invariant Feature Transform) features of template image and query image is proposed in [10], [11]. Similarly, SURF (Speeded up robust feature) based matching is done in [12] and [13] for pose estimation of different objects.

After object detection and recognition, a machine vision guided robotic manipulator can automate the assembly of constituent objects. A comprehensive survey of such vision system in industrial environment is given in [14].

In the above examples of recognition through matching, the operation is performed on the whole image while searching for particular objects and most of them considering some representation of object shape. Humans through a mechanism called visual attention, perform recognition more efficiently [2]. In the mechanism, only the informative parts in an image /scene are selected for further processing [3], which in the context of recognition, is the matching operation. In computer vision, salient object detection approaches are used to select the informative parts [15]. The computation involved in detecting salient objects and then performing matching can be significantly less compared to searching for an object in the whole image.

Most salient object detection approaches are highly parametric and significantly sensitive to the parameter values [16]. A challenge in detecting multiple salient objects is to appropriately set the parameters involved. If we consider

[1]S. Bhuyan is with the Department of Advanced Technology Development Center. sudiptabhuyan1@gmail.com

[2]D. Sen is with the Department of Electronics and Electrical Communication Engineering. dsen@ece.iitkgp.ernet.in

[3]S. Deb is with the Department of Mechanical Engineering. sankha.deb@mech.iitkgp.ernet.in Indian Institute of Technology Kharagpur, West Bengal 721301, India

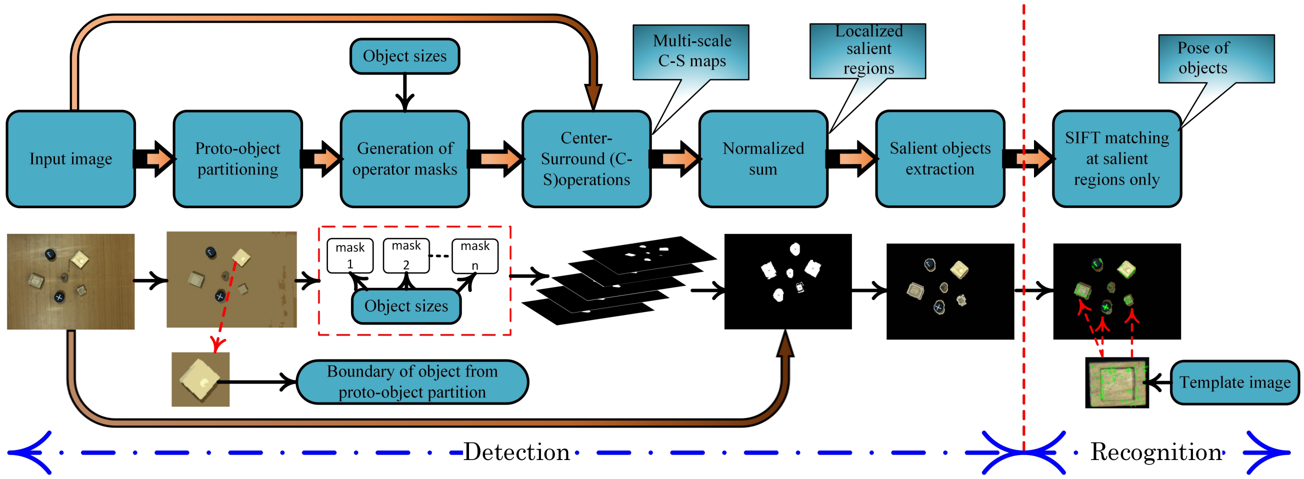

Fig. 1. Proposed approach for detecting multiple salient regions

the task of mechanical assembly by a vision-guided robot, it is essential that machine vision should detect all objects to be processed as salient. A second challenge, which will have significant consequences during automated assembly, is accurate detection of object boundaries and hence its proper localization.

The current paper presents a novel local structure aware method to detect multiple salient objects in order to perform fast object detection and localization for automated mechanical assembly. Once objects are localized, a feature point matching (or any other) recognition technique can be adopted, which will be applied only at the detected object regions thus making the recognition process significantly faster, as the total area of all the salient regions is always much smaller than the entire image. Parameters of our proposed method are based on the objects involved whose image templates are considered available. Our approach estimates a desired object′s position and orientation. The detection of multiple salient objects along with their boundaries is based on generation of center-surround operator masks adapted to the local structure obtained through proto-object partitioning. Recognition of each salient object is then done by feature point matching technique applied only at the detected salient object regions. To demonstrate performance in terms of angular and positional accuracies, and detection and recognition rates, an industrial robot manipulator is deployed for performing mechanical assembly. Our saliency based system is found to be significantly more efficient than that not based on saliency, without compromising on the performance.

II. THE PROPOSED APPROACH FOR DETECTING MULTIPLE SALIENT OBJECTS FOR ASSEMBLY

The framework of our application-specific and locally adaptive multiple salient object detection approach is shown in Fig.1. First, to make our approach local structure aware, we perform proto-object partitioning considering homogeneity based on color similarity (RGB vector). We employ the mean shift clustering algorithm [17] for the purpose, which is known to partition appropriately at object boundaries. By using the image partitions and knowledge of object sizes, operation masks adapted to local region around pixels are generated. Multi-scale center-surround operations are performed around each input image pixel using the corresponding generated mask. A map of salient objects for an image is generated through normalized addition of the maps obtained considering all the scales. Multiple salient regions are detected with object boundaries, where subsequently recognition is performed.

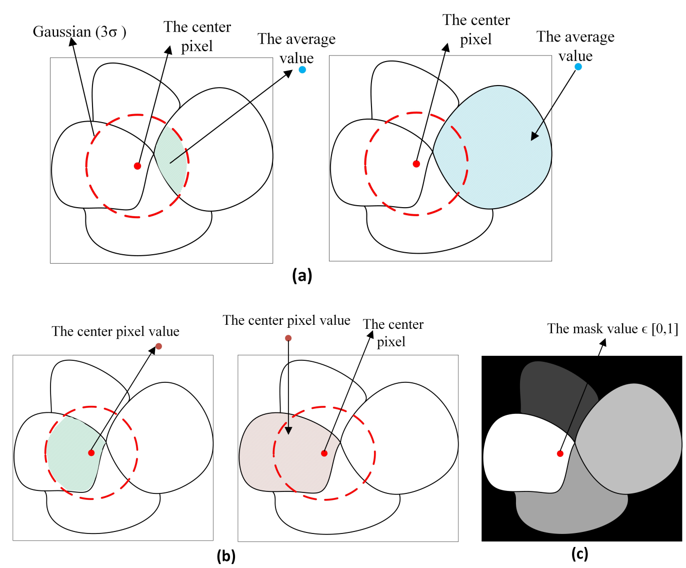

Fig. 2. (a) Assignment of mask values to a proto-object partition. (b) Assignment of mask value to the center partition. (c) The generated mask value $\in$ [0,1]

A. *Generation of operator mask for object detection*

Here, we describe the process of generating application specific center-surround operator masks adapted to local image structure. First, the sizes (diameter / side length) of the objects involved in a specific application are considered. Different circular Gaussian functions with standard deviations

(σ) one-third of the object sizes are generated. A mask is generated from such a circular Gaussian function as follows:

- Consider such a Gaussian function centered at a particular image pixel (center pixel) [See Fig.2(a)]. Consider all proto-object partitions having pixels within 3σ extent of the Gaussian function. Mathematically, we have:
  Image:$I(x,y)$, where $(x,y) \in L$, the set of all pixels.
  N proto-object partitions: $P_1, P_2, .., P_N$ such that

$$P_i \cap P_j = \emptyset \quad \forall\ i,j \in \{1,2,...,N\}, i \neq j \text{ and}$$
$$\bigcup_{i=1}^{N} P_i = L \tag{1}$$

  Let (x_c,y_c) be the center pixel. For a standard deviation σ, 2D circular Gaussian function centered at (x_c,y_c) is given by

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-\left((x-x_c)^2-(y-y_c)^2\right)}{2\sigma^2}} \quad \text{where } (x,y) \in L \tag{2}$$

  Pixels present within 3σ extent of the Gaussian function are:

$$P(x_c,y_c) = \{P_i, P_i \cap E(x_c,y_c) \neq \emptyset\} \forall i \in \{1,2,...,N\} \tag{3}$$
$$\text{where } E(x_c,y_c) = \{(x,y), \forall (x,y) \in L \text{ and}$$
$$(x-x_c)^2 + (y-y_c)^2 \leq 9\sigma^2\}$$

- To get the mask, all pixels in such a partition are assigned the average value from the part of the Gaussian function (upto 3σ extent) intersecting with it [See Fig.2(a)]. Assigning a single value to all the pixels in a partition is based on the notion that elements within a partition are indistinguishable.
  Set of pixels in the intersection region of Gaussian function and a proto-object partition P_i.

$$A_i(x_c,y_c) = \{(x,y), (x,y) \in \{P_i \cap E(x_c,y_c)\}\} \ \forall\ i \tag{4}$$

  Then the required average value is

$$V_{avg}(i) = \frac{1}{|A_i(x_c,y_c)|} \sum_{\forall (x,y) \in A_i(x_c,y_c)} G(x,y) \tag{5}$$

- However, if such a partition contains the center pixel (center partition), the value of the Gaussian function at the center pixel is assigned to all pixels in that partition [See Fig.2(b)]. Here, the center pixel value is considered instead of the average, so that we have a larger difference between the center partition and the neighboring ones yielding richer representation of local structure.
  Let P_{i0} be the center partition, that is, $P_{i0} \in P(x_c,y_c)$ and $(x_c,y_c) \in P_{i0}$
  So, we get the mask as follows:

$$Mask_{i0,(x_c,y_c)}(x,y) = G(x_c,y_c), \forall (x,y) \in P_{i0} \tag{6}$$

$$Mask_{i,(x_c,y_c)}(x,y) = V_{avg}(i), \forall (x,y) \in P_i,$$
$$\forall P_i \in P(x_c,y_c) \text{ and } i \neq i_0 \tag{7}$$

$$Mask_{i,(x_c,y_c)}(x,y) = 0 \ \forall\ P_i \notin P(x_c,y_c) \tag{8}$$

- A typical mask formed following the procedures mentioned in the above three steps is shown in Fig.2(c). A mask created as shown in Fig.2(c) does not consider the spatial extent of a partition, as emphasis is given on color indiscernibility among partition elements. It is well known that spatial distance of image contents from center pixel plays an important role in determining whether the pixel is informative (salient) or not [18]. A neighboring partition with a larger spatial extent will contain pixels farther from the center pixel. Therefore, farther the spatial extent of a neighboring partition from the center pixel, the value of the mask for that partition is multiplied with a smaller weight. This inverse relation is invoked as follows:
  Let the leftmost and rightmost values of x and y, where (x,y) is a pixel, for a partition P_i be $x_l^{P_i}$, $x_r^{P_i}$, $y_l^{P_i}$, $y_r^{P_i}$.
  Let the extent measure of a partition with respect to center pixel be

$$D_i = 1 - \frac{T_i}{(M+N)} \ \forall\ i \ \&\ i \neq\ i0 \tag{9}$$
$$\text{where } T_i = \max[\left|x_c - x_l^{P_i}\right|, \left|x_c - x_r^{P_i}\right|] +$$
$$\max[\left|y_c - y_l^{P_i}\right|, \left|y_c - y_r^{P_i}\right|], \ \forall\ i \ \text{ and } i \neq i0$$

  The normalization by M+N (the maximum possible extent) ensures D_i lies in [0,1]. Now the mask is weighted as

$$Mask_{i,(x_c,y_c)}(x,y) = V_{avg}(i) \times D_i,$$
$$\forall (x,y) \in P_i, \ \forall\ P_i \in P(x_c,y_c) \text{ and } \quad i \neq i0 \tag{10}$$

  In one expression, we write the mask obtained for a Gaussian function with σ as:

$$Mask^{\sigma}_{(x_c,y_c)}(x,y) = Mask_{i,(x_c,y_c)}(x,y),$$
$$\forall\ (x,y) \in P_i \ \forall\ i \in \{1,2,...,N\} \tag{11}$$

  Finally, the center-surround operator mask at location (x_c,y_c) is obtained as follows:

$$CSMask^{\sigma_k,\sigma_t}_{(x_c,y_c)}(x,y) = \frac{Mask^{\sigma_k}_{(x_c,y_c)}(x,y)}{\sum_{(x,y)\in L} Mask^{\sigma_k}_{(x_c,y_c)}(x,y)} - \frac{Mask^{\sigma_t}_{(x_c,y_c)}(x,y)}{\sum_{(x,y)\in L} Mask^{\sigma_t}_{(x_c,y_c)}(x,y)}, \ \sigma_t > \sigma_k \tag{12}$$

  We get the object saliency at the location (x_c,y_c) is

$$S_{\sigma_k,\sigma_t}(x_c,y_c) = \left|CSMask^{\sigma_k,\sigma_t}_{(x_c,y_c)} \times I\right| \tag{13}$$

  We calculate the object saliency at all image pixels considering all scales as

$$S(x_c,y_c) = \sum_{\sigma_k,\sigma_t} \frac{S_{\sigma_k,\sigma_t}(x_c,y_c)}{\max_{(x_c,y_c)} S_{\sigma_k,\sigma_t}(x_c,y_c)} \tag{14}$$

An image pixel where we get a non-zero object saliency value is considered to be a part of a salient object to achieve multiple salient object detection.

Some significant attributes of our operator mask generation are:

- As the values of a mask at pixels in a partition (center or neighboring) are the same and a partition captures the shape of locally homogenous regions, our masks are adapted to local image structure.
- As the Gaussian functions used to compute the masks depend on object′s size to be detected as salient, the mask is tuned to be application-specific guaranteeing proper performance.
- The image partitioning also provides accurate boundaries of the objects to be detected as salient for proper localization.

To validate our algorithm and to check its accuracy of detecting multiple objects, various types of mechanical components with largely distinct shapes and sizes are considered. The obtained multiple salient regions for different objects are shown in Fig.3. Due to the local structure awareness of our algorithm, in each salient region, the boundary of corresponding object is preserved which is proved advantageous in further processing of objects for robotic manipulation like grip point selection, shape and size estimation.

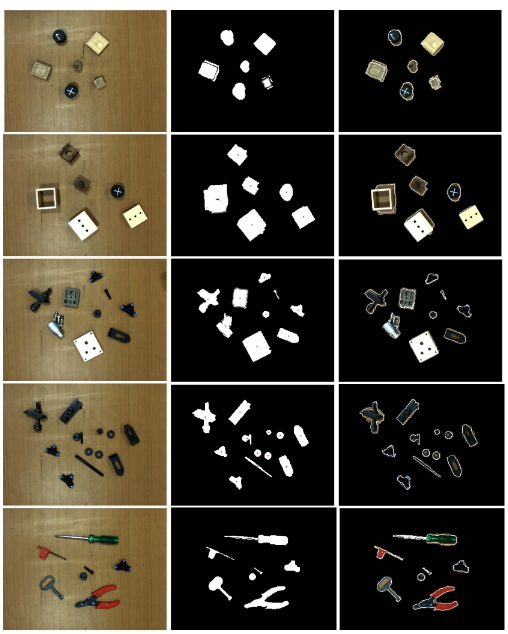

Fig. 3. (a) Original Image (b) Detected Salient Regions (c) Objects Present in Detected Salient Regions

III. OBJECT RECOGNITION AND ASSEMBLY

For automatic assembly by robot, after successful detection, objects must be recognized along with their locations and orientations (localization). We consider the centroid of the detected object as its location. For recognition we use feature point matching technique between template and detected image regions as all the objects to be assembled in a specific application are known a priori. Feature point matching has been used as an effective tool for recognition [10], [11], which also gives the orientation detail. In conventional feature matching technique, all feature points in the entire image is exhaustively searched for a matching with that of template's key points. This makes the system computationally expensive and unreliable due to presence of background clutter. However, we perform the matching only at regions where salient objects are detected. It is a very reasonable assumption that an object to be manipulated will be among the salient ones in a scene. This approach of ours, in spite of the added "burden" of saliency detection, makes recognition significantly efficient / faster.

A. *Object recognition by feature point matching*

In our proposed system, we use SIFT [19] matching for recognition of objects to be assembled. Recognition by SIFT is a point feature matching technique in which key points are extracted along with their descriptors, and these descriptors are matched with the already stored database descriptors by nearest neighbor technique. Instead of processing the entire image, in our method, SIFT feature point extraction and nearest neighbor matching are done only at the detected salient regions as shown in Fig. 4 and Fig. 5 which makes the system significantly faster. Matching at detected salient regions also increases pose estimation accuracy by eliminating the chance of false positive matches that arises due to the presence of background feature points. Further improvement in matching and outlier elimination is done by Hough transform clustering. Hough transform clustering clusters matching correspondences with consistent poses thus eliminating outliers. Possible object pose is obtained by taking the maxima of the clusters in Hough space. Final pose of the object is obtained by taking RANSAC homography.

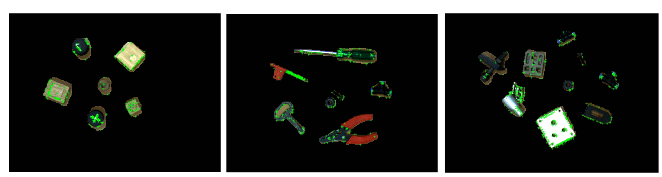

Fig. 4. SIFT key point extraction only at detected salient regions

B. *Automated mechanical assembly*

After successful detection and recognition of all objects with their locations and orientations, robotic manipulator moves to pick up the objects one by one in a predefined sequence from their detected locations at the estimated orientations, and places the object in the designated position on the assembly fixture. The sequence of objects to be placed depends upon the assembly structure. As in case of any automated assembly, our robotic assembly does the various operations through a well-defined strategy, which especially helps to ensure a collision free assembly. As the dimension of each object is known, the grasp height and placing height information is provided to the system.

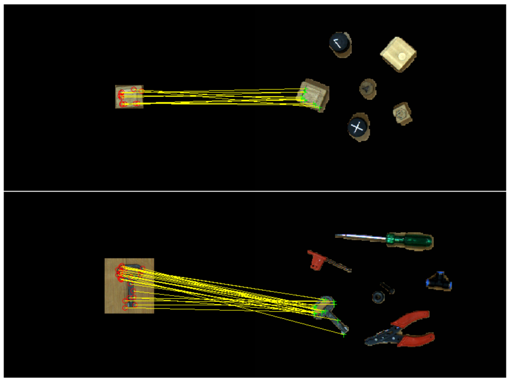

Fig. 5. Candidate point matches between a template image and a salient object

IV. RESULT AND DISCUSSION

The system that we consider to demonstrate our saliency based robotic assembly comprises of a 6-axis industrial robot manipulator (YaskawaMotoman MH5) equipped with a two finger pneumatic gripper. The vision system used consists of a Basler acA1300-22gc GigE camera with Sony ICX445 CCD sensor of 1.3 MP resolution along with Edmund Optics lens of 6 mm focal length. The camera is calibrated to map image coordinates to real world coordinates of robot. All computations are done using Matlab R2016b in a system with intel core i5 processor at 3.30GHz and 8GB RAM, the size of image processed is 1078×958.

The six components assembly structure that we consider for our experiment is shown in Fig.6 (a). Various stages of assembly performed by the robot working on our saliency based approach is shown in Figures 6(b), (c) and (d).

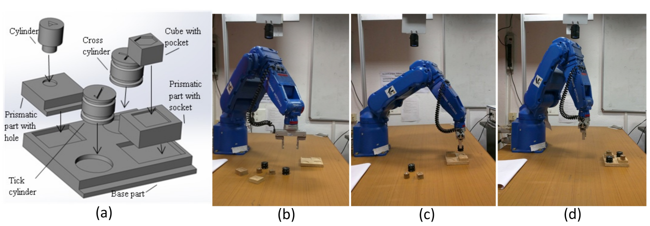

Fig. 6. (a) Assembly Structure (b)-(d) Stages of assembly

Some specifics of our approach applied to mechanical assembly are:

- For the center-surround operations to detect multiple salient objects, five different standard deviations (σ) of the Gaussian functions involved are considered based on the object sizes.
- The order in which the templates are fed for recognition is decided by the assembly sequence.
- The parameters proposed by [19] for eliminating low contrast regions key points gives excellent results in images with high local dynamic range or for textured images. But with that same parameters number of key points decreases for low dynamic range images or images with less texture. Thus, some parameters are modified after experimentation. The threshold value is taken as 0.02 instead of 0.03 as proposed by [19]. Similarly, distance ratio which is the ratio between the distance of the closest neighbor to the second closest one is taken as 1.5 instead of 0.8.

A. Performance of the proposed approach

Our algorithm is designed for fast object localization in unstructured robotic workspaces without compromising in accuracy. To validate our algorithm and to check its accuracy of multiple object detection, we have tested by taking various types of mechanical components with largely distinct shapes and sizes. Taking 20 observations by placing objects at different positions and orientations in the robotic work place, the approximate positional error, angular error, detection rate and recognition rate are computed for each object that is to be assembled by robot. The first part in Table I summarizes these results. Additionally, only for position and angle accuracy evaluation, 150 observations are taken by considering distinct shaped objects placed at different positions, and the localization and orientation error histograms are shown in Fig.7. The errors are computed by taking absolute differences from actual values. The error values are shown in Table II.

As our approach is possibly one of the first working on object saliency to expedite automatic mechanical assembly, we compare the performance of our entire system with the case where the proposed multiple salient region detection approach is not used in the same system. That is, when, instead of performing recognition only at the salient regions, feature point extraction and matching is executed over the entire image. For reliable comparison, Lowe′s [19] SIFT algorithm is implemented in Matlab. The second part of the table summarizes the results of such a system for the objects to be assembled.

As it was pointed out earlier, the primarily goal was to expedite robotic assembly employing object saliency, we present in Table III, the efficiency gained in terms of reduction in computation time involved in the entire process (saliency detection + recognition).

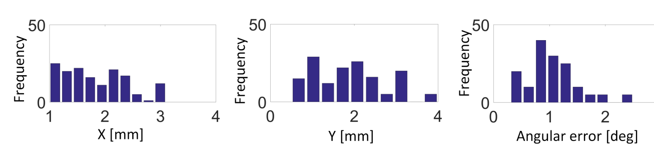

Fig. 7. Histogram of (a) positional error along X-axis (b) positional error along Y-axis (c) angular error

B. Discussion

From Tables I, II and III summarizing results of the experiments, we find that multiple salient object detection can be effectively employed to make existing matching based object recognition for robotic assembly significantly efficient without compromising on performance. Our paper provides a proof of concept of our proposal, whereas a comprehensive

TABLE I

EVALUATION OF PROPOSED (SALIENCY+SIFT) AND CONVENTIONAL (SIFT) METHODS

Assembly parts	Detection			Recognition	
	Positional error (mm)		Detect-ion rate (%)	Angular error (degree)	Recog-nition rate (%)
	X	Y			
Proposed (saliency+SIFT)system					
Prismatic part (pocket)	2±0.5	1.8±0.45	100	1.2±0.5	85
Cross cylinder	1.6±0.35	2.1±0.5	100	0.9±0.3	100
Prismatic part (with hole)	1.5±0.3	1.6±0.35	100	0.6±0.2	100
Tick cylinder	1.4±0.49	2.3±0.55	100	0.8±0.4	100
Cube with pocket	2.4±0.6	2.1±0.52	100	0.8±0.5	80
Cylinder	1.6±0.35	1.9±0.4	100	1.1±0.6	100
Conventional system (SIFT)					
Prismatic part (pocket)	2.5±0.6	2.7±0.5	NA	2.4±0.6	80
Cross cylinder	1.8±0.4	2.3±0.5	NA	1.6±0.4	100
Prismatic part (hole)	2.9±0.6	2.4±0.5	NA	1.8±0.5	90
Tick cylinder	2.2±0.3	2.9±0.5	NA	2.3±0.6	100
Cube with pocket	3.1±0.5	3.8±0.4	NA	1.94±0.7	75
Cylinder	3±0.6	3.2±0.7	NA	3.4±0.3	80

TABLE II

ABSOLUTE LOCALIZATION AND ORIENTATION ERROR

X (mm)	Y (mm)	Angular error (degree)
1.83±0.56	1.88±0.85	1.05±0.46

TABLE III

COMPARISION OF COMPUTATION TIME (AVERAGE) OF CONVENTIONAL MATCHING WITH OUR PROPOSED METHOD

Time Sec	Ours (Saliency + Recognition)	Conventional (SIFT Recognition)
	36	97

experimentation of different robots, object and in separate environments remains to be conducted in future. Certain issues that we observed that can addressed to improve the performance:

- The minor positional errors obtained are due to shadows around objects, where proper illumination or shadow removal can help.
- Reduction in recognition rate results from intra-object color variations due to difference in illumination at different places.

C. Limitation

Here, we have considered the objects to be assembled are placed isolated from each other in the work space. If objects placed are on top of or occludes each other, object localization and recognition through template matching can be erroneous. Use of multiple cameras and/or depth sensors can prove beneficial in such cases, where designing a fast saliency based system would be interesting.

V. CONCLUSION

A novel method for multiple salient object detection is presented in order to perform fast object recognition and localization for vision guided robotic mechanical assembly. Through multiple experiments, it has been observed that the proposed approach is significantly efficient than when a technique for detecting multiple salient objects is not used. The performance in terms of recognition and detection rates, and positional and angular error was found comparable to the existing. The proof of concept of our proposal is established with the help of an industrial robot and an example assembly sequence. Future potential for performance improvement in object saliency based mechanical assembly is also discussed.

REFERENCES

[1] N. Herakovic, Robot vision in industrial assembly and quality control processes, In Robot Vision, InTech 2010.

[2] A.M. Treisman, and G. Gelade, A feature-integration theory of attention, Cognitive psychology, vol. 12, no.1, pp. 97-136, 1980.

[3] C. Koch and S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry, In Matters of intelligence, pp. 115-141, Springer, Dordrecht, 1987.

[4] M. Berger, G. Bachler S. Scherer, Vision guided bin picking and mounting in a flexible assembly cell,17 In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, Berlin, Heidelberg., pp. 109-117, Jun 2000.

[5] MY Liu , O. Tuzel, A. Veeraraghavan, Y. Taguchi, TK Marks,R. Chellappa, Fast object localization and pose estimation in heavy clutter for robotic bin picking, The International Journal of Robotics Research, vol. 31, no. 8, pp. 951-73, 2012.

[6] G.M. Bone and D. Capson, Vision-guided fixtureless assembly of automotive components, Robotics and Computer-Integrated Manufacturing, vol. 19, no. (1-2), pp. 79-87, 2003.

[7] G. Wells, C. Venaille and C. Torras, Vision-based robot positioning using neural networks, Image and Vision Computing, vol. 14, no.10, pp. 715-732, 1996.

[8] I. Lpez-Jurez, M. Casteln, F.J. Castro-Martnez, M. Pea-Cabrera and R. Osorio-Comparan, Using object′s contour, form and depth to embed recognition capability into industrial robots, Journal of applied research and technology, vol. 11, no. 1, pp. 5-17, 2013.

[9] J.H. Su, Z.Y. Liu, and G. Yang, Pose estimation of occluded objects with an improved template matching method. In First International Workshop on Pattern Recognition (Vol. 10011, p. 1001115). International Society for Optics and Photonics July 2016.

[10] A. Collet, D. Berenson, S.S. Srinivasa and D. Ferguson, Object recognition and full pose registration from a single image for robotic manipulation, In Robotics and Automation, 2009.ICRA'09. IEEE International Conference on(pp. 48-55). IEEE, May 2009.

[11] P. Azad, T. Asfour and R. Dillmann, Combining harris interest points and the sift descriptor for fast scale-invariant object recognition, In Intelligent Robots and Systems, IROS 2009, IEEE/RSJ International Conference on (pp. 4275-4280). IEEE, October 2009.

[12] J. Stckler and S. Behnke, Integrating indoor mobility, object manipulation, and intuitive interaction for domestic service tasks, In Humanoid Robots, 2009.Humanoids 2009. 9th IEEE-RAS International Conference on(pp. 506-513). IEEE.

[13] V. Seib, M. Kusenbach, S. Thierfelder and D. Paulus, Object recognition using hough-transform clustering of surf feature, In Workshops on Electronical and Computer Engineering Subfields (pp. 169-176), 2012.

[14] E.N. Malamas, E.G. Petrakis, M. Zervakis, L. Petit and J.D. Legat, A survey on industrial vision systems, applications and tools, Image and vision computing, vol. 21, no. 2, pp. 171-188, 2003.

[15] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on pattern analysis and machine intelligence, vol. 20, no. 11, pp. 1254-1259, 1998.

[16] A. Borji, M.M. Cheng, Q. Hou, H. Jiang and J. Li, Salient object detection: A survey, arXiv preprint arXiv:1411.5878, 2014.

[17] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 5, pp. 603-619, 2002.

[18] Q. Yan, L. Xu, J. Shi and J. Jia, Hierarchical saliency detection. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (pp. 1155-1162). IEEE, Jun 2013.

[19] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, vol. 60, no. 2, 2004, pp. 91-110.