Concatenate Feature Extraction for Robust 3D Elliptic Object Localization

Yuichi Motai
University of Vermont
315 Votey Bldg, 33 Colchester Ave
Burlington, VT 05405, U.S.A.
+1-802-656-9660
ymotai@emba.uvm.edu

Akio Kosaka
Purdue University
1285 Electrical Eng Bldg
West Lafayette, IN 47907, U.S.A.
+1-765-494-3451
kosaka@ecn.purdue.edu

ABSTRACT
Developing an efficient object localization system for complicated industrial objects is an important, yet difficult robotic task. To tackle this problem, we have developed a system consisting first of a vision model acquisition editor, where the object salient features are acquired through a human-in-the-loop approach. Subsequently, two feature extraction algorithms, region-growing and edge-grouping, are applied to the object scene. Finally, by Kalman filter estimation of a proper ellipse representation, our object localization system successfully generates ellipse hypotheses by grouping edge fragments in the scene. The proposed system is validated by experiments using actual industrial objects.

Categories and Subject Descriptors
AI-04 [**Artificial Intelligence**]: Image Analysis and Feature Extraction – Image Segmentation, Feature Matching.

General Terms
Algorithms, Measurement, Performance, Reliability, Experimentation, Human Factors.

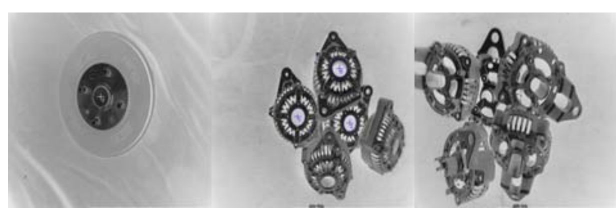
Keywords
3D robot vision system, Salient feature extraction, Human-in-the-loop segmentation, Elliptic edge grouping, Kalman filter estimation.

1. INTRODUCTION
The Robust localization of industrial objects for assembly tasks is an important issue in the robot vision community. A major component of automobile assembly tasks is bin-picking of clustered objects as shown in Figure 1. As shown in Figure 2, the objects in automobile industries, such as several types of alternator covers, hub rotor, and tire etc have shape of a large class of curved objects. The each of these objects can be uniquely defined by a 3D ellipse. In this paper, we propose a new system that is capable of localizing these complex industrial parts by a single camera mounted on a robot hand that carries the camera to multiple viewing points.

Numerical model-based vision systems have been developed to estimate object pose for robotic manipulation. In 3D vision systems [1, 5, 19], the correspondence search between model features and scene features needs to be solved first, and the precise estimation of 3D object pose needs to be then accomplished second. To reliably automate tasks, it is essential to develop robust feature extraction and correspondence matching. In our previous approach [13] for example, the split and merge feature segmentation method was applied for feature extraction of the scene images. Although those extracted features were prominent enough to represent the object, the previous system had some limitations for the particular industrial objects. More specifically, there still existed imperfect extraction of the salient features, due to lighting illumination, object shading, surface material cracks, and imaging noise.

(a) (b) (c)

Figure 1. Robotic manipulation sequences (a) (b) (c) for the localized Alternator Covers 1.



(a) (b) (c)

Figure 2. There are pictures of acquired industrial objects (a) Hub, (b) Alternator Covers 1, (c) Alternator Covers 2.

Although many studies have attempted to solve feature extraction problems, a completely successful algorithm is still not available particularly for objects composed of many circular or elliptical features. In ellipse extraction, the least mean squares method is frequently used to fit an accurate ellipse boundary [2, 6], although this method is very weak for outliers. Main alternative methods are to utilize Hough transformation [9, 16] and moments method [18], where these methods have shown to extract ellipses under some limited cluttered scenes. In spite of the difficulty of feature extraction problems, most vision systems do not provide any backup process for mis-extraction by the selected algorithm.

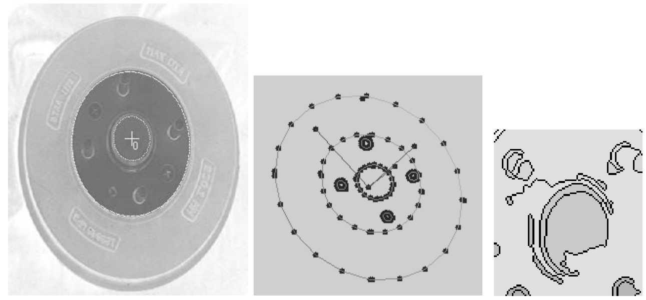
The studies on this frame work also have pointed out that edge grouping methods [10, 11] are promising to extract salient features. Such algorithms aim to utilize global salient structures of object contours, inspired perceptual organization. Based on a set of edges, optimal curve fitting using Kalman estimation is an important extension [7, 17]. One of our feature extraction methods inherits these edge grouping approaches. In order to cope with partial occluded poses of the cluttered objects as shown Fig. 1, grouping is essential to estimate salient/high-level features from local/low-level edges. Regardless of the complexity of the object, edge grouping approaches are possible [11], and several partial edges are only cues to extract the salient features in severe conditions [7]. However when a target object is too complicated, it is difficult to extract features by other extraction methods such as region growing [13]. The first contribution in this paper is to extract salient ellipse feature to represent this complex object class through a new edge-grouping.

The second contribution is to utilize a newly developed model acquisition system. Our vision model for the target object helps to solve the localization problem by providing a feature extraction strategy. More specifically, the vision model is used for choosing a method of extracting features: 1) region-growing segmentation 2) edge-grouping segmentation. The vision model instructs/drives the feature extraction methods and corresponding matching for the cluttered scene, for example, when the first matching does not provide any appropriate feature correspondence, then the system automatically generates up the second feature extraction and matching attempts for the scene.

In the rest of this paper, we will first present the overall strategy for our object localization system. We will then present the way to represent ellipse as salient feature. Following feature extraction methods as our main focus in this paper is described, which include the object model acquisition and the feature extraction in the scene by introducing a new edge-grouped feature extraction in details. The two-phase feature matching will be then presented. Finally, experimental evaluation results will be shown.

2. OVERALL STRATEGY OF OUR SYSTEM

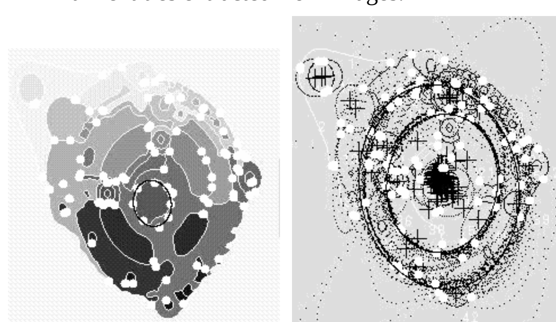
In our new robot learning system, target industrial objects are placed in the work area by a human who then teaches the computer into establishing registration of image-to-image and pose-to-pose correspondences. This object model registration is required by human-in-the-loop with a graphical editor. After the vision model is acquired in this manner, the same object in a random pose is captured from two or three viewpoints for automatic 3D pose calculation. The goal of on-line localization system described here is robust 3D pose calculation of each object. Once the feature matching between model and scene is achieved, then the 3D translation and rotation from the model coordinate to the scene coordinate is computed using quaternion approach [5]. The robot gripper can pick up the object or perform a peg-in-hole alignment through its 3D localization.



(a) (b) (c)

Figure 3. (a) 2D view of an automobile part Hub, (b) 3D object model generated by Human Computer Interaction (HCI) editor with grasping coordinates, (c) An example of broken smooth convex curves.

Among many modules in our system, one of the main focuses in this paper is feature extraction module by a robust edge grouping technique. The feature extraction essentially determines robustness and accuracy of the object localization [13]. In our formalism, a *primitive* is defined as a salient entity of features by which the matching between models and scenes can be directly established. For a representation of the vision model of an industrial object Hub, shown in Fig. 3 (b), ellipses are such primitives. A salient/high-level feature representation helps the system to make stable feature correspondences for estimating the 3D object pose, since such feature representations easily establish a dominant characteristic of the object -- but of course the system may not extract a salient feature in the object scene. For example, the segmented ellipses may be broken into smooth convex curves in the scene as shown in Fig. 3 (c). This fault is improved through the next compensation module by introducing the alternative feature extraction method. The local/low-level feature extraction is here used for subsequent correspondence to increase a stable performance. In our formalism, we call *fragments* for such local image features which correspond to immediate minimum entities extracted from images.



(a) (b)

Figure 4. Feature extraction strategy for an image scene of Alternator Cover 2. (a) region-growing segmentation result (grey level regions), (b) edge-grouping segmentation result (black dot curves).

In the model acquisition procedure of Alternator Cover 2 as shown in Fig. 4, segmentation results are displayed to a human. A human observes and chooses which feature extraction method could be better to apply the specific object. Using the reference of the segmented image, for example of Fig. 4 (b) he/she may choose to represent ellipses on the top of the edge-grouping result by drawing with red color in the figure, where black dot curves are the results of the system's edge-grouping procedure, white edges are fragments, and yellow points are end points of the fragments. In order to increase the robustness of feature extraction and matching, we propose a new strategy of combining two feature extraction methods. Using Fig. 4 (a), the vision model also includes supporting features in addition to salient features. These secondary features are used for verification of salient feature matching later of the localization phase. The vision model for the target object (Alternator Cover 2) has a strategy to extract salient features using edge-grouping segmentation and then to apply secondary feature extraction using region-grouping segmentation. Please note that this strategy is object specific. In the case of Alternator Cover 1, the salient features are extracted by region-growing through the human's inspection of the segmentation results. And edge-grouping is used for extract supporting features by aggregating the fragments in the scene until generating compatible salient features (called *group of fragments*).

For object localization phase, vision model operates feature extraction method and the corresponding matching between model and scene. For example, the system executes the first extraction method to find out salient feature, and subsequently the feature matching is carried out, the system also applies a different (sometimes alternative) extraction method to find out supporting feature. For example of Alternator Cover object 2 shown in Fig. 4, the strategy in the object model describes the following matching procedure: Using few representative numbers of salient features, search the correspondence by checking intrinsic attributes of the representative salient features in both model and scene frames. Among hypotheses of the first matching, apply a different feature extraction method to extract the supporting features to verify those hypotheses.

The advantage of our two-phase feature matching approach is that the system has the capability of compensating the loss of salient feature extraction in the scene, in addition to verify the localization result using different feature extraction results. Two major feature extraction approaches, edge grouping and region growing will provide the dual chances to attempt feature matching between model and scene. Table. 1 summarizes a characteristic comparison of the two feature extraction methods.

Table 1. Two feature extraction methods comparison

Method	Advantage	Disadvantage
Region-based	Stable Accuracy	Low extraction rate
Edge-based	High extraction rate	Unstable accuracy

Our edge grouping strategy is first to gather the fragments based on the attributes, such as size, convexity, and gray level similarity. The system then checks the elliptic curves by the number of fragments participating in forming the ellipses. For each group of fragments, the system estimates the parameters of hypothesized ellipses using iteration of Kalman filtering.

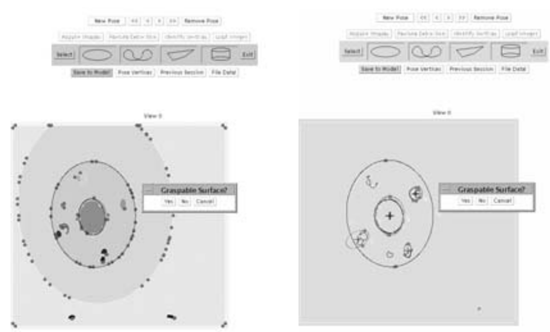
3. FEATURE EXTRACTION METHOD

First of all, features in object model are acquired through HCI editor described in Section 3.1. The acquired vision model of each object has the strategy of feature extraction methods as well. Using this model the corresponding features in object scene are extracted, described in Section 3.2.

3.1 Model Feature Extraction

In the model object feature representation through HCI editor, a human selects one of the dominant shapes shown in the red-highlighted top portion of Fig. 5, and specifies the points along the preprocessed contours by mouse clicking. A human gets to know which salient features are good to extract on the top of image with reference of the candidate features extracted by the system. For example of Hub object, Fig. 5 (a) shows the case of region-growing segmentation, on the contrary, Fig. 5 (b) shows edge-grouping segmentation. For the Hub case, from human's observation, two ellipses are salient features and region-growing segmentation clearly extracts those regions. The edge-grouping also extracts those two ellipses, therefore he/she chooses these identical ellipses by edge-grouping segmentation method as feature verification. The features are acquired in this manner, those features hold several important attributes, such as perimeter, area, shape complexity, and gray level mean. The 2D primitives are then integrated by human's assisted stereo matching process so that the 3D salient primitives can be computed using stereo triangulation.

The output of learning objects is not only a 3D geometrical shape of the object, but also a strategy of feature extraction methods. The system will apply the identical feature extraction methods for automatic localization, which is described next Section 3.2.



(a) **(b)**

Figure 5. Teaching features to represent an object model through Human Computer Interaction (HCI) editor with reference (a) region-growing segmentation result, (b) edge-grouping segmentation result.

3.2 Scene Feature Extraction

For the object scene, the system automatically extracts the salient features using either the region-based segmentation method or the edge-grouping method. The other method is used for extraction supporting features. In this section, we will describe the two methods in detail.

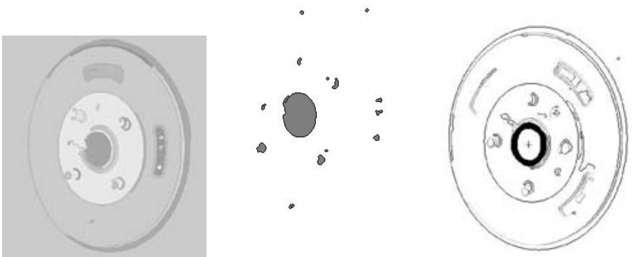
3.2.1 Region Growing Method

The region-based segmentation extracts the segmented areas with not only simple edge tracking but also edge linking to close contours by performing splitting and merging [8] as shown in Fig. 6 (a). This region extraction algorithm has an advantage that the shape of regions are well-preserved, especially on the surface boundaries. The outline of our developed algorithm, inspired from [8], can be described as followings:

1. apply Canny edge detector [4].

2. apply edge linking, especially around T-junctions to produce continuous edge contours.

3. apply split-and-merge segments of homogeneous regions if edge pixels do not exist within its interior.

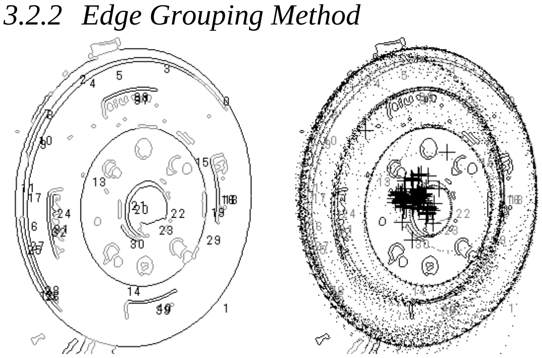


(a) (b) (c)

Figure 6. [Region Growing Method] Feature extraction procedure for the object scene (a) region-based segmentation, (b) pruning the segmented regions by attribute constraints, (c) fitting boundary of the regions for primitive extraction.

The outputs of the above procedures are regions with their complete description given by the convenient quadtree structure and the required attribute measures that will specified from the object model. After applying attribute constraints based on the model, such as area, circularity, shape complexity, perimeter, and average gray level with deviation, many segmented regions in the scene are pruned out as shown in Fig. 6 (b). The boundary under the constraints, if still exist, is fitted to a 2D ellipse shape in the primitive model as shown in Fig. 6 (c).

3.2.2 Edge Grouping Method



(a) (b)

Figure 7. [Edge Grouping Method] Feature extraction procedure for the object scene (a) Curve Fragments extraction (indexed) along edges, (b) Group of Fragments extraction (dotted) by aggregation through elliptical parameter estimation.

Our feature edge grouping method also starts from Canny edge images. After Canny edge detector is applied to an image, the edges are finely sampled to local segments--called *fragments*. These sampled edges are defined by tracking edges. Fragments along curves in the 2D image scene are automatically generated by the system in the following manners:

1. thin the edge pixels so that the edge tracking can be performed.

2. extract endpoints and junction points in the thinned edge map.

3. track the edge map and extract high curvature points along edge curves.

4. divide the long smooth curves into at least two components to avoid the accidental coincidence of merged curves.

5. register curve segments as fragments.

Note that the selection of high curvature points is done by smoothing the curve along its original form. If the deviation of the smoothed curve from the original curve is higher than some threshold and is maximal, then the system registers these points as high curvature points. As a base of low-level feature, the system decomposes the curve into smaller pieces if the curve is long enough and occupies a large angle for the ellipse formation. The decomposing is very useful for avoiding accidental coincidence -- by chance, two different curves are merged due to the viewpoint ill-conditions.

Fig. 7 illustrates the procedure to extract ellipse candidates in the scene image based on grouping fragments. The system checks the elliptic curves by the number of fragments participating in forming the ellipses. As we have discussed in the previous section, a single curve extracted from the image may not necessarily correspond to a perfect ellipse. An ellipse may be broken into several fragments. Therefore, we deal with the grouping of the fragments which potentially constitute an ellipse. The grouping of fragments is decided on the following constraints:

1. *size* The size of the ellipse in the image is limited. For each group of fragments, the combined curves must be smaller in size than some threshold based on the object model.

2. *convexity* Any pair of fragments must not violate the convexity when these fragments are combined.

3. *gray* level similarity Any pair of fragments must possess gray level similarity. Either internal or external region has the similar gray level. Note that this is a typical case for an industrial object when the object is composed of parts of homogeneous color. If the object region is homogeneous along the elliptic curve, then two fragments i and j must satisfy the gray level similarity constraint:

$$\frac{|\mu_i^{Internal} - \mu_j^{Internal}|}{\sigma_i^{Internal} + \sigma_j^{Internal}} < \varepsilon \text{ or } \frac{|\mu_i^{External} - \mu_j^{External}|}{\sigma_i^{External} + \sigma_j^{External}} < \varepsilon \quad (3)$$

In our current implementation, the system generates an ellipse hypothesis based on how many fragments are chosen for grouping, from a single fragment to four fragments. The system first estimates initial parameter of each ellipse and then updates for verifying that ellipse. For each group of fragments for an ellipse candidate, the system verifies whether or not these fragments certainly constitute an ellipse in terms of the parameters. we represent the ellipse by two focal points P_1 and P_2 and the sum *s* of the distances (s_1, s_2) from the two focal points to any boundary points P_k. Let (u_1, v_1) and (u_2, v_2) be image coordinates of two focal points P_1 and P_2 and (u, v) be the image coordinate of arbitrary boundary point P_k. Then

$$f \equiv \sqrt{(u-u_1)^2+(v-v_1)^2}+\sqrt{(u-u_2)^2+(v-v_2)^2}-s=0 \quad (3)$$

Our contribution using Kalman filter approach includes this proper ellipse representation to derive criterion function. More specifically, our procedure is as follows:

Generation of Initial Parameter Estimation

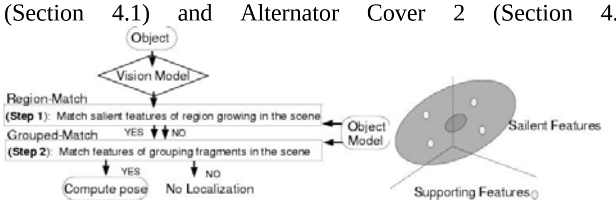
Given a set of points along group of the fragment curves, $\mathbf{p}=(u_1, v_1, u_2, v_2, s)$ is to be estimated. First of all, the system generates an initial estimate of **p**, and then applies the Kalman filter to update the parameter **p** [12]. In order to compute the initial estimate of **p**, we first compute the centroids and the moment of inertia for the image points participating in the fragment set. The initial estimates of (u_1, v_1) and (u_2, v_2) are computed as the above centroids. The initial estimate of *s* is computed by the sum of the lengths of two axes spanned by the moment of inertia. We also associate the covariance matrices for these parameters. The covariance matrices are assigned on the basis of the experiments.

Verification of Ellipse Formation

After the system obtains the initial estimate of the ellipse parameter **p**, the system selects representative points from the fragments. This is done by equally selecting points along the boundary curves. In our current implementation, the system selects at least 16 points for each fragment. By applying the Kalman filter to the constraint equation of Eq. (2) for every selected boundary point *(u, v)*, the system updates the ellipse parameter **p**.

4. TWO-PHASE FEATURE MATCHING STRATEGY

To increase the robustness of localization, we build the two-phase feature matching as shown in Fig. 8, corresponding to the two feature extraction methods for Hub object case. The salient features are used for hypotheses for localization of the object. Also the supporting features are used for verification of the hypotheses. As described before, it is the object model that decides which feature extraction modules are applied and which matching modules are subsequently applied. Therefore in the two-phase matching, we will show two object case studies, Hub (Section 4.1) and Alternator Cover 2 (Section 4.2).



(a) **(b)**

Figure 8. Overall matching feature strategy based on the object model for Alternator Cover case, (b) Salient feature (two ellipses) and identical supporting features (two ellipses).

4.1 Matching Hub Object Case

For Hub, the salient features and supporting features are built by learning feature procedure as shown previously in Fig. 5 (a) and (b). The features for Hub model is illustrated in Fig. 8 (b). Fig. 8 (a) shows the matching procedure of the corresponding features between model and scene.

Region-Match (**Step 1** in Fig. 8) aims to match the primitives in the object model with the segmented region boundaries in the object scene. Using the salient features in the model, the system searches for the correspondence by checking the intrinsic attributes of the scene image. The constraints of the model attribute, such as area, circularity, shape complexity, perimeter, and average gray level, are used for pruning out the candidates of salient features in the scene. After pruning, the boundary of the region in the scene is fitted to the 2D primitive. Those 2D primitive-fits in the scene are then reconstructed using the stereo construction of the multiple-view images. Our reconstruction from 2D to 3D utilizes a model-fit for a 2D image scene first, and then computes the 3D coordinates using the 2D model-fit, which we call model-based stereo reconstruction [13]. In the final process, the reconstructed 3D salient features are verified by the attributes in 3D space to be matched. As illustrated in Fig. 8 (a), if the hypotheses are generated by Region-Match, the following Group-Match is carried for verifying the hypotheses.

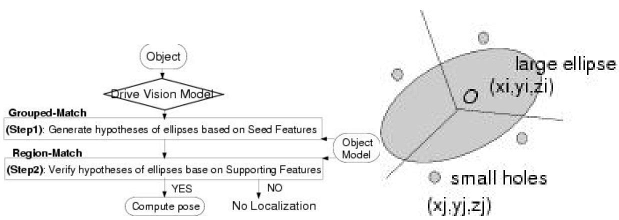
Group-Match (**Step 2** in Fig. 8) aims to match the primitives in the model with the groups of fragments in the scene. For the example of Hub object, the object model executes the other feature extraction method even if Region-Match does not provide localization result. In that case, the groups of fragments are utilized as the compensated primitives in the scene. Under two-phase feature matching shown Fig. 8, Group-Match is carried out by applying the intrinsic attributes constraints, model-based stereo reconstruction, and the coordinate transformation computation, which are all same as Region-Match. This concatenate process is built to increase the robustness of overall matching performance by additional matching even if Region-Match does not provide any feature matching results. If the Region-Match generates the hypothesis, these Group-Match is used for verifying the hypotheses.

The salient feature with region growing method in **Region-Match** is reliable if the features in the scene are successfully extracted, although the extraction rate of the salient feature is lower than that of the fragmentary feature. On the other hand, the feature with edge groping method in **Group-Match** enables the system to back up robustly for further feature correspondence, which then directly increases robustness of object localization. For Hub object, the grouping of fragments will provide the second chances to attempt matching between model and scene. For Alternator Cover object, the grouping will provide many hypotheses of ellipses to extract salient features. The Alternator Cover 1 is the same matching procedure to Hub. On the other hand, the Alternator Cover 2 is the opposite as described in the next section.

4.2 Matching Alternator Cover Case

For a different example object, Alternator cover shown in Fig. 10 (a), ellipses are salient features to generate object hypotheses in the images. Our algorithm described in the previous edge grouping procedures generates all potential ellipse candidates for each image as a hypothesis from a 2D image. Our next task is to verify such a 2D hypothesis based on 3D geometric constraints; more specifically it is based on feature matching of ellipses (Group-Match). In addition, the object model further deals with supporting features that region-based segmentation extracts and applies the secondary matching (Region-Match). Note that the procedures of Alternator cover are different order of the procedures of Hub.

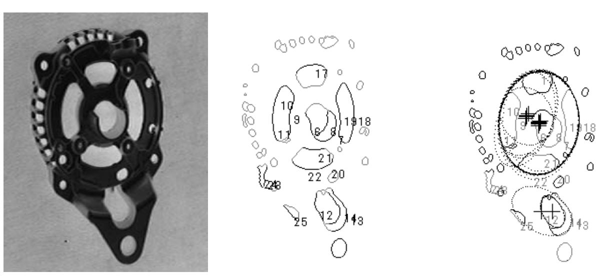
To make the algorithm robust enough for bin-picking, the more detail steps of automatic feature matching are illustrated as the two steps in Fig. 9 (a). Also illustrated in Fig. 9 (b), we utilize an object model, consisting of ellipse (salient feature) and small holes (supporting feature).



(a) **(b)**

Figure 9. (a) Overall matching feature strategy based on the object model for the Alternator Cover case, (b) Salient feature (3D ellipse) and supporting features (small holes).

Group-Match (Step 1 in Fig. 9) Hypothesis Generation of Objects Based on Salient Features (Large Ellipses): In the first step, the 2D results of edge grouping extraction are reconstructed in 3D, and then used for matching with model as following procedures:



(a) **(b)** **(c)**

Figure 10. (a) Object scene, (b) Curve fragments extraction along edges, (c) Group of fragments extracted by aggregation through elliptical parameter estimation.

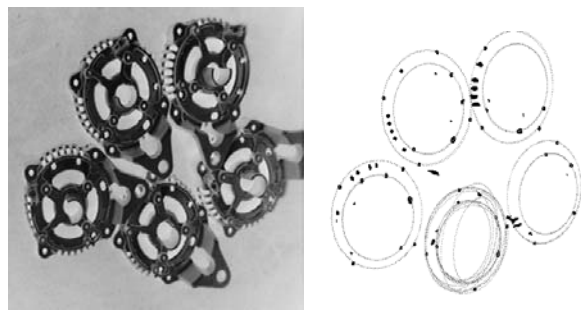
[**Step 1-1**] Given groups of fragments, generate hypothesis that optimally fits to the each hole by computing elliptic parameters.

[**Step 1-2**] For each hypothesized ellipse in the left, look for an ellipse in the right image which will correspond to the left one by considering the epipolar constraint.

[**Step 1-3**] By epipolar constraint, estimate the salient feature position in the 3D space. Check whether or not this feature will support the estimated pose of the object.

[**Step 1-4**] Apply attribute constraints of the correspondent model, such as area, circularity, shape complexity, perimeter, and average gray level with deviation, many hypothesized regions are pruned out.

Region-Match (Step 2 in Fig. 9) Hypothesis Verification of Objects Based on Supporting Features (Small Holes): As you expect, there exist several mismatches of generated hypothesis, which should be removed by verification using supporting features. Two fine-line ellipses in Fig. 11 (b) represent where the small holes should be, given the estimated poses. These supporting features are extracted by a different extraction method, such as region-based method called split and merge segmentation process [13], because this method is just useful to extract small holes as you see Fig. 11 (b). The second steps are described more details as follows:



(a) **(b)**

Figure 11. (a) 2D captured view (b) Small holes extracted along hypothesized regions.

[**Step 2-1**] Based on the 3D object pose hypothesis, generate constraint regions of 3D supporting features where the supporting features should be in the left and right images, by projecting the 3D supporting features onto the 2D images with a given pose associated with the hypothesis.

[**Step 2-2**] Select 2D features associated with 3D supporting features in the left and right images within the constraint regions.

[**Step 2-3**] Estimate the 3D positions of 3D supporting features based on the stereo correspondences of 2D features obtained in previous **Step 2-2**.

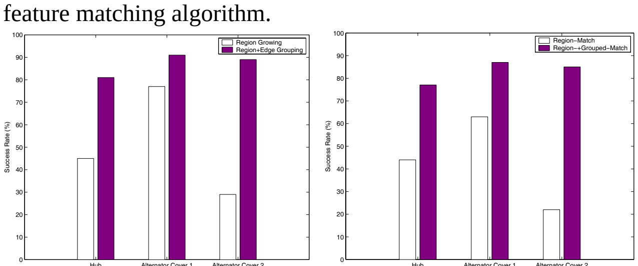
[**Step 2-4**] Verify the hypothesis by considering the geometric constraints of 3D supporting features: These 3D supporting features should be compatible with the hypothesized 3D object pose given in **Step 1**, as well as the 3D supporting features satisfy the geometric constraints, e.g., distance between 3D supporting features, orientation between 3D supporting features. Also the number of supporting features should exceed a user specified threshold.

[**Step 2-5**] Find an optimal solution, if multiple solutions exist within a certain portion of the workspace. (This may happen due to the edge grouping). For each solution, we approximate the object space occupancy. If multiple solutions share the space occupancy, then select the optimal solution from such shared solution sets. The optimality is based on the geometric constraints of the 3D salient features and 3D supporting features, which are associated with the fitting error.

As shown in next experiment section, the 3D matched features are used for computing the 3D coordinate transformation from the model coordinates to the object scene coordinates using quaternion [5].

5. EXPERIMETAL RESULTS

We implemented the off-line Human Computer Interaction editing process for learning object model in a SUN Workstation. It took several minutes for a human and a computer to interactively generate each object model to learn features. Our wrist-mounted robotic vision systems consisted of a Sony DC-47 monocular 1/3 inch CCD camera with Pulnix Lens of focal-length 16 mm, a PUMA 761 or Kawasaki JS10, and a PC. We mounted a monocular camera on the robotic manipulation gripper for capturing multiple images of the object from different viewpoints. For stereo reconstruction, at an arbitrary viewpoint, calibration matrix was computed [15]. We used two or three views to automatically localize the object in the testing phase, and five views to generate the object model in the learning phase. In the automobile assembly tasks, our target shape, ellipse is a large class of curved objects in industry. Currently our focus will be limited to 3D planar elliptical features, as exemplified by the curved silhouettes on the industrial object shown in Fig. 12. The three automobile industrial objects, Hub and two different type of Alternator Cover 1 and 2 were analyzed for localization 35 times, total around 140 pieces for each type of Alternator cover. We evaluated the **robustness** and the **accuracy** of our two-phase feature matching algorithm.



(a) (b)

Figure 12. (a) Feature Extraction success rate, (b) Feature Matching success rate, Previous method vs. Our method of Hub and Alternator Cover 1 and 2.

The **robustness** was examined by counting the successful cases of 1) the features extraction in the scene (Fig. 12 (a)), and then 2) the features matching by applying Region-Match and Group-Match (Fig. 12 (b)). When the feature extraction strategy of the object model was applied to Alternator Cover 2, for example, the detail success rates of feature matching were the following categories:

successful localization case: localization completed with a success robotic manipulation (within tolerance range) [84.7%]

inaccurate localization case: localization completed with false manipulation (outside tolerance range) [2.1%]

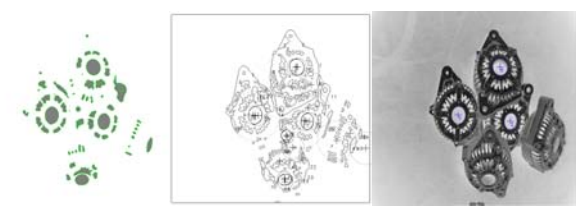
incomplete localization case: no localization output [13.2%]

Fig. 12 shows that our overall success extraction rate was 81% for Hub, and 91% and 89% for Alternator Cover 1 and 2. Also our localization success rate was 77% for Hub, 87% and 85% for Alternator Cover 1 and 2.

Table 2. Localization result error of three objects

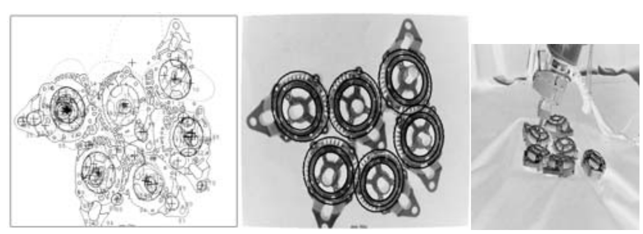
Object	Process	Translation(mm)	Rotation(0)
Hub	Region-Match	3.4±1.9	11.3±5.7
Drum	Group-Match	3.5±2.0	12.0±5.1
Alternator	Region-Match	1.4±0.5	4.3±1.6
Cover 1	Group-Match	1.1±0.4	4.0±1.0
Alternator	Group-Match	1.9±0.7	4.4±1.8
Cover 2	Region-Match	2.2±1.1	5.1±2.2

The localization **accuracy** was also evaluated by translation and rotation in the world coordinates. The quantitative error, shown in Table 2, was computed with the successful cases in Fig. 12 (b). The outline diameter was 240 mm for Hub and 107 mm, 30mm and 65 mm for Alternator Cover 1 and 2 of symmetric circle outline, respectively. These localization results were verified through the robotic manipulation. For example, the error of 5 mm and 10 degree was within the tolerance range for our robot manipulation of Alternator Cover. Since we already calibrated all the robotic coordinate transformations [15], the robot hand can manipulate the localized object, as shown in Fig. 1.



(a) (b) (c)

Figure 13. [Alternator Cover1] (a) region-base feature boundary (green color), (b) edge grouping extraction (blue color), (c) model superimposition based on Region-Match.



(a) (b) (c)

Figure 14. [Alternator Cover 2] (a) edge grouping, (b) 2D superimposition of model onto scene, (c) robot manipulation demonstration from close view.

For Alternator Cover 1, Fig. 13 (a) shows a typical result of Region-Match, and Fig. 13 (b) is a result of Group-Match. These results provide the localization parameters and superimposed by the object 3D model onto the 2D scene. Similarly for Alternator Cover 2, Fig. 14 (a) is shown hypotheses of ellipses by edge grouping. Obviously for these multiple complex objects, our proposed system achieved improvements in both robustness and accuracy.

6. CONCLUSIONS

We proposed a new localization system (1) to generate a vision model through a human-in-the-loop editor and (2) to establish a two-phase feature matching algorithm. Our contributions are (1) to establish a competent edge-grouping method to generate ellipse hypotheses in a complex object, and (2) to derive an efficient ellipse representation for Kalman estimation. Using each group of edge fragments, the system estimated the parameters of ellipse hypotheses using an extended Kalman filtering. The advantage of Kalman estimation was that a desired feature could be robustly extracted regardless of ill-condition of partial occlusions and outlier noises. For optimizing the criterion, we introduced a proper parametric representation of an ellipse feature to achieve a stable result.

The desired advantage was that the system was capable of compensating the loss of salient feature extraction in the scene. Our feature extraction methods were based on a concatenation of both region growing and edge grouping. At the matching stage, the primitives by region growing method in the scene were matched with the model through the attribute constraints. Subsequent matching was carried out, after the fragments generated along the edges were grouped together through a new parameter estimation using the extended Kalman filter. The evaluation results of three different objects verified that the extraction methods of combining region growing and edge grouping improved the success rate of salient features, thus increased the cases of making correspondence of the features between model and scene.

A next extension is to enhance strategy of the feature extraction procedures. Because human "draw" features on the top of the images, the system gets to know which feature extraction method is close to the choice of the human's feature extraction. In this manner, the feature extraction should be adapt itself to the human's choices with respect to threshold and parameter setting.

REFERENCES

[1] Bolles, R. C. and Horaud, P. 3DPO: A Three-Dimensional Part Orientation System, International Journal of Robotics Research, Vol. 5, No. 3, 1986. 3-26.

[2] Chen, C. H. and Kak, A. C. Robot vision system for recognizing 3D objects in low-order polynomial time Systems, IEEE Transactions on Man and Cybernetics, Vol. 19, No. 6, 1989. 1535-1563.

[3] Chatterjee, C. and Chong, E. K. P. Efficient algorithms for finding the centers of conics and quadrics in noisy data, Pattern Recognition, Vol. 30, No.5, 1997. 673-684.

[4] Canny, J. A computational approach to edge detection, IEEE Trans. of Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, 1986. 679-698.

[5] Faugeras, O. D. Three-Dimensional Computer Vision, MIT Press, 1993.

[6] Fitzgibbon, A. Pilu, M. and Fisher, R.B. Direct least square fitting of ellipse, IEEE Trans. of Pattern Analysis and Machine Intelligence, Vol. 21, No. 5, 1992. 476-480.

[7] Guichard, F. and Tarel, J. Curve finder combining perceptual grouping and a Kalman like fitting, Proc. of IEEE International Conf. on Computer Vision, Vol. 2, 1997. 1003-1009.

[8] Horowitz, S. L. and Pavlidis, T. Picture segmentation by a tree traversal algorithm, Journal of ACM, Vol. 23, No. 2, 1976. 368-388.

[9] Hough, P. V. C. Method and means for recognizing complex patterns, US Patent, No. 3 069 654, 1962.

[10] Huttenlocher, D. P. and Wayner, P. C. Finding convex edge groupings in an image, International Journal of Computer Vision, Vol. 8, No. 1, 1992. 7-27.

[11] Jacobs, D. W. Robust and efficient detection of salient convex groups, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 18, No. 1, 1996. 23-37.

[12] Kosaka A. and Kak, A. C. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties, Computer Vision, Graphics, and Image Processing - Image Understanding, Vol. 56, No. 3, 1991. 271-329.

[13] Kosaka A. and Kak, A. C. Stereo vision for industrial applications, Handbook of Industrial Robotics, edit. S. Y. Nof, John Wiley & Sons, Inc., 1999. 269-294.

[14] Motai, Y. and Kak, A.C. An interactive framework for acquiring vision models of 3D objects, IEEE Trans. of Systems, Man, and Cybernetics - Part B: Cybernetics, in press, 2003.

[15] Motai, Y. and Kosaka, A. SmartView: Hand-eye robotic calibration for active viewpoint generation and object grasping, Proc. of IEEE Int. Conference of Robotics and Automation, 2001. 2183-2190.

[16] Olson, C. F. Improving the generalized Hough transform through imperfect grouping, Image and Vision Computing, Vol. 16, 1998. 627-634.

[17] Porrill, J. Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter, Image and Vision Computing, Vol. 8, 1990. 37-41.

[18] Voss, K. and Suesse, H. Invariant fitting of planar objects by primitives, IEEE Trans. of Pattern Analysis and Machine Intelligence, Vol. 19, No. 1, 1997. 80-84.

[19] Yoshimi, B. H. and Allen, P. K. Active uncalibrated visual servoing, Proceeding of IEEE International Conference on Robotics and Automation, Vol. 1, 1994. 156-161.