MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Finding a Needle in a Specular Haystack

Shroff, N.; Taguchi, R.; Tuzel, O.; Veeraraghavan, A.; Ramalingam, S.; Okuda, H.

TR2011-082 May 2011

Abstract

Progress in machine vision algorithms has led to widespread adoption of these techniques to automate several industrial assembly tasks. Nevertheless, shiny or specular objects which are common in industrial environments still present a great challenge for vision systems. In this paper, we take a step towards this problem under the context of vision-aided robotic assembly. We show that when the illumination source moves, the specular highlights remain in a region whose radius is inversely proportional to the surface curvature. This allows us to extract regions of the object that have high surface curvature. These points of high curvature can be used as features for specular objects. Further, an inexpensive multi-flash camera (MFC) design can be used to reliably extract these features. We show that one can use multiple views of the object using the MFC in order to triangulate and obtain the 3D location and pose of the shiny objects. Finally, we show a system consisting of a robot arm with an MFC that can perform automated detection and pose estimation of shiny screws within a cluttered bin, achieving position and orientation errors less than 0.5 mm and 0.8 respectively.

IEEE International Conference on Robotics and Automation (ICRA)

Finding a Needle in a Specular Haystack

Nitesh Shroff*, Yuichi Taguchi†, Oncel Tuzel†, Ashok Veeraraghavan†, Srikumar Ramalingam†, and Haruhisa Okuda‡

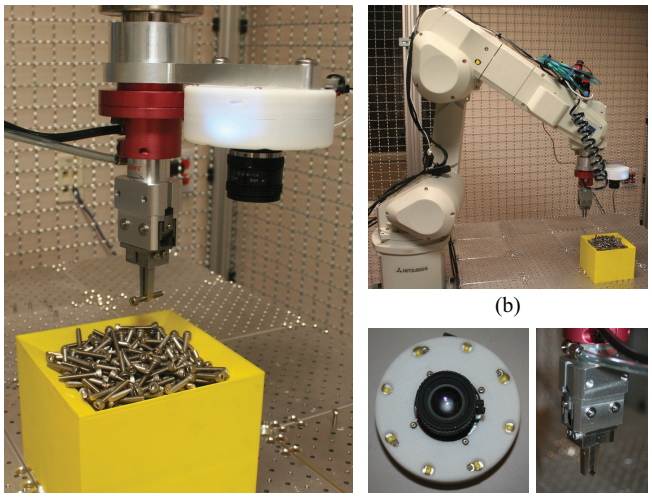
*University of Maryland †Mitsubishi Electric Research Labs ‡Mitsubishi Electric Corporation

***Abstract*— Progress in machine vision algorithms has led to widespread adoption of these techniques to automate several industrial assembly tasks. Nevertheless, shiny or specular objects which are common in industrial environments still present a great challenge for vision systems. In this paper, we take a step towards this problem under the context of vision-aided robotic assembly. We show that when the illumination source moves, the specular highlights remain in a region whose radius is inversely proportional to the surface curvature. This allows us to extract regions of the object that have high surface curvature. These points of high curvature can be used as features for specular objects. Further, an inexpensive multi-flash camera (MFC) design can be used to reliably extract these features. We show that one can use multiple views of the object using the MFC in order to triangulate and obtain the 3D location and pose of the shiny objects. Finally, we show a system consisting of a robot arm with an MFC that can perform automated detection and pose estimation of shiny screws within a cluttered bin, achieving position and orientation errors less than 0.5 mm and 0.8° respectively.**

I. Introduction

Steady advances in machine vision algorithms have resulted in the adoption of these techniques in industrial automation and assembly. Nevertheless, several challenges still persist in tackling objects that do not have a near-Lambertian reflectance. Objects with mirror-like, transparent or translucent surfaces possess material properties that are currently viewed as noise sources and traditional techniques attempt to suppress the impact of these effects. This means that objects which are either highly specular or have significant translucency cannot be handled by such methods since these material effects cannot be completely suppressed.

In this paper, we show that the effect of specularities need not be treated as a source of noise but rather as a feature that can and does enhance our ability to detect, recognize and interact with specular objects. We show that regions of high curvature on a specular object lead to a very consistent and robust feature that can be used to reliably perform machine vision tasks. The basic idea is very simple. Since high curvature regions in an object have normals that span a wide angular extent, these regions almost always produce specular reflections irrespective of the lighting position. The detection of these specular reflections serves to indicate regions of very high curvature on the surface of the object. The use of a multi-flash camera (MFC) [20] aids in the detection and extraction of these specular features. These high curvature regions provide a distinct signature for several industrial objects, which can be used for object detection, recognition and pose estimation.



(a) (b) (c) (d)

Fig. 1: The robotic grasping platform. *(a) Our system enables the grasping of screws from a cluttered bin using a multi-flash camera (MFC). (b) System overview. MFC is mounted on the robotic arm. Screws are placed in a part container. (c) Close up of MFC, employing* 8 *LEDs uniformly placed on its circumference. (d) Close up of the gripper.*

In this paper, we primarily focus on demonstrating our approach to grasp screws from a cluttered bin, as shown in Figure 1. Arguably, screws form the most fundamental class of objects used in manufacturing systems. More threaded screws are produced each year than any other machine elements [6]. In conventional assembly lines, the screws have to be placed into a part holder with known position and orientation before the robot grasps and manipulates them. This operation requires either a specially designed hardware for each screw type such as a part feeder or performed using manual labor, which can be avoided by our bin picking system.

Majority of screws are made from shiny materials; therefore they cannot be handled easily by traditional machine vision algorithms. In addition, pose estimation of a screw in a bin is a very challenging problem because of clutter and occlusion. We show that the specular features can be robustly extracted even in the cluttered bin. We estimate the pose of screws by matching these features from the same screw in multiple views. This matching is particularly difficult because the entire bin contains many instances of the same screw in a wide variety of poses. To solve this, we employ three-plane rank two constraint for establishing correspondences and two/three-view triangulation for pose estimation.

A. Prior Work

Vision-based bin picking has attracted a lot of interest in the last decade [13], [19], [1] where the main problem is to localize an object and estimate its pose. Development of such systems have been challenging mainly because of (a) the specular reflections from the metallic surfaces of industrial parts and (b) occlusions in a cluttered bin.

Model-based pose estimation algorithms using 3D model to 2D image correspondences can be found in [15], [16], [7]. Unfortunately, 3D-2D point correspondences are hard to obtain for industrial parts due to their textureless surfaces. The situation is particularly severe when multiple identical objects are placed together and overlap each other.

Object contours provide rich information about object identities and their poses. Various contour matching algorithms have been proposed in [8], [3], [14]. However, for specular objects the contour information is difficult to obtain in a cluttered bin since these objects do not have appearance of their own but reflect the surrounding environment.

Range sensors have been widely used for the pose estimation problem. Bolles and Horaud [2] use range data to group the surface features which is then used to generate and verify the hypotheses of object location. Wang et al. [22] use 3D range data to compute shape of flexible industrial parts such as cables. However in the presence of specularities, range sensors fail to produce accurate depth maps and they are comparably more expensive than camera-based solutions.

Active illumination patterns can greatly assist vision algorithms for extracting robust features. Horn et al. [13] use brightness of patches observed with varying illumination condition to estimate orientation of surface patches and then match them with the 3D model. In our approach, we exploit MFC [20] to extract specular features in the regions of high curvature. MFC has 8 LEDs that are uniformly placed on its circumference and flash one by one. As the illumination sources move, the specular reflections also move by a distance that is inversely proportional to the surface curvature.

Specularities have generally been treated as sources of noise in machine vision algorithms. Most vision algorithms identify specularities and remove them before inference. Brelstaff and Blake [4] identify regions with specular highlights which deviate from Lambertian reflectance. Nayar et al. [18] use polarization filters to identify and discard specular highlights. Mallick et al. [17] transform the color space such that it becomes invariant to changes due to specular highlights.

Robust feature extraction in the presence of strong specularities has always been a challenging task. Specular highlights have been used in vision and robotics for object detection and pose estimation [5]. Healey and Binford [12] use specular highlights to reconstruct the local orientation and principal curvatures of a surface. Similarly, Gremban and Ikeuchi [10] use specular highlights for object recognition, and plan novel views that are discriminative between objects with similar highlights.

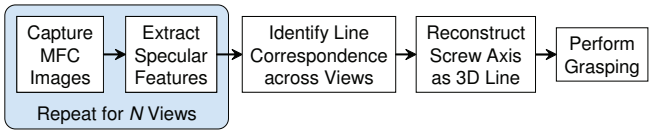


Fig. 2: Overview of our algorithm.

Sankaranarayanan et al. [21] propose a novel image invariant for highly specular and mirror-like surfaces exploiting the fact that the image gradients exhibit degeneracy at regions where at least one principal curvature is zero. This property is used to detect points with parabolic curvature on the object surface. However, this is not a very practical feature for industrial objects, since parabolic curvature points rarely exist on the surface of such objects. In this paper, we consider the exact opposite: i.e., the regions on the surface of the object that have very high curvature. Such regions have a large variety of normals present within a small area and therefore produce specular reflections irrespective of the illumination position.

B. Contributions

The main technical contributions of this paper are

- We show that regions of high curvature in highly specular objects produce specular reflections largely independent of the location of the illumination sources, which can be used as a distinct features for object detection and pose estimation.
- We show that an inexpensive multi-flash camera can be used to reliably detect such specular features.
- We use screws as an example part and develop two view and three view based algorithms for 3D pose estimation using triangulation.
- Finally, we implement the algorithm on an industrial robot and show very high location and angular accuracy for 3D pose estimation.

II. System Overview

Figure 1 shows our system performing bin picking of screws. We use a 6-axis industrial robot arm, on which an MFC is mounted. The MFC was calibrated both internally and externally using a checker board. Hand-eye (gripper-camera) calibration was also performed so that the robot arm can interact and grasp objects using the gripper.

Figure 2 shows the overview of our algorithm. We employ a multi-view approach to compute the pose of screws in a bin. We first capture a set of images using the MFC and extract specular features, as we describe in Section III. We repeat these processes for two or three capture positions. We then identify the lines corresponding to the same screw across the multiple views and estimate the pose of a screw by reconstructing the 3D axis of the screw, as described in Section IV. Our system finally grasps the screw using the estimated pose and performs the subsequent assembly task.

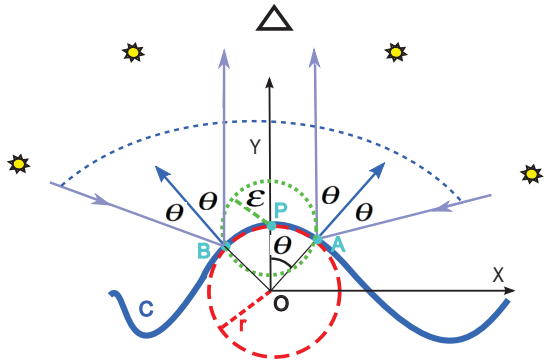


Fig. 3: Analysis of the relation between curvature and the cone of rays. *Point P has curvature κ. If the illumination direction is within the cone (shown by blue dotted curve) of $[-2\theta \quad 2\theta]$, then the specular highlight is captured by the camera within the ε neighborhood (shown in green).*

III. Specular Feature Extraction

In this section, we analyze the reflection of light rays from the surface of a specular object and derive a feature extraction algorithm for detecting high curvature regions.

A. Reflection Analysis

Consider a point P on an arbitrary one dimensional curve C as shown in Figure 3. Let r be the radius of the osculating circle at P. Then the curvature κ at P is given by

$$\kappa = \frac{1}{r}. \tag{1}$$

Consider an ε neighborhood at point P as shown in Figure 3. As ε is small enough, we assume the curvature in this neighborhood to be constant for our analysis. Without loss of generality, we consider the two-dimensional coordinate axes with origin as the center of the osculating circle O, Y-axis passing through P and X-axis orthogonal to it. This ε ball meets the curve C at A and B with θ given by

$$\theta = 2\sin^{-1}\left(\frac{\varepsilon\kappa}{2}\right). \tag{2}$$

Now, consider the ray towards A from a camera placed on Y-axis. The camera center is assumed to be at $(0, y_0)$ such that $y_0 \gg \varepsilon$. This implies that the ray coming from camera center can be considered parallel to the Y-axis as shown in the figure. This ray subtends an angle θ with the normal at A and gets reflected at angle θ from the normal. Symmetrical analysis holds true for point B. This shows that if the light source is placed anywhere within this cone (shown by blue dotted curve) of $[-2\theta \quad 2\theta]$, then the camera will receive specular reflection from this ε neighborhood.

For a fixed cone $[-2\theta \quad 2\theta]$, the size of the ε neighborhood is inversely proportional to the curvature of the point, $\varepsilon = \frac{2}{\kappa}\sin\frac{\theta}{2}$, by using the distant camera assumption. As the curvature increases, the reflections are visible within a small neighborhood of the point. In contrast, when the curve is almost flat (curvature is close to zero), the reflection is not visible within the neighborhood of the point. In the next section we illustrate methods that exploit this fact to detect features points on the high curvature regions of the object. This analysis assumed mirror-like reflection. When the specular lobe is considered, this cone of rays can be increased by an additional 2σ where σ is the width of specular lobe.

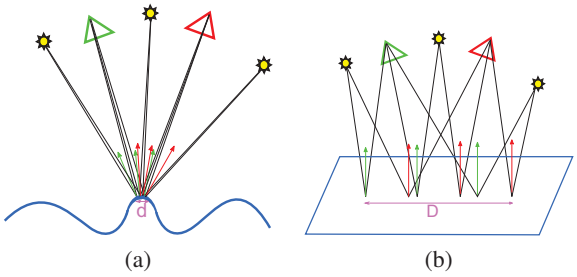


Fig. 4: *(a) High curvature region. The specular highlights shift by a small amount with the changing light location. (b) Low curvature region. The specular highlights shift by a large amount.*

The analysis can be extended for a two dimensional surface S. The principal curvatures are defined as the minimum and maximum values of the curvature measured along various directions at a given point. The Gaussian curvature K of a surface is given by the product of principal curvatures κ_1 and κ_2 of the point

$$K = \kappa_1 \kappa_2. \tag{3}$$

Similarly for a two dimensional surface, the reflections are visible within a smaller neighborhood of the point as the Gaussian curvature of the point increases. Note that both principle curvatures have to be large to observe the reflection within a small neighborhood. For example, a sphere (with small radius) is a surface with both its principal curvatures large ($\kappa_1 = \kappa_2 = 1/r$ for all the points on the sphere), therefore the reflection is visible within a small neighborhood. In contrast, the Gaussian curvature of a point on a cylindrical object is zero since the surface has bending only in one direction. Hence the reflection may not be visible within an immediate neighborhood.

B. MFC-Based Feature Extraction Algorithm

Our feature detection algorithm finds points on the specular surface which have *large Gaussian curvature* (which has bending in both directions) and *normals towards the camera*. Although the second requirement seems like a restriction, in fact the high curvature regions span a large set of surface normals. These features provide sufficient information for 3D reconstruction and pose estimation which are explained in the Section IV.

An MFC is an active illumination based camera that contains 8 point light sources (LEDs) arranged in a circle around the camera as shown in Figure 1. As the different LEDs around the camera flash, the specular highlights on the surface move depending upon the local curvature. From the analysis presented in the previous section, it is clear that the specular highlights remain in a very local neighborhood for all points that have high surface curvature. We exploit this cue to detect

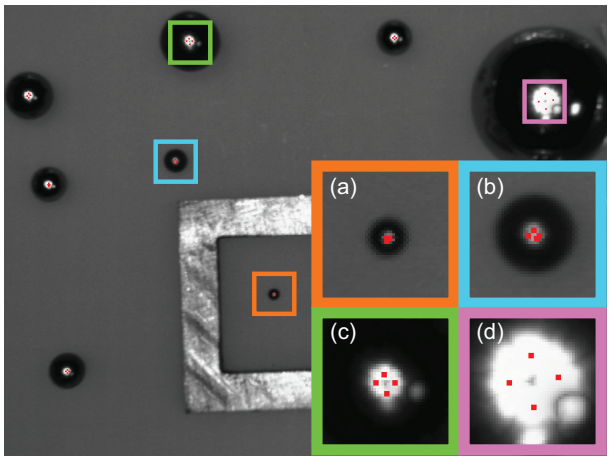


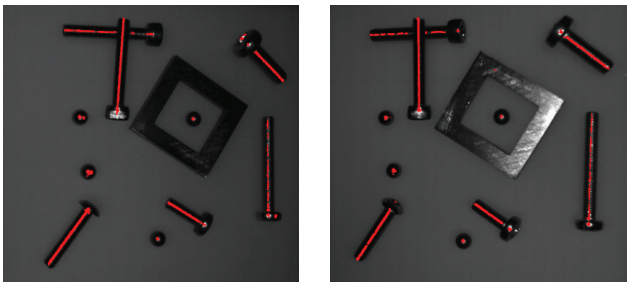
Fig. 5: Motion of specular reflections for different sizes of spheres. *The centers of specular highlights corresponding to* 4 *different light positions are superimposed as red dots on the maximum image. Close-ups of* 4 *spheres are shown in bottom right. The diameter of sphere and the diameter of the motion of specular highlight are respectively (a)* 1.2 *mm–*2 *pixels, (b)* 3.2 *mm–*4 *pixels, (c)* 8 *mm–*7 *pixels, and (d)* 19 *mm–*18 *pixels. The smaller the diameter (the larger the curvature), the smaller the motion of specular highlight.*

pixels that correspond to the high curvature regions on the object having normals towards the camera. These pixels serve as a feature that is both characteristic of the object's 3D shape and its relative pose with respect to the camera.

Our specular feature detection algorithm consists of the following steps. We first take 8 images corresponding to 8 flashes, along with a 9^{th} image with no flash. This image with no flash (ambient image) is then subtracted from each image to remove effects of ambient illumination. Let us denote the image captured (after ambient subtraction) during the flashing time of the i^{th} LED as I_i. The minimum intensity values at each pixel location among these ambient subtracted images are found and used to construct the minimum illumination image $I_{min}(x,y) = \min_i I_i(x,y)$. The minimum illumination image will be similar to the surface albedo. Since the surfaces we are considering are highly specular and the specular highlights move across the images, the minimum illumination image appears to be dark for all the specular regions. We compute the ratio images of the ambient subtracted images to the minimum illumination image, $RI_i = \frac{I_i}{I_{min}}$. Ideally, the ratio values in the non-highlight regions remain close to one while the ratio values in regions seeing a specular highlight are much greater than 1. This information is utilized to detect the highlight regions at each flash image.

As discussed above, with the changing flash position specular highlights at high curvature regions stay within the small neighborhood (Figure 4(a)) whereas they shift by a large amount in low curvature regions (Figure 4(b)). This effect is explained in Figure 5.

We choose a neighborhood ε (set to 1 pixel in our experiments) and compute the number of flash images in which a specular highlight was observed within this ε neighborhood.



(a) View 1 (b) View 2

Fig. 6: Specular feature detection using MFC. *The high curvature regions facing towards the camera are robustly identified across two views whereas low curvature regions and Lambertian surfaces are filtered out.*

For pixels corresponding to high curvature regions, the specular highlight remains within this ε neighborhood and therefore a specular highlight is observed in all 8 MFC images within this ε neighborhood. For pixels corresponding to low curvature regions, the specular highlight moves outside the ε neighborhood and therefore the number of images in which the specular highlight is observed within the ε neighborhood is less than 8. Pixels that correspond to Lambertian surfaces are filtered out automatically because of normalization with respect to the minimum image. An example of specular feature detection using MFC is shown in Figure 6. The algorithm robustly identifies high curvature specular regions facing towards camera across two views while filtering out Lambertian surfaces and low curvature specular regions.

IV. Pose Estimation of Screws

This section presents a practical application of the specular features described in Section III, in conjunction with a novel pose estimation algorithm for bin picking of screws.

A. Specular Features on Screws

Screws are cylindrical in shape. As discussed in Section III, curvature of cylinder in direction perpendicular to the axis of the cylinder is high but in the direction parallel to the axis is small. This renders the Gaussian curvature of cylinder to be small. Let us now consider the effect of adding threads on the surface of screws. The threads (typically approximated as a helix or a conical spiral) provide high curvature to all the points on the screw body even in the direction parallel to the axis of the cylinder. This, in turn, ensures that the specular highlights are visible in an ε neighborhood independent of the illumination direction.

The specular features are extracted using the algorithm described in Section III. It is important to analyze the location of the features detected on the screw. The specular features are detected on high curvature regions having normals towards the camera. Approximating the screw shape with a helix, the points on the helix facing towards the camera lie on a line. In addition, this line lies on the plane joining the camera center and the axis of the screw as shown in Figure 7.

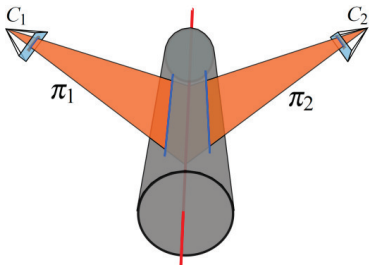
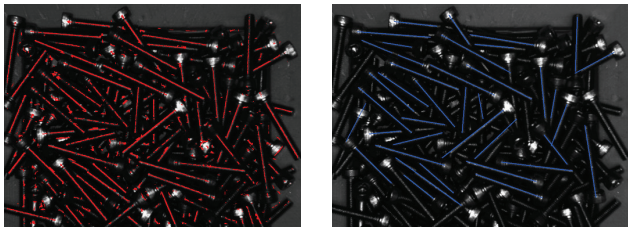


Fig. 7: Central axis reconstruction. *For simplicity the screw shape is shown as a cylinder. Blue lines shown on the surface of the cylinder represent the specular features detected. π_i represents the planes formed by these lines with the respective camera centers. π_1 and π_2 intersect at the axis of the cylinder (shown in red).*



(a) (b)

Fig. 8: *(a) Detected specular features on screws. (b) Fitted lines.*

We therefore represent the specular features on the screws with line segments. For line fitting we use a variant of RANSAC [9] algorithm. The algorithm initially hypothesizes a variety of lines by selecting a small subset of points and their directions. The support of a line is given by the set of points which satisfy the line equation within a small residual and form a continuous structure. The line segment with the largest support is retained and the procedure is iterated with the reduced set until the support becomes smaller than a few points. The RANSAC algorithm provides inlier points for each line segment. We then refine each line segment using least square estimation on the inlier points. In Figure 8, we show detected specular points from a bin of screws and their line segment representation.

B. Pose Estimation

In order to estimate the pose of a screw, we reconstruct the 3D line corresponding to the axis of the screw which uniquely determines the orientation and the position of the screw. We compute the specular features from multiple camera positions. As described before, the detected specular line segments lie on the plane π_i joining the center of projection of the i^{th} camera and the axis of the screw as shown in Figure 7. The reconstruction of the 3D line corresponding to the axis of the screw is given by the intersection of these planes. In the remainder of this section, we present algorithms for reconstruction of the central axis for three view and two view configurations.

1) 3D Line Reconstruction using Three Views: We first look into the well studied relation between a 3D line and its corresponding projections in multiple views. We then use a geometric constraint to find the lines corresponding to the same screw across three views and reconstruct the 3D line.

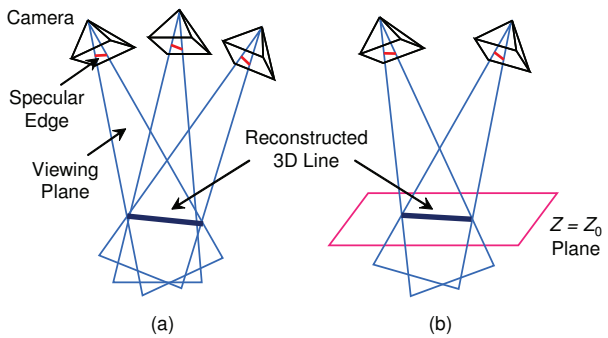


Fig. 9: *(a) Intersection of three planes becomes a line only if the correspondence is correct. (b) Two viewing planes and $Z = Z_0$ plane can be used for the two view case.*

Geometry of Line Projection: Consider a line **L** in 3-space which is imaged in three views. Let l_i be the projection of **L** in view i and P_i be the projection matrix of view i. Using the projection matrix, an image point can be back-projected to a ray (passing through the camera center) in the world coordinate system. Both end-points of the line in the image plane can be back-projected as individual rays. Alternatively, the back projection of an image line l_i corresponds to a plane in view i as π_i

$$\pi_i \equiv A_i X + B_i Y + C_i Z + D_i = 0 \tag{4}$$

Finding Correspondences: Since there are multiple screws in the bin, we need to identify the lines corresponding to the same screw from different views. In order to find correspondences, planes (formed by back projecting the lines) from all three views are transformed to a common coordinate system. This transformation is easily obtained as the camera poses are known. The common coordinate system is chosen as the camera coordinate system of the first view. We then use the property that three arbitrary planes do not meet at a single line to obtain our correspondences. This intersection constraint is expressed algebraically by the requirement that the 4×3 matrix defined by the coefficients of three planes [11]

$$\left(\underbrace{\begin{matrix} A_i & B_i & C_i \\ A_j & B_j & C_j \\ A_k & B_k & C_k \end{matrix}}_{\mathbf{A}} \underbrace{\begin{matrix} D_i \\ D_j \\ D_k \end{matrix}}_{\mathbf{b}} \right) \tag{5}$$

should have rank two. This constraint is satisfied if the determinants of four 3×3 sub-matrices of (5) are zero. In practice, due to noise in image measurements, even for lines in correspondence the determinants of these sub-matrices are not zero but small. We therefore compute as the correspondence cost the sum of four determinant values for each triplet of lines, and select the triplets that have the cost smaller than a threshold.

Reconstruction of the 3D Line: After finding the three lines corresponding to the same screw, we compute the 3D line passing through the axis of the screw. The line equation in

3D can be written as

$$\mathbf{X} = \mathbf{X}_0 + \lambda \mathbf{X}_1, \tag{6}$$

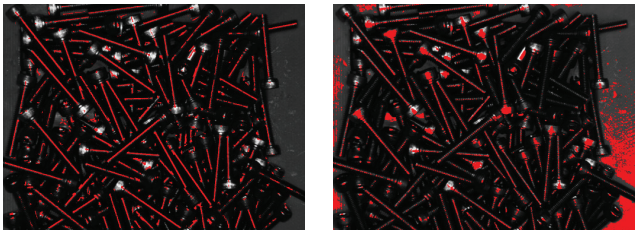
where $\mathbf{X}_0$ is a point on the line and $\mathbf{X}_1$ is the direction of the line. The direction of the line should be computed as perpendicular to the all the plane normals as possible. This can be obtained by solving $\mathbf{A}\mathbf{X}_1 = 0$. The optimum solution (in least squares sense) is given by the right singular vector corresponding to the smallest singular value of matrix $\mathbf{A}$. We select $\mathbf{X}_0$ as the point which is closest to the three planes. This point can be found via the least squares solution of $\mathbf{A}\mathbf{X}_0 = -\mathbf{b}$.

Degenerate Configurations: If two planes are close to being parallel, the rank of the matrix (5) becomes close to two regardless of the other plane, leading to difficulty in finding correspondence. We therefore ignore such pairs of planes by checking the angle between the normals of the two planes. This primarily happens when the axis of screw is aligned with the translation between two camera positions. If we randomly choose three camera positions, all three translational directions between camera pairs become degenerate directions. To avoid this, we move the camera on a straight line so that there is only a single degenerate direction. In addition, we change this direction of motion to handle screws with varying poses during subsequent pickups.

2) 3D Line Reconstruction using Two Views: Since two non-degenerate planes from two views always intersect on a line, finding correspondences using two views is typically difficult. However, the correspondence can be computed if we assume that the 3D lines are located around some plane in the world coordinate system. As shown in Figure 9 (b), we use $Z = Z_0$ plane together with the two viewing planes and compute the correspondence cost in the same way as the three view case. This Z_0 value is obtained by approximately measuring the physical distance between the camera and the bin with screws. The proposed cost favors screw (a) whose Z position is close to Z_0 and (b) whose angle is close to horizontal. After finding the correspondence, we reconstruct a 3D line as the intersection of two viewing planes (without using $Z = Z_0$ plane).

3) Position Estimation: After reconstructing the 3D line (6), we compute the segment of the line corresponding to the screw by back-projecting the end points of 2D lines in each view and finding the intersection point between the back-projected viewing ray and the reconstructed 3D line. We further validate the reconstructed line segment by comparing the length of the reconstructed line segment with the physical length of screw. The center of the line segment is chosen as the 3D gripping position.

Finally, we determine the end point of the line segment corresponding to the head of the screw. We compute the maximum image in all the views, which is simply a pixel-wise maximum over the 8 MFC images. The head of the screw is a smooth region with relatively lower curvature compared to the screw body. Therefore the specular highlights on the screw head move in a larger neighborhood with the alternating flashes and produce brighter patches in the maximum image compared to the screw tail.



(a) Proposed (b) Threshold

Fig. 10: *Comparison of the threshold-based highlight detection with the proposed algorithm. Detected edges superimposed (in red) on the image. Best viewed in color.*

V. Experiments

We performed an extensive evaluation of the proposed method on the robot platform shown in Figure 1. We refer the readers to the accompanying video, which demonstrates bin picking of screws using the robot arm[1].

A. Specular Feature Detection

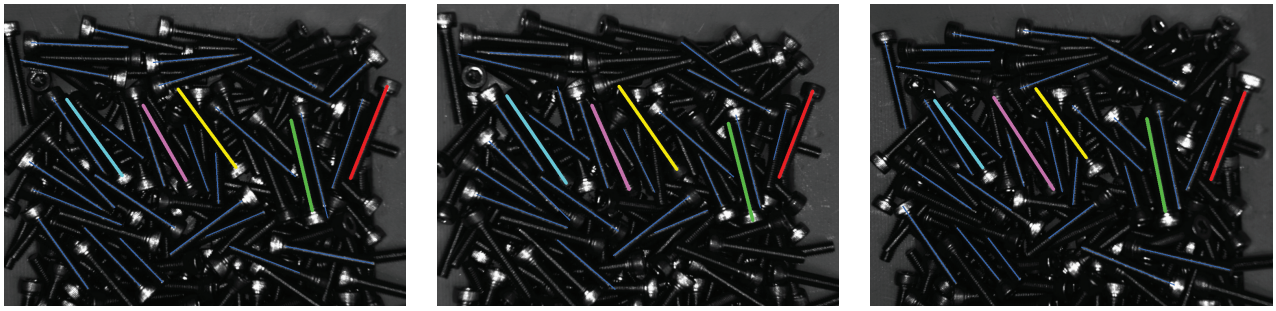
We first compare the proposed specular feature detection algorithm with the basic approach for detecting specularities in the scene using simple thresholding on the maximum image. The results are shown in Figure 10 where detected features are superimposed (in red) on the images. Threshold-based highlight detection is very sensitive and needs to be tuned every time scene or any parameter of scene e.g., illumination, viewpoint, etc. changes. Even with possibly the best choice of threshold, highlights from the glossy background cause a lot of spurious detections as seen in Figure 10(b). The proposed algorithm robustly detects specular features in the regions of high curvature as shown in Figure 10(a).

Figures 11(a)–11(c) show the correspondence results for three view approach. Pixels in blue show lines fit to the detected specular features. The top 5 line correspondences across three views are labeled with different colors. This visually validates the accuracy of our algorithm. The algorithm successfully locates multiple correspondences across the views and hence can simultaneously estimate multiple poses within a single capture cycle. Figure 13 shows histograms of the correspondence costs for true and false matches. There is a large gap between true and false matches.

B. Single Screw Pose Estimation

In order to statistically evaluate the accuracy of the proposed system, we devised a method to evaluate the consistency of pose estimate irrespective of the viewpoint of the camera. Towards this evaluation, we first placed a single (isolated) M4 25 mm length screw in the bin. This is then imaged

[1]The video was captured with unoptimized implementation. Please see Table III for the optimized system performance.



(a) View 1 (b) View 2 (c) View 3

Fig. 11: *Pixels in blue show the line fitting results on specular highlights detected from three views. Top* 5 *line correspondences across views have been shown respectively in green, yellow, cyan, magenta and orange.*

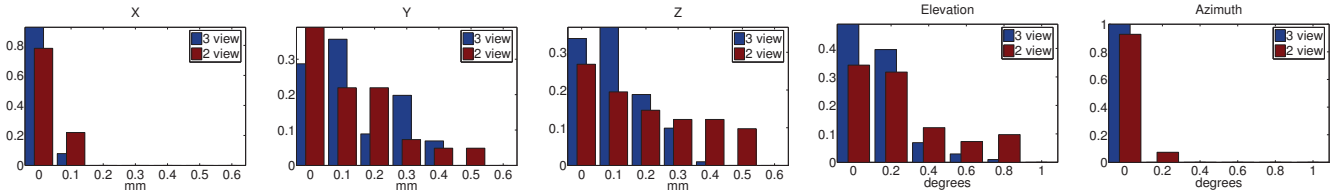
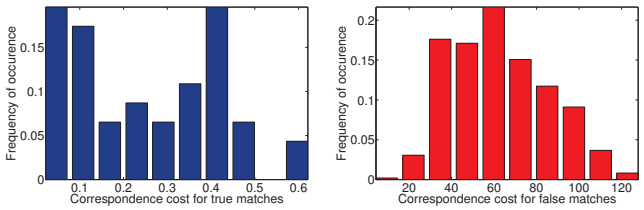


Fig. 12: Histograms of deviations of the pose estimates from their median. *Left three figures show errors in translations* (X,Y,Z) *and right two figures show errors in rotations. The* 3 *view approach (blue bars) has less deviation than the* 2 *view approach (red bars).*



(a) Histogram for true matches (b) Histogram for false matches

Fig. 13: The correspondence cost for true and false matches. *Note that the scale on the horizontal axis for false matches is* $200\times$ *larger than that for true matches.*

TABLE I: *Average absolute pose estimation errors for the 2-view and 3-view approaches.*

Average absolute error	X mm	Y mm	Z mm	Elevation degree	Azimuth degree
2-view	0.06	0.18	0.27	0.38	0.08
3-view	0.05	0.18	0.16	0.23	0.04

TABLE II: *Statistical comparison of screw pickup success rate between* 2*-view and* 3*-view in highly cluttered scenes.*

	Trials	Pickup failure	Flip failure	Success rate
2-view	100	7	6	87%
3-view	100	4	5	91%

using MFC from 15 different locations arranged in a 3×5 grid. This setup produced 50 image pairs for 2-view and 100 image triplets for 3-view pose estimation where the camera positions have reasonable baseline with respect to the screw orientation. Since the object is static, the estimated pose of the object in the world coordinate system should be identical irrespective of the viewpoints. The median pose estimate of each method is selected as the ground truth.

Figure 12 shows the histogram of absolute deviations from the median pose estimates and Table I presents the mean absolute deviations. The results demonstrate that both (3-view and 2-view) algorithms result in highly consistent estimates with mean absolute deviation of less than 0.5 mm in all the three directions (X,Y,Z) and less than 0.8 degrees in the estimates of elevation and azimuth angles. In this experiment, the screw was placed perpendicular to X axis and the motion of the camera had larger baseline along Y axis, which resulted in slightly smaller translational error in X axis. The 3-view approach achieves smaller estimation deviations than the 2-view approach.

C. Bin of Screws

We then quantitatively evaluated the accuracy of the proposed system to detect and localize objects in highly cluttered scenes. More than 200 screws (M4, 25 mm length) were thrown together in a bin to create cluttered scenes just as shown in Figure 1(a). The gripper was made up of two vertical steel pins as shown in Figure 1(d). In its default position, the gripper is open with its pins apart. In order to pick up the screw, the gripper first moves to the estimated location/orientation and then closes the grip with the screw held horizontally between the two pins. This location is assigned such that the midpoint between the two pins coincide with the reconstructed axis and the gripper is perpendicular to the screw direction.

The width of the gripper, in its open position, is kept at 5 mm, which is 1 mm larger than the diameter of the screw. Therefore, to successfully grip the screw (before lifting the object) the error in pose estimate should be less than 0.5 mm. When the pose estimate error is greater than 0.5 mm

TABLE III: *Average time required for each process of the proposed approach. The first* 3 *processes are repeated for* 2 *or* 3 *views. All numbers in seconds.*

MFC Cap.	Spec. Edges	Line Fitting	Corres. (2view)	Corres. (3view)	Total (2view)	Total (3view)
0.87	0.11	0.15	0.01	0.04	2.3	3.4

the gripper fails to pick up the object. We achieved a 96% grasping rate for 3-view and 93% for 2-view approach evaluated over 100 trials each (Table II). Among these successful grasps there were a few cases where the algorithm failed to correctly detect the head direction of the screw which resulted in pose flip failures. There were 5 flip failures in 3-view estimates and 6 flip failures in 2-view estimates, which resulted in overall correct pose estimates of 91% and 87% respectively.

The pickup failures were mainly caused when two screws (say S_1 and S_2) in parallel or anti-parallel orientations are very close to each other. In such scenario, sometimes line correspondence ends up matching S_2 in one of the view to S_1 in the other two views. Hence, pose estimate is completely off and leads to the failure in grasping of the screws. The additional assumption about the $Z = Z_0$ plane in 2-view case increased the number of faulty correspondences and hence led to a slight deterioration in the success rate.

Complexity: Table III shows the average processing time required for each process of the proposed approach in extremely cluttered environments. This was measured on an Intel 3.06Ghz Quad-Core CPU with a C++ implementation. Capturing MFC images is the bottleneck of the current system. The processing times are only 0.53 and 0.82 seconds for the 2-view and 3-view approaches respectively, excluding the capturing time. Although in the current version the robot picks up a single screw for each capture cycle, the algorithm simultaneously estimates poses of multiple screws. It is possible to pick up multiple screws within a single capture cycle, which would drastically accelerate the system.

VI. Conclusion and Future Work

We presented a system for specular object detection and pose estimation using a multi-flash camera. We introduced a novel feature that is present on the high curvature regions of specular objects. We developed an algorithm to reliably detect these features by exploiting the changing lighting positions. We demonstrated our approach for pose estimation of screws in a highly cluttered bin, where we used specular features detected at multiple views to reconstruct the 3D axis of the screw. We implemented our algorithm on a robot arm and achieved highly accurate pose estimation with location and orientation errors less than 0.5 mm and 0.8° respectively. The proposed specular feature is useful in detecting any shiny object with high curvature. Developing pose estimation algorithms using specular features for a wider class of objects would be an interesting future direction.

Acknowledgments: We thank the anonymous reviewers for their helpful comments. We also thank Jay Thornton, Keisuke Kojima, John Barnwell, Ramesh Raskar, Amit Agrawal, Tim K. Marks, and Rama Chellappa for their help and support.

References

[1] A. Agrawal, Y. Sun, J. Barnwell, and R. Raskar. Vision-guided Robot System for Picking Objects by Casting Shadows. *The International Journal of Robotics Research*, 29(2–3):155–173, 2010.

[2] R. Bolles and P. Horaud. 3DPO: A three-dimensional part orientation system. *The International Journal of Robotics Research*, 5(3):3–26, 1986.

[3] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):849–865, 1988.

[4] G. Brelstaff and A. Blake. Detecting specular reflections using Lambertian constraints. In *ICCV*, 1988.

[5] J. Chang, R. Raskar, and A. Agrawal. 3D pose estimation and segmentation using specular cues. In *CVPR*, 2009.

[6] E. P. Degarmo, J. T. Black, and R. A. Kohser. *Materials and Processes in Manufacturing (9th ed.)*. Wiley, 2003.

[7] D. DeMenthon and L. S. Davis. Model-based object pose in 25 lines of code. In *ECCV*, pages 335–343, 1992.

[8] P. F. Felzenszwalb and J. D. Schwartz. Hierarchical matching of deformable shapes. In *CVPR*, pages 1–8, 2007.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381–395, 1981.

[10] K. Gremban and K. Ikeuchi. Planning multiple observations for object recognition. *International Journal of Computer Vision*, 12(2):137–172, 1994.

[11] R. Hartley and A. Zisserman. *Multiple view geometry*. Cambridge university press Cambridge, UK, 2000.

[12] G. Healey and T. Binford. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42(1):62–86, 1988.

[13] B. Horn and K. Ikeuchi. Picking Parts Out of a Bin. Technical Report AIM-746, MIT Artificial Intelligence Laboratory, 1983.

[14] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, R. Chellappa, A. Agrawal, and H. Okuda. Pose estimation in heavy clutter using a multi-flash camera. In *ICRA*, 2010.

[15] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987.

[16] D. G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 13(5):441–450, 1991.

[17] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstructing specular surfaces using color. In *CVPR*, 2005.

[18] S. Nayar, X.-S. Fang, and T. Boult. Removal of Specularities using Color and Polarization. In *CVPR*, 1993.

[19] K. Rahardja and A. Kosaka. Vision-based bin-picking: recognition and localization of multiple complex objects using simple visual cues. In *IROS*, 1996.

[20] R. Raskar, K.-H. Tan, R. Feris, J. Yu, and M. Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics*, 23(3):679–688, 2004.

[21] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Image Invariants for Smooth Reflective Surfaces. In *ECCV*, 2010.

[22] S. Wang, R. Cromwell, A. Kak, I. Kimura, and M. Osada. Model-based vision for robotic manipulation of twisted tubular parts: using affine transforms and heuristic search. In *ICRA*, 1994.