# Tracking, Recognizing, and Estimating Size of Objects Using Adaptive Technique

Fazal Noor[1(✉)] and Majed Alhaisoni[2]

[1] Islamic University of Madinah, Madinah, Kingdom of Saudi Arabia
mfnoor@gmail.com
[2] University of Hail, Hail, Kingdom of Saudi Arabia
majed.alhaisoni@gmail.com

**Abstract.** The detection and tracking of object in a video is an important problem in many applications. In surveillance and in robotic vision tracking and recognition of objects and it's size is desired. In this paper, an algorithm to obtain size of an object in image or video is presented based on pixel relationship to actual size. The object is mainly tracked by the Kalman filter and Log Polar Phase Correlation method is used to more precisely recognize objects in a video. The tracking of objects is performed from frame to frame. As the image of an object gets deformed in a video due to motion of either the camera or the motion of an object a dynamic template for matching is proposed to minimize the error. Simulation results are presented showing the errors in determining the size of objects in an image.

**Keywords:** Arduino microcontroller · Object size · Kalman filter · Log-polar phase correlation · Robotics · Sonar

## 1 Introduction

In robotics, tracking and recognizing objects is a common task [1]. The Kalman filter has been used in numerous applications and in object detection and tracking [2–5]. Similarly, in self driven cars computer vision object tracking requires processing in real-time [6–8]. Many different types of sensors are used in the self-driven cars, each having their own capabilities and limitations. Sensors such as Sonar, LiDAR are used to measure distances and mapping of the surroundings very precisely [8]. In certain applications, robots need to know an object's size, such as width and height [13]. In case an object is stationary it is less difficult to calculate the object's dimensions, however, in case when an object is moving it is very challenging to calculate an object's size. This is so because as the camera moves or as the object moves in the field of view, the object's image changes or deforms. Recognizing an object may help in determining an object's original size. However, there are situations in which prior information about the object is unknown and therefore determining objects dimensions from tracking in a video becomes even more challenging. Also, there are situations when a moving object of a specific dimension has to pass through pathways openings of another dimension so care must be taken not to collide if the pathway opening is smaller.


M. H. Miraz et al. (Eds.): iCETiC 2019, LNICST 285, pp. 283–291, 2019.
https://doi.org/10.1007/978-3-030-23943-5_22

In this paper a reference image with known dimensions is used within a video to obtain dimensions of unknown objects. The number of pixels of an object is camera dependent and varies with precise distance to an object. Practical experiments are performed using Arduino microcontroller, with camera and sonar assembled to it.

The paper is presented as follows, in Sect. 2, the Kalman filter, Log polar phase correlation, and object size determination methods are presented. In Sect. 3, experimental setup of hardware is presented and in Sect. 4 results and discussions are presented and last in Sect. 5, conclusion and future directions are discussed.

## 2 Methods

### 2.1 Kalman Filter

In 1960, Rudolph E. Kalman published a paper describing a recursive solution using state space methods, later called a Kalman filter [2–5]. Many applications use Kalman filtering such as tracking, navigation, and estimation. Given a linear stochastic difference equation shown as,

$$\hat{x}_{t|t-1} = A_t \hat{x}_{t|t-1} + B_t u_t + v_{t-1} \tag{1}$$

and a measurement equation given by

$$z_t = H x_t + w_t \tag{2}$$

where A is defined as the state transition matrix, B is defined as a control matrix, H is the measurement matrix, $\hat{x}$ is the estimated state, and u is the control variables. The variables w, and v are assumed to be random, independently distributed with Gaussian white noise. The variables w, and v denote the process and measurement noise, respectively. Assuming the time dependent matrices A, B, and H to be constant, the two part Kalman filter can be written in two parts, the predictor and measurement as shown below.

Predict (process time update equations):

$$\hat{x}_{t|t-1} = A_t \hat{x}_{t|t-1} + B_t u_t \tag{3}$$

$$P_{t|t-1} = A_t P_{t-1|t-1} A_t^T + Q_t \tag{4}$$

The predicted state is given by Eq. (3) and the predicted estimate covariance is given by Eq. (4). Update (measurement or correction update equations):

Feedback

$$\hat{z}_t = y_t - H_t \hat{x}_{t|t-1} \tag{5}$$

$$S_t = H_t P_{t|t-1} H_k^T + R_t \tag{6}$$

$$K_t = P_{t|t-1} H_t^T S_t^{-1} \quad (7)$$

$$\hat{x}_{t|t-1} = \hat{x}_{t-1|t-1} + K_t \hat{z}_t \quad (8)$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} \quad (9)$$

where in Eq. (5), $\hat{z}_t$ denotes the measurement residual and y denotes the measurement variables. In Eq. (6), $S$ gives the innovation covariance and R is the measurement variance matrix. In Eq. (7), $K_t$ is the Kalman gain, P is defined as the state variance matrix. In Eq. (8), $\hat{x}_{t|t-1}$ is the updated (a posteriori) state estimate and in Eq. (9), $P_{t|t}$ is the updated (a posteriori) estimate covariance.

## 2.2 Log Polar Phase Correlation

A powerful signal processing technique is the correlation filter and has been used in object detection and registration problems. The use of correlation filter appears in numerous applications such as signal matching, automatic target detection, missile guidance systems, in medical imaging, and many others [4, 5]. The basic idea of the phase correlation method is to see the similarity between two images. Phase correlation is used in methods such as motion estimation, image registration, object tracking by template matching. In particular the log polar phase correlation technique has been shown to be robust to image scaling, rotation, and translation and provides a good solution to image recognition.

Suppose f(x, y) represents a scaled, translated, and rotated version of an original image h(x, y), given by [4, 5, 12],

$$\begin{aligned} f(x,y) = h(&s \cdot x \cos\theta - s \cdot y \sin\theta - x_0, \\ &s \cdot x \sin\theta + s \cdot y \cos\theta - y_0) \end{aligned} \quad (10)$$

After taking the Fourier transform of f(x, y) transforms it in frequency domain as,

$$F(u,v) = \frac{1}{|s^2|} H(u' \cos\theta - v' \sin\theta, u' \sin\theta + v' \cos\theta) \rightleftarrows \cdot e^{-j(ux_0 + vy_0)} \quad (11)$$

where u′ = u/s and v′ = v/s. Taking the magnitude of F(u, v) results in the following,

$$M_F(u,v) = w \cdot M_H(u' \cos\theta - v' \sin\theta, u' \cos\theta + v' \sin\theta) \quad (12)$$

where $w$ is defined as a weighting factor. Utilizing the polar coordinates $(r, \varphi)$ and using the relation $u = r\,cos\,\varphi$ and $v = r\,sin\,\varphi$. Inserting in Eq. (12) above will result in,

$$\begin{aligned} M_F(u,v) =& w \cdot M_H(r'\,cos\,\varphi\,cos\,\theta - r'\,sin\,\varphi\,sin\,\theta, \\ & r'\,cos\,\phi\,sin\,\theta + r'\,sin\,\phi\,cos\,\theta)\cdot \end{aligned} \tag{13}$$

and using trigonometric identities results in the following,

$$M_F(u,v) = w \cdot M_H(r'\,cos(\varphi+\theta) - r'\,sin\,(\varphi+\theta)) \tag{14}$$

which can be denoted by,

$$M_F(r,\varphi) = w \cdot M_H(r', \varphi+\theta) \tag{15}$$

Transforming the above Eq. (15) to log-polar form, by using the logarithm results in

$$M_F(log\,r,\varphi) = w \cdot M_H(log\,r', \varphi+\theta) \tag{16}$$

where $r' = r/s$ and $\log r' = \log r - \log s$.

It is seen by the use of the Fourier properties, a phase shift is produced by a positional shift. Also, a linear scaling of the variables x, and y causes an inverse scaling of the spatial frequencies u and v. Using the log-polar transform and applying it to the magnitude spectrum, the scale and rotation are obtained by using the phase correlation. This is possible because the scaling and rotation in the Cartesian system become pure translation in the log-polar system.

## 2.3 Variation of Objects with Distance

The object appearance changes due to angle of viewpoint, change in shape, occlusion, daylight, etc. [7, 9, 10, 11]. For example, as the road sign comes in view of a camera the image of it changes with relationship to motion, becoming larger as camera is moved toward it and smaller as the camera moves out [10]. The size of a slanted image is more involved and needs the angle of camera or angle of object or both to determine the size of an object [9]. For an object forming a 90° angle to a camera, the size relationship of an object is:

$$\text{Size} = \text{Object size in pixels}/(\text{Number of pixels/cm}) \tag{17}$$

Where pixels/cm is obtained by Object size/Actual size. The pixels/cm in the equation is Camera dependent and needs to be determined beforehand, the distance to an object can be determined using either ultrasonic sensors or LiDAR [8]. A known object size can be inserted in an image and use it to calculate the pixels/cm relationship (Fig. 1).

The algorithm to determine the size of an object from a video can be summarized as follows:

**Algorithm:**

**Preliminary Steps: Calibrate the pixel relationship to actual object size**.

1. Assume original object dimensions are known for comparison and determining the error.
2. Given a video frame.
3. Detect an object of interest.
4. Calibrate the relation of object size to number of pixels in frame with distance obtained by sonar. Move object back and forth and left and right for obtaining the precise relationship of object scaling to number of pixels.

**Track, freeze frame, perform log polar phase correlation**.

5. Keep track of the object using Kalman filter, predicting the object's position.
6. Freeze the frame and use log polar phase correlation to recognize the object by matching a reference template of an object. If match is greater than a threshold then object is recognized else if match is less than threshold then object is not recognized.
7. If object is detected and recognized, approximate the dimensions using the pixel relation.
8. Compare the calculated size with original size to obtain the error.

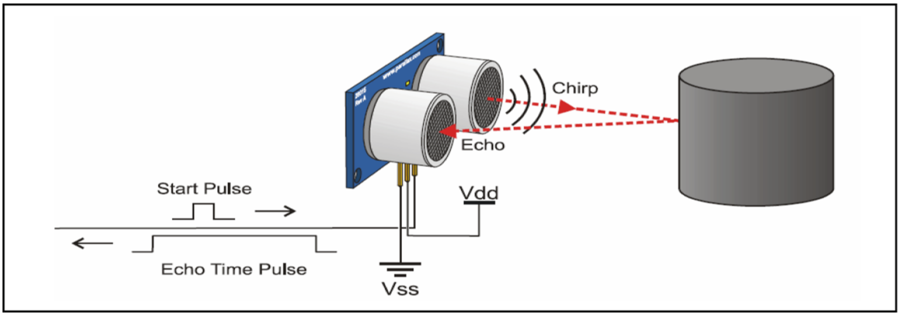



**Fig. 1.** Using sonar to measure the distance to an object.

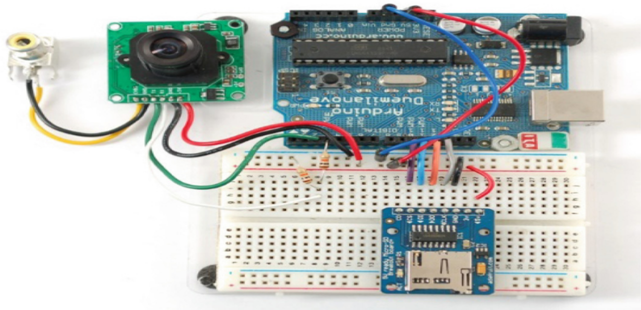


**Fig. 2.** Arduino microcontroller with a camera and sd card connected with wiring.

## 3 Experiment Setup

An Arduino micro-controller board with a camera, sonar, motor, and a sd card were assembled in the experiment setups. The sonar measures the range to an object, while the camera is used to take a video. Matlab 2017 was used to write a script to capture video and the Kalman Filter was used to track a moving object while the Phase correlation method was used to match an object within the image of a frame video. The algorithm uses the size to pixel relationship to calculate the size of an object and is camera dependent.

Figure 2 shows a sd card connection and a camera assembled to an Arduino microcontroller. The software is downloaded from the pc usb port via wire connection to the board.

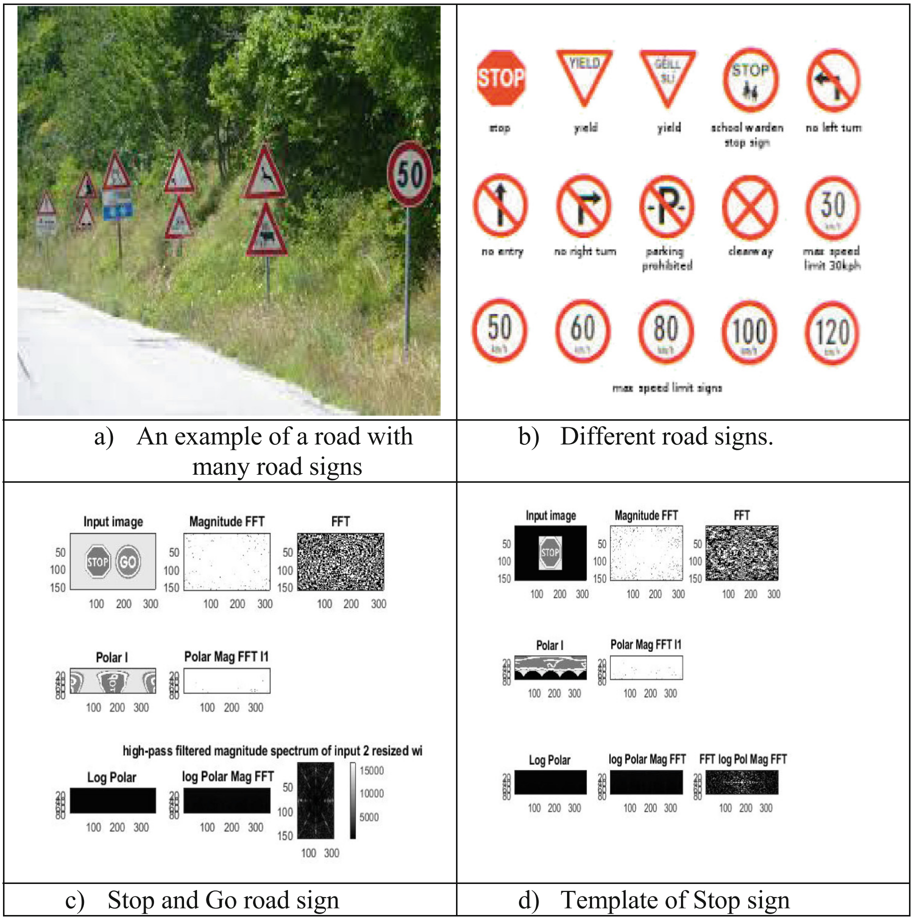
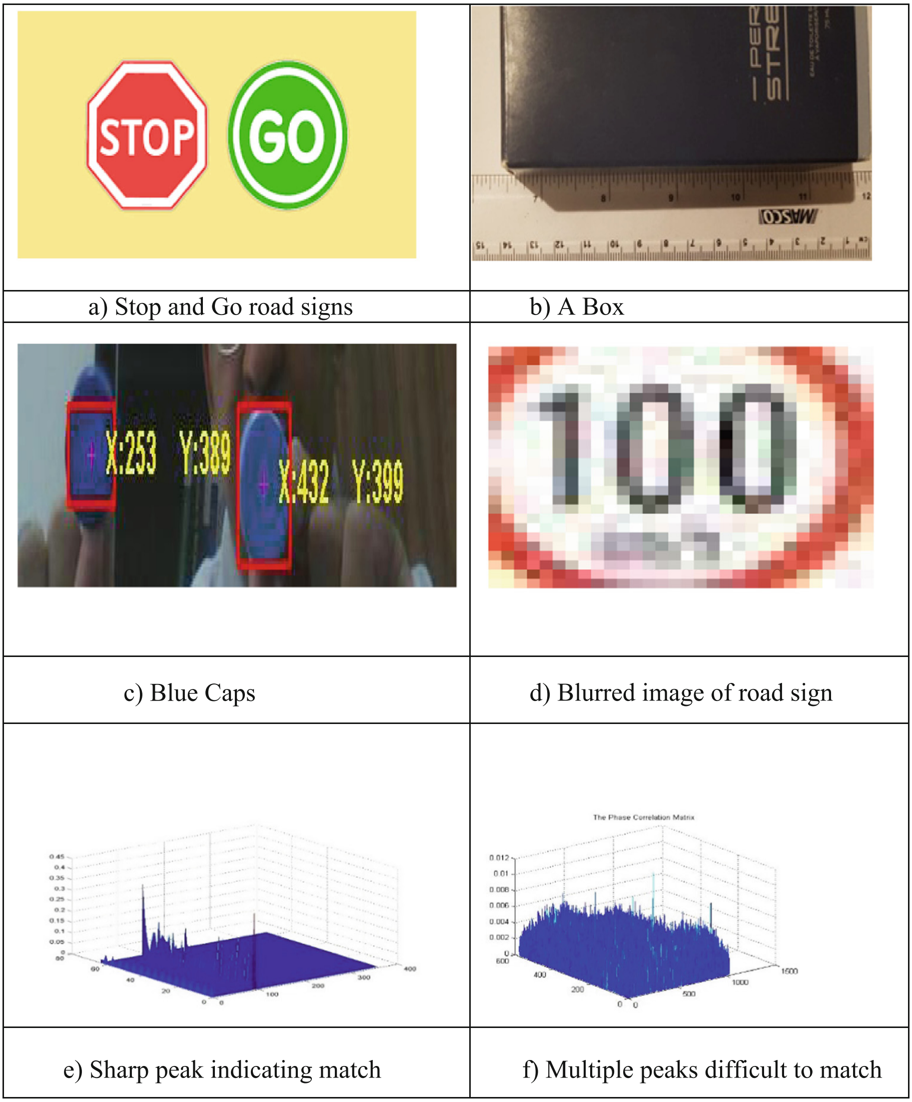



**Fig. 3.** Road signs and template matching using LPC.

# 4 Results and Discussion

## 4.1 Moving Camera with Stationary Objects

In this scenario, a car is mounted with a camera and road signs are detected under clear weather (not foggy weather) while the car is moving. As the video is captured, in real time the road signs are tracked from frame to frame. Ideally, an angle of 90° would give the best results for recognition, however practically it is not possible when the vehicle




a) Stop and Go road signs; b) A Box; c) Blue Caps; d) Blurred image of road sign; e) Sharp peak indicating match; f) Multiple peaks difficult to match

**Fig. 4.** Object size calculation and image matching.

is moving. Using the log polar phase correlation method discussed above results in good recognition when the sign is at a 90° angle. However, in case the camera is in motion the view of the road signs will appear to change with distance and angle. Objects will appear to scale, and/or rotate. The log polar phase correlation is robust to image scaling and rotation. Figure 3 shows the result of LPPC for a 90 snapshot of with image template with a scaled and rotated version. Figure 3a, b shows an image with many road signs and objective is to detect them and recognize them (Fig. 4 and Table 1).

**Table 1.** Comparison of objects size between actual width and calculated width.

| Object | Direction | Distance (cm) | Actual width | Calculated width | Percent error% |
|---|---|---|---|---|---|
| Stop sign | 90 | 15 | 5.5 cm | 5.3 cm | 3.63 |
| Box | 90 | 15 | 14 cm | 14.08 cm | 0.57 |
| Cap | 90 | 15 | 2.5 cm | 2.4 cm | 4.0 |

## 5 Conclusion and Future Work

In this paper, we have shown how to detect, track, recognize, and find the size of an object. The Kalman filter is very useful in predicting the movement of an object and the log polar phase correlation is robust to object scaling, translation, and rotation. The size of an object is calculated by knowing the image scaling relationship due to camera properties and a reference image. Also the distance and angle can be used to calculate the size of an object. In our approach, a known image is inserted in an image to use to determine the size of other unknown objects. As the camera moves toward an object, the image becomes larger therefore there is an increase in number of pixels of object's image. As the camera moves away from an object its image in number of pixels decreases. Knowing the precise relationship between actual size and dimensions of image in terms of number of pixels the size of an object can be determined. The size accuracy is fairly good when an object is moving back or forth with 90° angle with camera view point and deteriorates as the image becomes slanted. A good example of image slant is when viewing road signs from a moving car.

The size of a moving object is determined by tracking it from frame to frame. As the object's image changes due to motion, the tracking is done by correspondingly using an adaptive template (i.e. a dynamic template updated every N frames or depending on some threshold).

Future work involves work for cases of object slant and for feasibility of the method in real-time applications.

**Acknowledgement.** The authors would like to thank FCIS and the Deanship of Research at the Islamic University of Madinah and CSSE at the University of Hail for their support.

## References


1. Vargas, H., et al.: Human-robot interface using face detection and recognition, for the service robot, Donaxi. In: Workshop Proceedings of the 8th International Conference on Intelligent Environments. IOS Press (2012)
2. Raja, T.V., Tirupathamma, M.: Object detection and tracking in video using Kalman filter. Int. J. Res. Adv. Eng. Technol. **5**(5), 102–106 (2016)
3. Saravanan, S., Parthasarathy, K.A.: Video object detection and tracking using Kalman filter and color histogram-based matching algorithm. Int. J. Eng. Res. Dev. **9**(7), 31–39 (2014). www.ijerd.com. e-ISSN: 2278-067X, p-ISSN: 2278-800X
4. Jain, R.: Phase Correlation Algorithm for Video Tracking: Phase Correlation Algorithm with Kalman Filtering. LAP LAMBERT Academic Publishing (2012). ISBN 10 9783659210730. ISBN 13 978-3659210730
5. Noor, F., Alhaisoni, M.: Distinguishing moving objects using Kalman filter and phase correlation methods. In: 17th IEEE International Multi-topic Conference, Karachi, Pakistan, December 2014
6. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: 2000 Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 142–149 (2000)
7. Object detection and tracking in video. https://towardsdatascience.com/object-detection-and-tracking-in-pytorch-b3cf1a696a98. Accessed 19 Apr 2019
8. Hwang, S., Kim, N., Choi, Y., Lee, S., Kweon, I.S.: Fast multiple objects detection and tracking fusing color camera and 3D lidar for intelligent vehicles. In: Proceedings of 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 234–239 (2016)
9. Hou, L., Wan, W., Lee, K.-H., Hwang, J.-N., Okopal, G., Pitton, J.: Deformable multiple-kernel based human tracking using a moving camera. In: Proceedings of International Conference Acoustics, Speech and Signal Processing (ICASSP), pp. 2249–2253 (2015)
10. Sankaranarayanan et al.: Object detection, tracking, and recognition for multiple smart cameras. In: Proceedings of the IEEE, vol. 96, no. 10, October 2008
11. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
12. Srinivasa Reddy, B., Chatterji, B.N.: An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Trans. Image Process. **5**(8), 1266–1271 (1996)
13. Noor, F., Swaied, M., AlMesned, M., AlMuzini, N.: A method to detect an object's width using an ultrasonic sonar. In: Proceedings of IEEE ICCECE 2018 Conference, UK (2018)