CamMirror: Single-Camera-based Distance Estimation for Physical Analytics Applications

Vivek Yenamandra[†*], Akshay Uttama Nambi S.N.[‡], Venkata N. Padmanabhan[‡], Vishnu Navda[‡], Kannan Srinivasan[†]
[†]The Ohio State University [‡]Microsoft Research India

ABSTRACT

Distance estimation is key to many physical analytics applications in settings such as driving, shopping, and more. The goal is to tell where an object or person is. While specialized sensors such LIDAR and stereoscopic cameras can solve the problem, these tend to be expensive. In this paper, we present CamMirror, which performs distance estimation with a single camera. The key idea is to use a pair of carefully-positioned mirrors to provide a second view of the scene akin to what a second camera would have provided, which then enables disparity-based ranging. We present the design of CamMirror and two applications, one on vehicle ranging and the other on smart shelf.

Keywords

Computer vision; disparity-based ranging; physical analytics applications

1. INTRODUCTION

Distance estimation, or ranging, is key to many physical analytics application. For example, in a driving safety context, knowing how far a vehicle is from the vehicle in front would help detect and flag tailgating. In a smart shelf in a retail setting, knowing how far a user's hand is say from the top would help identify which shelf the user has reached out to pick up an item from.

Such distance estimation is typically performed using specialized sensors such as RADAR [11], LIDAR [1], or stereoscopic cameras [3]. While being accurate, these tend to be expensive. For instance, it will likely be a while before LIDAR is installed in all of our vehicles (including in the developing regions) or stereoscopic cameras mounted atop store shelves.

In this paper, we ask whether it is possible to do distance estimation using a single camera. Specifically, we consider how to enable stereoscopic ranging using a single camera. To this end, we present CamMirror, a novel design which uses a pair of mirrors to serve as a "second eye" for a camera. The mirrors provide a displaced view of the scene just as a second, physically-separate camera would. This enables disparity, or parallax, based ranging.

We first present the design of CamMirror and the tradeoffs involved. For instance, if the mirrors are placed far apart, ranging accuracy would improve, but the effective field of view would narrow. We then present two physical analytics applications that we have prototyped. The first enables ranging in a vehicular setting where we use the back-camera of a dashboard-mounted smartphone to pick out distinctive features such as taillights and match these across the direct and mirror views to calculate disparity and thereby range. The second enables a smart shelf, where a low-cost webcam mounted directly above the front edge of the shelf and looking down is used, along with pair of mirrors, to detect the hand of a user who reaches out for an item and determines where the reaching out happened (i.e., which shelf and where on that shelf).

Several works have studied the use of stereo-vision cameras to identify and track vehicles in front [4,5]. A summary of vision based techniques used for vehicle tracking can be found in [7]. These techniques generally rely on detecting a distinctive feature such as taillights [8, 10]. On the retail sector, deployment of static cameras to view user activities and behavior is well studied [6]. RFID tags and other low cost sensors have been deployed in retail stores to track which items or shelves user is interacting [2,9]. In this work, we show that using just a single camera we can estimate the distance to an object or person. The applicability of CamMirror is demonstrated on two applications, *viz.,* vehicle ranging and smart shelf.

2. DESIGN OF CamMirror

2.1 Overview

We seek a solution that keeps costs low and is broadly applicable. With regard to the latter point, we wish to make minimal assumptions about the object, person, and environment. We do not assume that we know the size of the object of interest (e.g., the car in front). We do not assume that the environment has been prepared with cues, e.g., well-marked, standard-width lanes (which is often absent in the the developing regions). We do not assume a sensor or other artificially-introduced marker has been placed on the object or person of interest (e.g., RFID tags on store shelves).

*The author was an intern at Microsoft Research India during part of this work.

WPA'17, June 19, 2017, Niagara Falls, NY, USA

DOI: http://dx.doi.org/10.1145/3092305.3092312

A vision-based approach approach is attractive since it could potentially satisfy many of the above requirements. In particular, disparity-based ranging, which is based on parallax shift (and is indeed how humans gauge distance), seems attractive since it could work for any "object" (car, bus, two-wheeler, etc. in a vehicular context; a child's arm reaching out to a shelf, an adult's arm enveloped in a full-sleeved jacket, etc. in a shopping context). The size of the object need not be known.

Disparity-based ranging poses two challenges. First, it requires binocular vision, which typically means having two, physically-separated cameras. Second, it requires matching the images obtained through the two cameras to calculate disparity and thereby distance.

To address the first challenge, in CamMirror we devise a novel technique comprising a single camera and a pair of carefully-placed mirrors instead of the second camera. The single camera obtains both a direct view of the scene and a displaced view via the mirrors, thereby enabling disparity-based ranging. To address the second challenge, we take a scenario-specific approach to identify specific features in the image and confine the disparity computation to these.

2.2 Disparity-based Ranging

Estimating the distance of an object using the parallax shift, or *disparity*, between the two camera views is a well-known technique. Indeed, even humans use disparity to estimate distance to objects. Here, we briefly describe the technique. Assume that the user places two identical cameras, as illustrated in Fig. 1. (Section 2.3 discusses how such a placement of cameras could be realized with a single smartphone plus a pair of mirrors.) The cameras are placed horizontally displaced relative to each other, such that their image planes are parallel and they are at the same height. Assume the *baseline* distance between the two cameras is b. Consider an object in front of the cameras that lies within the overlapping field-of-view of the two cameras. Assume this object is recognized as the same by both the cameras. Let the pixel location of the centroid of the detected common object be (x_{left}, y_{left}) and (x_{right}, y_{right}) on the left and the right cameras, respectively. Since the cameras are at the same height, $y_{left} = y_{right}$. Further, because of the arrangement of the cameras, $x_{left} > x_{right}$. Disparity, then, is defined as $\Delta = x_{left} - x_{right}$. Intuitively, the disparity decreases as the distance between the detected object and the image plane of the two cameras increases.

Specifically, using basic trigonometry, the distance between the object and the camera plane can be proved to be

$$d_{disp} = \frac{b \times \frac{x_{max}}{2}}{\Delta \times tan(\frac{\theta}{2})} \tag{1}$$

where θ is the field-of-view of the cameras and x_{max} is the width of the image (in pixels).

2.2.1 Challenges

Poor resolution at farther distances: The challenge of estimating distance using disparity is indicated in Fig. 2. The distance estimation resolution using the disparity based technique is worse the further away an object is. To see this mathematically, we can calculate the derivative:

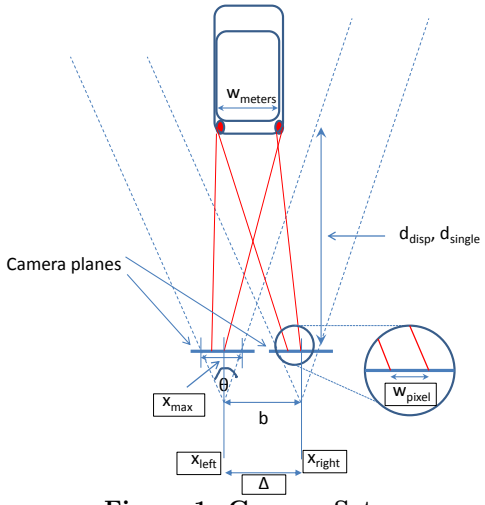

Figure 1: Camera Setup

$$\frac{dd_{disp}}{d\Delta} = -\frac{d_{disp}}{\Delta} \tag{2}$$

Now, the disparity Δ can be expressed as:

$$\Delta = \frac{\frac{b}{d_{disp}}}{tan(\frac{\theta}{2})} \times \frac{x_{max}}{2} \tag{3}$$

Combining equations 2 and 3, we have:

$$\frac{dd_{disp}}{d\Delta} = -\frac{d_{disp}^2 \times tan(\frac{\theta}{2})}{b \times \frac{x_{max}}{2}} \tag{4}$$

For instance, at a distance of $d_{disp} = 10m$, and with a horizontal field of view of the camera of $\theta = 70^o$, a horizontal resolution of $x_{max} = 2000$ pixels, and a baseline separation between the cameras of $b = 33cm = 0.33m$, we obtain a derivative of $\frac{dd_{disp}}{d\Delta} = -0.21m/pixel$. In other words, a 1-pixel change (or estimation error) in the disparity Δ, would result in an error of 0.21m in the distance estimation at a distance of 10m. A 3-pixel error would result in a distance estimation error of over 0.6m.

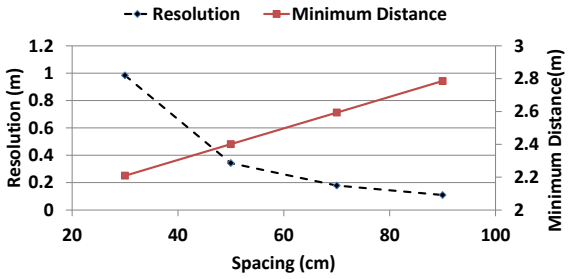

Figure 2: Disparity design tradeoffs: Increasing spacing between the cameras increases resolution at the cost of the minimum distance that can be ranged. (The resolution is plot at a distance of 10m.)

Limit on minimum ranging distance: The resolution of the disparity-based technique can be improved by increasing the baseline distance, b, between the cameras. Indeed,

Equation 4 shows an inverse relationship between the ranging error derivative and the baseline, b. This can also be seen in Figure 2, which shows that increasing the baseline distance between the cameras helps improve the resolution for any given distance. However, increasing the baseline distance shrinks the overlapping region between the two camera views, making it less likely that objects that are close would be seen in both views. In other words, increasing the baseline would also increase the minimum distance at which an obstacle in front of the vehicle can be detected and ranged.

Figure 2 confirms this trend of an increase in the minimum ranging distance as the baseline distance between the cameras is increased. The minimum ranging distance is estimated by finding the minimum distance at which the common field-of-view of the two cameras is at least 1.8m (this threshold is motivated by the vehicular ranging application, where the width of the typical car is $\approx$ 1.7m).

2.3 Single Camera based Ranging

Disparity-based ranging requires two camera views, which has generally meant using two, physically-separated cameras. However, we show that *disparity-based ranging can be performed using a single camera*. As long as two views can be obtained and their relative displacement is known, the location of the object of interest itself can be estimated using simple ray-tracing geometry techniques.

We employ this key insight to enable disparity-based ranging using a single camera, e.g., just the back camera of an off-the-shelf, dashboard-mounted smartphone or a webcam mounted above a shelf. We obtain two distinct views of the scene in front using the mirror arrangement depicted in Figure 3. (This figure depicts the vehicular ranging scenario, although a similar configuration also works for the smartshelf scenario, as shown in Figure 5.) As indicated in the figure, the smartphone is placed on the dashboard of the car such that its back camera is facing the road ahead. Although *Mirror2*, as can be seen in the figure, is placed in the field-of-view of the camera and thus occludes part of the view, the camera can still observe the scene that is directly ahead. *Mirror2*, on the other hand, contains a reflection of *Mirror1*, which is displaced laterally relative to the camera. Therefore, the camera can see both the direct view of the object (e.g., vehicle) in front and also a displaced view in *Mirror2*. Thus, using a single camera and a simple two mirror setup, we enable the smartphone to obtain two distinct views of the scene of interest.

We present a few details:

The mirror angles: Ideally, if the mirrors are placed at 45^o, then the virtual second view observed by the camera has an image plane that is parallel to its own. In such a case, the disparity-based ranging formulation from the previous section can be used as it is with minor modifications (to account for the fact that the mirror and the camera are not at the same depth). However, in our setup, where the camera is laterally displaced from the mirror, the overlap between the reflected view and the direct view of the scene directly in front of the camera is almost non-existent.In order to increase the overlap between the two views, we rotate $Mirror2$ beyond 45^o (50^o to be precise) so that it can direct the reflected rays from objects directly in front of the camera (the objects of interest) to the camera lens. While rotating $Mirror2$ beyond 45^o can direct reflected light from the front, the assumption that the two views are approximately parallel does not hold anymore. Thus, the simple similar triangles based formulation of disparity-based ranging does not hold anymore.

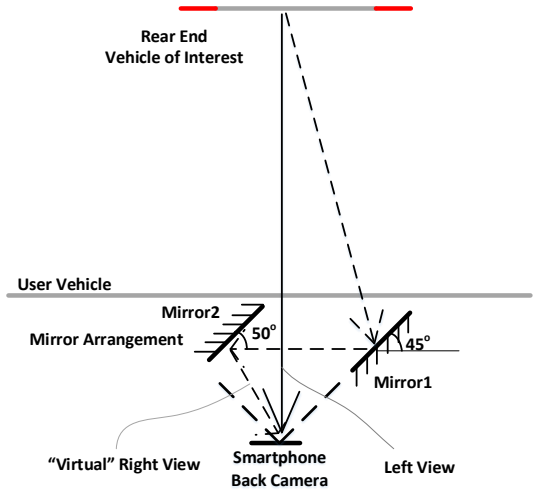

Figure 3: Mirror Arrangement for vehicle ranging

Calculating the depth of the object of interest: The object of interest is seen twice in each camera frame: directly by the camera and also in the reflection of *Mirror1* in *Mirror2*. Say the two pixel locations are x_{camera} and x_{mirror} respectively. Intuitively, the image location, x_{camera} implies that the direct ray of light from the object of interest makes an angle of $[(\frac{180-\theta}{2}) + \frac{x_{max}-x_{camera}}{x_{max}} \times \theta]$, where x_{max} is the horizontal pixel width of the frame and θ is the field-of-view of the camera, with the image plane. Similarly, the x_{mirror} implies that a ray of light from the virtual object makes an angle $[(\frac{180-\theta}{2}) + \frac{x_{max}-x_{mirror}}{x_{max}} \times \theta]$ with the image plane. Once the position and angles of mirrors are known with respect to the camera, the angle of the second ray with respect to the image plane can also be calculated. The depth of the object of interest is simply the intersection of these two rays.

Correspondence between the two image views: Typically, when the image planes are not parallel, image rectification is employed to ensure faster stereo correspondence between the two image views. However, since CamMirror computes disparity by focusing on just the scenario-specific features (see Section 2.4 below), we can simplify the process of establishing correspondence.

2.4 Computing Disparity

As noted in Section 2.3 above, establishing correspondence between the two views of the scene can be an expensive computation. Therefore, in CamMirror, we use an approach that involves picking out scenario-specific features, which are then used to compute disparity much more efficiently. Here we discuss the two scenarios that we have focused on — vehicular ranging and smart shelf — in turn.

2.4.1 Identifying and matching tail lamps

There are various features of vehicles that can be used to establish a correspondence across the camera views and thereby compute disparity, e.g., edges and corners of vehi-

(a) Left View. (b) Color Filter-Left. (c) Right View.

(d) Color Filter-Right. (e) Eliminating Noise In Color Filter output. (f) Match Shape, height, proximity

Figure 4: Identifying and matching corresponding tail lamps across camera views.

cles, tail lamps, license plates, etc. However, we picked tail lamps, since all vehicles, irrespective of the type or size have one. Furthermore, all three and four-wheeled vehicles, have a pair of tail lamps, which gives us two features per vehicle. Averaging disparities computed over multiple features of a same vehicle helps in reducing the effect of errors and noise on the distance estimation. Hence, we pick the tail lamps as the features to identify and match across the camera views.

To detect the tail lamps, we take advantage of the fact that these are typically a shade of red in color. We employ a color filter on both camera views and detect all objects that have a shade of red. We construct a Hue-Saturation-Value (HSV) image corresponding to each camera view. This essentially makes the color independent of the shade and lighting. Hue indicates the color. Different values of S and V essentially give different shades of the same color. This makes the HSV color scheme robust to changes in color due to different lighting conditions. We apply an HSV based *red* color filter on the frame of each camera view.

A colour filter alone could still result in spurious detections, e.g., red-coloured objects in the frame. To eliminate some of this noise, we find contours, enclosing a certain minimum area, in the binary output of the colour filtered frame in each camera view. Each contour is potentially the (approximate) outline of a tail lamp. We associate each identified contour with its centroid. We also leverage the observation that tail lamps across the two camera views (direct and mirrored) should be identified at the same height since the camera and the mirrors are placed at the same height, only displaced horizontally. Figure 4 illustrates the steps in identifying and matching corresponding tail lamps across camera views.

While the HSV based color scheme makes the color based tail lamp detection more robust under different lighting conditions, it is still possible that when the lighting is poor but the vehicle tail lamps are still not lit up (e.g., dawn or dusk), the tail lamps fail to be detected using the empirically set *red* color filter. To overcome this issue, we adaptively switch between two distinct color filters. Under normal conditions, we use the *red* color filter. However, when we identify that the lighting condition is dull for a continuous stretch of time (5 seconds), we switch to a second, empirically set darker color filter. We detect the lighting condition of the scene by periodically computing the histogram of the S and V values of the frame. If majority of the V values (which give a measure of darkness of the color, with a lower V value suggesting a darker shade) are lower than half the maximum value, then we identify the scene to be a dull scene.

2.4.2 *Identifying hand reaching out for item on shelf*

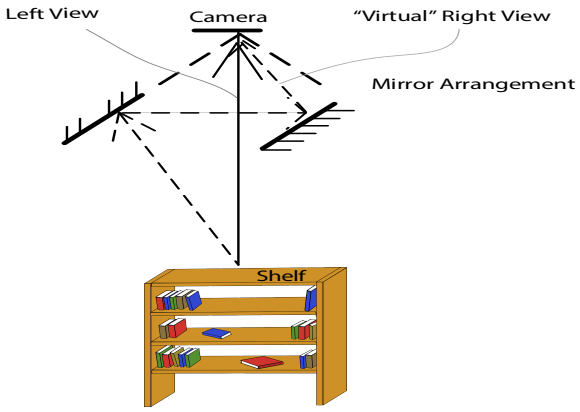

Figure 5: Mirror Arrangement for smart shelf

Consider a retail setting where items are placed in several shelves. In this section, we describe our smart shelf setup and how CamMirror design can be used to identify the shelf user has reached out to pick up an item. Fig. 5 shows the overview of our setup along with the mirror arrangement.

We follow the same design principles outlined earlier for mirror placement and calculating the depth of the object of interest. In order to identify which shelf and where on the shelf user hand is placed, we apply a simple background subtraction (BS) technique to determine the moving object (i.e., user hand) from the static camera feed. Background subtraction generates a foreground mask, i.e., binary image containing pixels that belongs to the moving object. Fig. 6 shows the masked frame where white pixel represents moving object (user hand). The two highlighted views show the

user hand viewed directly by the camera and from the mirror arrangement. Since our camera position and the scene is pretty static, we eliminate much of the background noise.

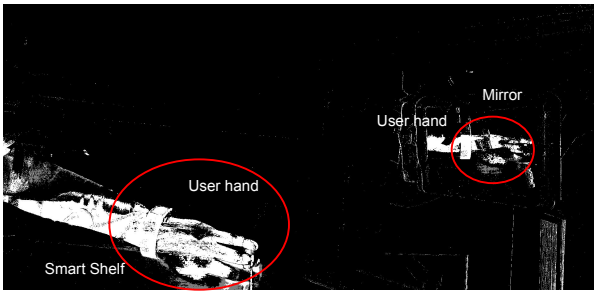

Figure 6: Background subtraction on a frame when user was picking up an item

3. EVALUATION

This section presents a detailed evaluation of various aspects of the CamMirror system. Specifically, we show distance estimates for the two applications considered.

3.1 Vehicle ranging

We collected several videos (854x480 pixels at 15 frames per second) and high resolution images (1920x1080 pixels) in burst mode during drives in two locations: a city in a developing region (India) and another in a developed region (USA). For evaluating the performance under various lighting conditions, we specifically gathered data at different times of the day including night times and evenings with deep shadow effects.

Figure 7: CamMirror setup for vehicle ranging

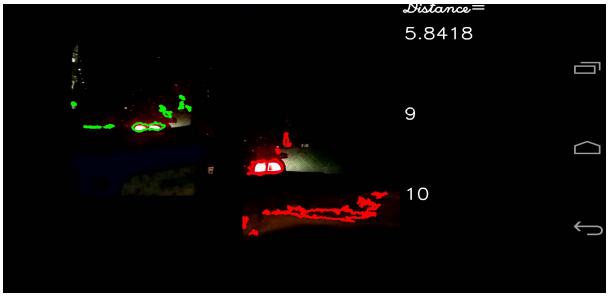

Figure 8: Screenshot of the frame

Setup: Fig. 7 illustrates our implementation mirroring the design in Fig. 3. The mirror sizes are 19cm by 15cm (Width * Height) and distance between mirrors is 25cm. The mirror plane is 18cms from the phone. *Mirror1* is angled at 45^o while *Mirror2* is angled at approximately 50^o. As indicated in the figure, the entire setup rests on a wooden board. For our experiments, the setup is held in position by the copassenger. The phone itself is clamped to position. Fig 8 illustrates a screenshot of the online implementation of CamMirror. As indicated in the figure, because of the mirror setup, the camera observes two distinct views (one real, one virtual). The virtual view is confined to the left half of the frame, while the camera view of the region of interest is confined to the right half. We adopt appropriate mask on the image to focus the field-of-view to the region of interest. As indicated in Fig. 8, due to the loose color threshold, nine contours are observed in the virtual observation, while ten are observed in the real camera view itself. However, using the principles in Section 2.4.1, the tail lamp contours correspondence is obtained and the distance is estimated online.

The mean processing time per frame is approximately 140ms with a variance of 30ms. Processing within each frame includes masking the frame to confine the field-of-view to the object of interest, color filtering, finding closed contours, and matching contours across the two images. Most of the computational complexity is alleviated by masking majority of the input frame to limit the processing to the object of interest.

Figure 9 illustrates the performance of the smartphone based real-time implementation of the disparity-based distance estimation. We seek to evaluate the effect of the additional rotation of the mirror on the disparity-based distance estimate. As indicated in Fig. 9, upto almost 7m, the mean distance estimate follows the ground truth. The variance in the disparity-based distance estimate is less than 50cms at 7m.

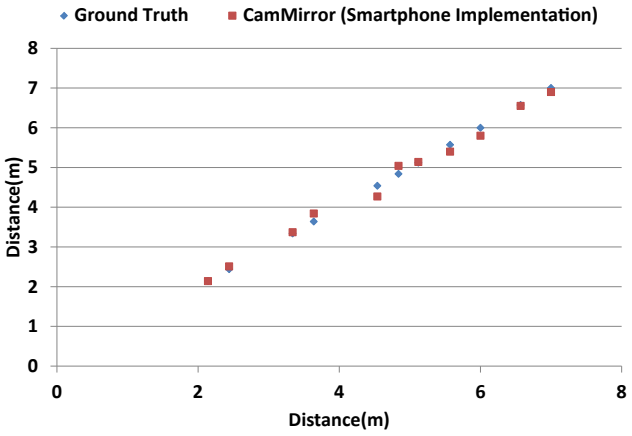

Figure 9: Disparity-based distance estimation using a single camera

3.2 Smart shelf

In smart shelf application, to determine the exact shelf from which the user picked an item, we need two distinct views of the object of interest (in this case the user hand picking an item). We achieve two distinct views of the object

of interest using the mirror arrangement shown in Fig. 10. In our experiments, we mounted a camera on the wall looking down to the shelf as shown in the figure. The markings 1,2,3,4 (in Fig. 10) indicates the items placed in each shelf. Furthermore, another distinct view of these objects can be seen in markings 5,6,7,8. Therefore, camera can see both the direct view of the object and also a displaced view in the mirror.

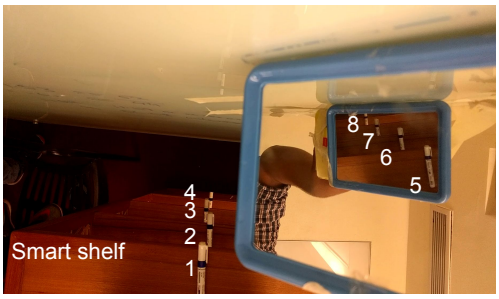

Figure 10: CamMirror setup for smart shelf

In this application, we only need to identify which shelf and where in the shelf user hand is present as opposed to vehicle ranging application, which requires fine-grained distance estimation. In our evaluation, for each frame we first identify the distance between the user hand and camera. Further, we cluster the distance values, where each cluster corresponds to the individual shelf. The appeal of this approach is to make CamMirror estimate less sensitive to setup mismatches. We then compute the true positives (tp): user hand present in a particular shelf both in ground truth and evaluation, true negatives (tn): user hand not present in both ground truth and evaluation, false positives (fp): user hand not present in a particular shelf but identified as present in our evaluation and false negatives (fn): user hand present in a particular shelf but identified as not present in our evaluation. We compute precision, recall and F-measure for identifying which shelf user hand was present. Precision is defined as $P = \frac{tp}{tp+fp}$, recall $R = \frac{tp}{tp+fn}$ and F-measure $F = 2 \cdot \frac{P \cdot R}{P+R}$. From the video collected during our trials, we obtained a 92% precision with 74% recall and f-measure value of 82%.

4. CONCLUSION

In this paper, we have presented CamMirror, which employs a novel single-camera technique to performance disparity-based ranging, using a pair of mirrors in the place of a second camera. Such ranging can be the building block for several physical analytics applications. We have presented two, vehicular ranging and a smart shelf, which show the effectiveness of CamMirror.

5. REFERENCES

[1] Google Self-Driving Car Project. https://www.google.com/selfdrivingcar/.

[2] Black, D., Clemmensen, N. J., and Skov, M. B. Pervasive computing in the supermarket: Designing a context-aware shopping trolley. *International Journal of Mobile Human Computer Interaction (IJMHCI) 2*, 3 (2010), 31–43.

[3] Broggi, A., Bertozzi, M., and Fascioli, A. Self-Calibration of a Stereo Vision System for Automotive Applications. In *IEEE International Conference on Robotics & Automation* (May 2001).

[4] Foggia, P., Limongiello, A., and Vento, M. A real-time stereo-vision system for moving object and obstacle detection in avg and amr applications. In *Seventh International Workshop on Computer Architecture for Machine Perception (CAMP'05)* (July 2005), pp. 58–63.

[5] Huang, Y., Fu, S., and Thompson, C. Stereovision-based object segmentation for automotive applications. *EURASIP Journal on Advances in Signal Processing 2005*, 14 (2005), 910950.

[6] Liciotti, D., Contigiani, M., Frontoni, E., Mancini, A., Zingaretti, P., and Placidi, V. Shopper analytics: A customer activity recognition system using a distributed rgb-d camera network. In *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage* (2014), Springer International Publishing, pp. 146–157.

[7] Mukhtar, A., Xia, L., and Tang, T. B. Vehicle detection techniques for collision avoidance systems: A review. *IEEE Transactions on Intelligent Transportation Systems 16*, 5 (Oct 2015), 2318–2338.

[8] O'Malley, R., Glavin, M., and Jones, E. Vehicle Detection at Night based on Tail-light Detection. In *Int. Symp. Vehicular Computing Systems* (2008).

[9] Pierdicca, R., Liciotti, D., Contigiani, M., Frontoni, E., Mancini, A., and Zingaretti, P. Low cost embedded system for increasing retail environment intelligence. In *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on* (2015), IEEE, pp. 1–6.

[10] Rezaei, M., Terauchi, M., and Klette, R. Robust Vehicle Detection and Distance Estimation Under Challenging Lighting Conditions. *IEEE Transactions on Intelligent Transportation Systems 16*, 5 (Oct 2015).

[11] Widmann, G. R., Daniels, M. K., Hamilton, L., Humm, L., Riley, B., Schiffmann, J. K., Schnelker, D. E., and Wishon, W. H. Comparison of lidar-based and radar-based adaptive cruise control systems. Tech. rep., SAE Technical Paper, 2000.