A REAL-TIME OBJECT DISTANCE MEASUREMENT USING A MONOCULAR CAMERA

Peyman Alizadeh 1, Meysar Zeinali 2

1Student Member, IEEE,
School of Engineering, Laurentian University, Sudbury, ON, Canada
E-mail:Px_alizadeh@laurentian.ca
2Member, IEEE
School of Engineering, Laurentian University, Sudbury, ON, Canada
E-mail:Mzeinali@laurentian.ca

ABSTRACT
Distance measurement for moving object using image processing is one of the important research areas in robotics and computer vision. In the current paper, an improved method is proposed to compute the distance of the moving object using a monocular camera and based on the feature extraction method. First, the specific object is tracked by a camera and then the distance from the object to the camera is measured using a computationally efficient algorithm that is implemented in MATLAB/ SIMULINK environment. Experimental results show that this method is fast, efficient, and accurate and can be used for real-time robotic applications.

KEY WORDS
Desired Object Tracking; Distance Measurement; Image Processing; Single Camera.

1. Introduction
For decades, researchers have tried to develop different techniques of relating computer vision capabilities to applications and products, such as automotive safety, manufacturing, video surveillance, and visual servoing (one of the important related robotics applications). Visual servoing is a technique which controls the motion of the robot by using feedback information that is sent from a vision sensor. In visual servoing, it is difficult to track the moving object if the object distance is not accessible [1]. There are two different types of visual servo control: position-based visual servo control (PBVS) and image-based visual servo control (IBVS). The reference inputs for PBVS are the 3D relative position and orientation between the object and the robot end-effector in Cartesian space. In the image-based visual servo control, the reference input is the 2D object position which can be obtained from the image plane. The IBVS approach fails without having an accurate estimation of object distance and motion, especially when it is carried out in dynamic environments [2]. Because of error propagation, when a camera is employed in PBVS servo method, the small measurement error would highly affect the servoing accuracy [2].

To estimate the distance from the object, the tracking of an object must be addressed first. Object tracking can be classified into four main categories: model-based, appearance-based (region-based), feature-based, and contour-based methods. Model-based tracking methods need previous knowledge of the object shapes for the matching process and in order to find the object in the scene; Model-based tracking methods apply the exact geometrical models of the object. Model-based tracking methods have two shortcomings: 1) an object that is not in the database cannot be recognized by using these models. And 2) establishing this model is complicated and sometimes impossible [3]. Appearance-based methods track an object using the 2-D shape of the connected region. This tracking approach relies upon the information that is provided by the entire region pixels. Motion, color, and texture are examples of such information. These methods are not robust with complex deformation [4]. Feature-based methods track the specific features of each object. Many feature-based matching techniques have been developed although these techniques are not specialized for video object tracking. Advantages of these methods are their simplicity and stability for tracking algorithms. However, clustering of these features to the same group of an object is costly, specifically in dealing with a large number of features. Contour-based methods track the contour of the object rather than tracking the whole pixels that make up the object. In contour-based methods, the contour for the next frame is determined using the motion information about the object. Thus, the shape and the position of that contour are enhanced to fit into the object. Furthermore, the motion information is updated by the change in the contour location [3]. There are some advantages for using the contour-based representation. Merit of such methods is that they can cut the computational complexity. Another advantage of such techniques is their ability to track rigid and non-rigid objects. However, the contour-based methods fail to track the objects that are partly occluded [4].

To measure the distance of a moving object several methods presented in the literature. Zhang et al. [5] presented the 3D positions of a target in the camera coordinate frame by measuring the distance between feature point and principal point based on the calculated area in the image. The Zhang et al. [5] process is divided into three stages. The first stage is the camera calibration in order to calibrate the intrinsic parameters. The second stage is to set up a model for distance measurement along the optical axis direction according to the mapping relationship between the
targets in the camera coordinates frame, and its projection in
the pixel coordinate frame. The last stage is the absolute
distance measurement. Yang and Cao [6] proposed a 6-D
object localization method based on a monocular vision
system which is obtained through decomposing the
homography matrix and finally refining the result using the
Levenberg–Marquardt algorithm. This approach showed an
absolute accuracy of 0.3 mm and the repeatability accuracy
of 0.1 mm for their robot’s arm pose estimation. Coman and
Balan [7] mentioned that their application of object distance
measurement can be a starting point of complex applications
as long as the object geometry remains a square. Coman and
Balan [7] also indicated that the surface area and a line in
the surface area for different objects are correlated to each
other. Yamaguti et al. [8] found the distance to the texture
surface of the object from the ratio of two images that are
taken by a monocular camera at different locations, and then
they used complex log mapping of these two images. In their
method, the processing time is long and that the object is
restricted to a plane which is vertical on the optical axis.
Zhang et al. [9] proposed a method of object distance
measurement based on a particle filter. This method is
reported to be highly reliable and have strong real-time
performance. However, the accuracy of the technique
proposed by Zhang et al. [9] is not mentioned. Dixon et al.
[10] presented a new continuous observer for determining
the range information of an object moving with affine
motion dynamics and known motion parameters using one
camera. The impact of this experiment was that a continuous
observer can be employed to make a monocular vision
system in order to determine the range parameter for a
moving object which moved with an affine or Riccati
motion dynamics even in presence of sensor noise.

In the present research, the distance to the object is
calculated using a feature extraction method. The proposed
algorithm is robust and reliable with respect to the
orientation and position of the object and could be used for
any moving objects of different size, and in any direction on
the ground.

The remainder of this paper is organized as follows:
Section 2 presents the problem formulation, Section 3
describes the Image-processing algorithm, experimental
setup is explained in Section 4, the experimental results
and detailed discussion are also presented in Section 5, and
Section 6 concludes the article.

2. Problem Definition

The main image formation technique is mostly based on
the pinhole lens model. In this model, the mapping
relationship between the 3-D world coordinates and the 2-D
image coordinates plays an important role. Let’s consider M
in Figure 1 to be a point in the world coordinate relative to
the camera frame. In this case, m would be the projection of
M onto the image plane. Under the pinhole assumption,
these two points and the origin of the camera frame (the
center of the projection) that is in the back of the image
plane at a distance µ will be collinear [11]. This model of
image formation is applied in the present research. Now, the
problem of mapping the object’s location is described using
measured distance and frame transformation.

For visual servoing purposes, there are five “standard”
frame names associated with a robot and its workspace: the
base frame, the station frame, the wrist frame, the tool
frame, and the goal frame. The base frame is shown as \( \{B\} \),
and it is the fixed part of the robot which is sometimes called
the Link 0. The station frame \( \{S\} \) is called the universe
frames since all actions of the robot are made corresponding
to this frame. The wrist frame \( \{W\} \) is the last link of the
manipulator, and it is normally defined relative to the base
frame. The tool frame \( \{T\} \) is assigned at the end of any tool
that a robot is holding. Finally, the goal frame \( \{G\} \) is the
location to which the tools need to move [12]. Figure 2
shows entire frame assignments to find the distance of the
moving object with respect to a robot hand. In Figure 2, the
object distance \( AG \) can be measured through the proposed
algorithm in the present paper. The camera’s location w.r.t.
base frame is shown with vector \( BA \) (known by
measurement), and \( BT \) is obtained through forward
kinematics and using DH parameter of the robot. Using this
information, it will be easy to find vector \( TG \), which is the
distance of the moving object with respect to the tool frame
of the robot, and can be calculated by equation (1).

![Figure 1. Camera coordinates frame](image1)

![Figure 2. Standard frame assignment](image2)
There are three ways of using image based distance computation techniques: 1) Stereo vision based technique; 2) Mono vision based technique, and 3) Time-of-Flight camera technique. The stereo vision based method uses two cameras to find the depth and the disparity map using a complex method. This technique is highly accurate, but it is time consuming because of many images of the same object needed to be processed at the same time. In addition, implementing such a technique is expensive as it requires two cameras. Moreover, the stereo vision accuracy fails with the increase of the distance to the object compared to the baseline distance between the two different views. On the other hand, mono vision based approach are comparatively less expensive as it requires only one camera[13]. The Time-of-Flight depth estimation technique is used to find the depth information by measuring the total time required in the light to transmit and reflect from the object. It is hard to separate the incoming signal since the signal depends on many parameters such as intensity of the reflected light, intensity of the background light, and the dynamic range of the sensor which makes detecting the incoming light a hard task [3]. Besides, the accurate range of distance measurements for the time-of-flight camera is usually between 1–4 m [14].

Therefore, researchers are looking for inexpensive, uncomplicated, and accurate techniques. Applying such superior techniques requires researchers to tackle with several other challenges such as object detection, obstacle avoidance, and location finding. There are two approaches to estimate the location of any object: contact and non-contact methods. Nowadays, the non-contact distance measurement system becomes useful in a wide variety of applications, though, most of the time it is impossible to have physical contact with the object to make the measurement [15]. High accuracy and time saving are advantages of using a non-contact measurement technique. Clarke and Williams [16] discuss the benefits of using a non-contact measurement technique. The Image-Processing Algorithm In this section, some parts of the image-processing algorithm are explained. The color space conversion is used to change the color information to different color spaces. Figure 4(a) demonstrates the R. G. B. - three primary color of red, green and blue- image. Auto threshold technique is applied in the system in order to identify different objects in the camera’s field of view. The Otsu method [17] is one of the widely referenced threshold techniques. As a result, this method would be satisfactory especially when dealing with a large number of pixels within each class with close pixel’s value [17]. The image of a sample threshold by the auto-threshold technique is given in Figure 4(b).

\[
\begin{align*}
BA + AG &= BG \\
BT + TG &= BG \\
TG &= BG - BT
\end{align*}
\]

(1)

\[
\begin{align*}
\alpha &= \arctan\left(\frac{480 - y_2}{x_2 - 320}\right) \\
\theta &= \frac{\pi}{2} + \alpha \\
b &= \sqrt{(320 - x_2)^2 + (480 - y_2)^2} \\
c &= \sqrt{a^2 + b^2 - 2ab\cos(\theta)} \\
d &= c^2 + h^2
\end{align*}
\]

(2)

3. The Image-Processing Algorithm

In this section, some parts of the image-processing algorithm are explained. The color space conversion is used to change the color information to different color spaces. Figure 4(a) demonstrates the R. G. B. - three primary color of red, green and blue- image. Auto threshold technique is applied in the system in order to identify different objects in the camera’s field of view. The Otsu method [17] is one of the widely referenced threshold techniques. As a result, this method would be satisfactory especially when dealing with a large number of pixels within each class with close pixel’s value [17]. The image of a sample threshold by the auto-threshold technique is given in Figure 4(b).

3.1 2-D FIR Filter Filtering is another important technique that has to be accurately chosen to remove the noise from the image without reducing its sharpness.
3.2 Image Complement

Image complement is used to compute the complement of a binary or intensity image. The image complements switches the values of the zero and one valued pixels.

3.3 Averaging Subsystem

The averaging subsystem method is used to stabilize the system by applying four consecutive measurement values for the system. This technique would mediate the fluctuation arising from the noise into the system [18]. The completed algorithm to the system is shown in Figure 5.

![Image Complement](image1.png)

Figure 5. The Algorithm for object distance measurement

4. Experimental Setup

In this research, Logitech Quick Cam® Communicate STX™ is used to get the R. G. B. image with resolution of 640 × 480 pixels. The captured image is then converted to intensity by color space conversion since most of the applications require an intensity of the image. To reduce the noise from the image of the desired object, a 2-D FIR Filter is applied. MATLAB Functions are used to carry out the required code in the subsystem for the desired object tracking algorithm. In the next step which is implemented in the subsystem, desired object distance is measured using the resulting image data. From implementing the object distance measurements that are shown as \(d_1\) and \(d_2\) for moving object at time \(t_0\) and \(t\) in Figure 6, the horizontal distance traveled by the object on the ground could be calculated using the law of cosines suggested in Equation (3). The initial values are \(d_1=1160\) mm and \(\theta_0=55^\circ\).

\[
d_2^2 = d_1^2 + d^2 + 2d_1d_2 \cos(\theta) = 0
\]

\[
d = \frac{2d_1d_2 \cos(\theta) \pm \sqrt{4d_1^2 \cos^2(\theta) - 4(d_1^2 - d_2^2)}}{2}
\]  

(3)

5. Experimental Result and Discussion

In this experiment, the distance of the desired object is extracted and used for the second subsystem to find the position and orientation of the end-effector with respect to the object. The tracked object is shown with green bounding box in Figure 4(a). Tables 1 and 2 compare the results from the experiment. These tables respectively show the object distance average errors for two different samples of one rectangular block of size 25.43 mm×79.22 mm (W×L) and one irregular toy car of size 70.21 mm ×145.54 mm (W×L) which are measured at any location in the camera’s field of view. In addition, Figures 7 and 8 show the tabulated results in Tables 1 and 2 in the graphical form. Some related works for object distance measurement are shown in Table 3. It should be mentioned that this experiment is implemented with the low resolution camera to check the proposed algorithm. The accuracy of the measurement could improve by using a high-resolution camera.

Image noise is the main unavoidable reason for producing errors during the image acquisition stage. Such error can occur in finding the exact point of contact with the object on the ground. Another potential cause of error would be the variation in image illumination across the camera’s field of view. Using the method presented by [19], it is possible to measure an object distance when the optical axis of the camera is parallel to the ground. However, in case of dealing with finding object distance anywhere in the field of view of camera, Joglekar et al [20] method is found to be the possible solution.

Joglekar et al [20] define the in-path object distance and the oblique object distance as the distance to the object that is on the optical axis and as the distance from the object that is not on the optical axis, respectively.
Table 1
Average distance error measurements for the rectangular block

<table>
<thead>
<tr>
<th>Trial</th>
<th>Real measured distance (mm)</th>
<th>Proposed measured distance (mm)</th>
<th>Joglekar et al [20] measured distance (mm)</th>
<th>Proposed distance error (mm)</th>
<th>Joglekar et al [20] distance error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1181</td>
<td>1176</td>
<td>1186</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1213</td>
<td>1231</td>
<td>1131</td>
<td>18</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>1215</td>
<td>1190</td>
<td>1150</td>
<td>25</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>1272</td>
<td>1302</td>
<td>1622</td>
<td>30</td>
<td>350</td>
</tr>
<tr>
<td>5</td>
<td>1320</td>
<td>1340</td>
<td>1134</td>
<td>20</td>
<td>186</td>
</tr>
<tr>
<td>6</td>
<td>1323</td>
<td>1320</td>
<td>1399</td>
<td>3</td>
<td>76</td>
</tr>
<tr>
<td>7</td>
<td>1367</td>
<td>1414</td>
<td>2275</td>
<td>47</td>
<td>908</td>
</tr>
<tr>
<td>8</td>
<td>1432</td>
<td>1438</td>
<td>1424</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>1436</td>
<td>1468</td>
<td>1209</td>
<td>32</td>
<td>227</td>
</tr>
<tr>
<td>10</td>
<td>1481</td>
<td>1529</td>
<td>1433</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>11</td>
<td>1513</td>
<td>1521</td>
<td>1206</td>
<td>8</td>
<td>207</td>
</tr>
<tr>
<td>12</td>
<td>1548</td>
<td>1380</td>
<td>1183</td>
<td>32</td>
<td>365</td>
</tr>
<tr>
<td>13</td>
<td>1585</td>
<td>1617</td>
<td>1227</td>
<td>32</td>
<td>358</td>
</tr>
<tr>
<td>14</td>
<td>1621</td>
<td>1633</td>
<td>1156</td>
<td>12</td>
<td>465</td>
</tr>
<tr>
<td>15</td>
<td>1630</td>
<td>1615</td>
<td>1200</td>
<td>15</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Average absolute error</strong></td>
<td><strong>23.43</strong></td>
<td><strong>269.64</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Average distance error measurements for the toy car

<table>
<thead>
<tr>
<th>Trial</th>
<th>Real measured distance (mm)</th>
<th>Proposed measured distance (mm)</th>
<th>Joglekar et al [20] measured distance (mm)</th>
<th>Proposed distance error (mm)</th>
<th>Joglekar et al [20] distance error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1179</td>
<td>1173</td>
<td>1184</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1206</td>
<td>1183</td>
<td>1160</td>
<td>23</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>1250</td>
<td>1262</td>
<td>1155</td>
<td>12</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>1253</td>
<td>1259</td>
<td>1204</td>
<td>6</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>1263</td>
<td>1238</td>
<td>1196</td>
<td>25</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>1295</td>
<td>1328</td>
<td>1953</td>
<td>33</td>
<td>658</td>
</tr>
<tr>
<td>7</td>
<td>1363</td>
<td>1352</td>
<td>1341</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>1373</td>
<td>1404</td>
<td>1196</td>
<td>31</td>
<td>177</td>
</tr>
<tr>
<td>9</td>
<td>1420</td>
<td>1458</td>
<td>2057</td>
<td>38</td>
<td>637</td>
</tr>
<tr>
<td>10</td>
<td>1480</td>
<td>1489</td>
<td>1344</td>
<td>9</td>
<td>136</td>
</tr>
<tr>
<td>11</td>
<td>1495</td>
<td>1542</td>
<td>1370</td>
<td>47</td>
<td>125</td>
</tr>
<tr>
<td>12</td>
<td>1503</td>
<td>1529</td>
<td>1202</td>
<td>26</td>
<td>301</td>
</tr>
<tr>
<td>13</td>
<td>1527</td>
<td>1570</td>
<td>1310</td>
<td>43</td>
<td>217</td>
</tr>
<tr>
<td>14</td>
<td>1580</td>
<td>1583</td>
<td>1236</td>
<td>3</td>
<td>344</td>
</tr>
<tr>
<td>15</td>
<td>1583</td>
<td>1613</td>
<td>1164</td>
<td>30</td>
<td>419</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Average absolute error</strong></td>
<td><strong>16.50</strong></td>
<td><strong>381.50</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We have investigated whether this method can be used to estimate the object distance anywhere on the camera’s field of view with the provided setup. The results from some experiments using the Joglekar et al [20] method shows that as the object gets closer to the vanishing point, the object distance would surprisingly rise, and the error would significantly increase. Therefore, when the camera has an angle with optical axis, [19-20]’s methods are no longer valid as the object distance measurement would be a non-linear function of distance. Moreover, due to the fact that the oblique object distance is calculated from the result of the in-path object distance divided by the cosine angle between the optical axis and the line joining the camera point to the object’s point of contact. Hence, the error is relatively similar for the in-path object distance as obtained. The proposed vision-based object distance measurement technique is principally different from above-mentioned optical techniques.

Table 3
Comparison of different methods for object distance measurement

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics (in terms of distance)</th>
<th>Nature of the method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gat et al [19]</td>
<td>Along the optical axis</td>
<td>Linear</td>
</tr>
<tr>
<td>(Optical axis is parallel to the ground)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joglekar et al [20]</td>
<td>In-path and oblique</td>
<td>Non-Linear</td>
</tr>
<tr>
<td>(Optical axis is parallel to the ground)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proposed Method (Optical axis is not parallel to the ground) | Any-where on the field of view of camera | Non-Linear

Figure 7. Object distance measurement for the rectangular block
6. Conclusion

In this paper, an improved method was proposed to calculate the object distance using a single camera even if the surface of an object is not parallel to the camera and is not restricted to be vertically intersecting the optical axis. The proposed method is able to identify the desired object and extract the object feature for a moving and a static object. The experimental results show that the object distance average error for the rectangular block, and the irregular toy car are 23.43 mm and 16.50 mm, respectively. Moreover, the discrete sample time to run the proposed algorithm is estimated to be 0.033. This inexpensive distance measurement method could be applied to many applications such as pick and place task in robotics.

References


