

Solid or not solid: Vision for radar target validation

Amir Sole	Ofer Mano	Gideon P. Stein	Hiroaki Kumon	Yukimasa Tamatsu	Amnon Shashua
MobileEye Vision Technologies Ltd.			DENSO Corporation.		Hebrew University
Jerusalem, Israel.			Japan.		Jerusalem, Israel
amir.sole,ofer.mano,gideon.stein@mobileye.com			kumon,tamatsu@rd.denso.co.jp		shashua@cs.huji.ac.il

Abstract

In the context of combining Radar and Vision sensors for a fusion application in dense city traffic situations, one of the major challenges is to be able to validate Radar targets. We take a high-level fusion approach assuming that both sensor modalities have the capacity to independently locate and identify targets of interest. In this context, Radar targets can either correspond to a Vision target — in which case the target is validated without further processing — or not. It is the latter case that drives the focus of this paper. A non-matched Radar target can correspond to some solid object which is not part of the objects of interest of the Vision sensor (such as a guard-rail) or can be caused by reflections in which case it is a ghost target which does not match any physical object in the real world.

We describe a number of computational steps for the decision making of non-matched Radar targets. The computations combine both direct motion parallax measurements and indirect motion analysis — which are not sufficient for computing parallax but are nevertheless quite effective — and pattern classification steps for covering situations which motion analysis are weak or ineffective. One of the major advantages of our high-level fusion approach is that it allows the use of simpler (low cost) Radar technology to create a combined high performance system.

1 Introduction

Radar systems have been shown to provide good performance in Adaptive Cruise Control (ACC) applications. They work well in the highway environment where the host vehicle and the valid targets (other vehicles) are all moving at highway speeds and can be separated from stationary targets. Automotive engineers are faced with two challenges: to increase the performance envelop to include urban driving and stop & go situations and to reduce system cost so as to penetrate the mid-range. Unlike highways, urban environments have significant clutter of radar reflecting targets. Many stationary targets such as slow moving

vehicles, parked cars, railings and lamp posts are relevant targets and cannot simply be ignored. They must be accurately located and tracked to determine if there are in the vehicle path.

One way to meet both these challenges is to combine a low cost radar together with a vision system. One approach for radar cost reduction is to exclude the moving parts from the radar, this design will also make the radar smaller. DENSO Corporation followed this approach by developing a Phase Array Radar which uses an electronic scan instead of mechanical scanning. The angular resolution of a phase array radar is lower than that of a mechanical scanning radar and thus has stronger side lobes. These side lobes can cause false detection both multi-path false (reflections from walls) and other reflectors on the road. It is possible to solve these problems for particular applications such as ACC using data processing treatment without changing the hardware. However, to support a more expanded range of applications it is necessary to supplement the radar with additional sensors in a sensor fusion approach.

The vision system can more than compensate for the lower angular resolution of the low cost radar and the increased appearance of ghost radar targets. The camera has very good angular resolution and can be used to determine both height, width and lateral speed of the target. Pattern recognition can be used to classify the object and even non-radar (or weakly) reflective targets such as pedestrians can be detected. Moreover, the cost of a vision system is significantly lower than the cost saved by using the simpler radar. A vision system, in addition to overcoming cost reduction problems, can contribute to the system features such as road analysis and scene understanding. Thus taken together the fused system can be less expensive and have more capabilities than high-end radar systems. A complete fusion system must address the following problems:

1. Match radar (approved and candidate) and Vision targets. The matching process is based on range, angular position, range-rate and must consolidate such information

across multiple images.

2. Using shape and texture characteristics provide a rich target description which includes width, height, length (when possible — such as when target is in an adjacent lane).
3. Validate target as to "Solid" or "Ghost" where the Ghost target arises from radar reflections.
4. Provide target classification for Solid targets: this includes vehicle (and ordinal size) whether moving or stationary, and motorcycles.
5. Detection of non-reflecting or weak reflecting radar targets: these includes mostly Pedestrians whether moving (longitudinally or laterally) or stationary.
6. Target lock when target moves out of the radar viewing field but is still visible to the Vision system (which typically has a much wider field of view).
7. Integration with lane mark analysis: this includes refinement of yaw rate sensing with visual input for better designation of in-path targets and for path prediction.
8. Integrate *visual motion analysis* (optic-flow) for early target cut-in detections.

The list above is by no means complete but provides an idea of the complexity of the task. We take a "high level" fusion approach in which each of the sensory modalities (radar and vision) has the capacity for independently locating and identifying objects of interest.

The radar sensor provides accurate longitudinal and approximate (and somewhat rather poor) lateral positioning of a radar reflecting target. The Vision sensor can provide an accurate lateral positioning with precise object boundaries and an approximate longitudinal position (yet sufficient for distance control — see [8]) — therefore in that sense the two sensing modalities complement each other. But the fusion can go much deeper as the visual processing provides shape and texture cues for pattern classification, motion cues for visual motion analysis and image processing for lane mark analysis. The shape and texture cues can be used by the fused system for target typing, target validation, collection of additional features (like object length) and certain elements of "scene understanding" which includes detection of guard-rails, bridges and over-passes, pedestrian walking zones, and so forth; the visual motion analysis can be used for early detection of adjacent lane targets and for cut-in analysis; and the lane analysis used for in-path designation.

In this paper we focus on one of the items in the list above which is related to target validation with the focus on distinguishing Solid and Ghost radar targets. A ghost target is weakly correlated with physical features in the world (and

thus with image features) so it may seem that the strongest validation cue would be based on motion parallax measurements. The motion parallax can distinguish between a target on the road surface (be it a ghost reflection or a flat metallic object on the road) and a target positioned perpendicular to the road surface. We show in this paper that motion parallax alone is not sufficient for making a reliable validation since often the detected targets are relatively narrow or small in the image. We develop alternative motion analysis cues which are not sufficient for parallax recovery yet are effective for separating ghosts from physical upright objects. In addition to visual motion cues we collect texture and shape cues for classifying typical non-vehicle targets that tend to be picked up by the radar such as guard rails and poles. Furthermore, we focus on vehicle-related targets that are not systematically detected by the the radar sensor such as laterally moving vehicles (in a "T-bone" configuration).

We describe below some basic characteristics of the radar and vision sensors we used for this project, discuss related work and then proceed with the target validation processing description.

1.1 Related work on radar/vision fusion

The advantages of fusing together data from radar and vision together with the complementary advantages and disadvantages of each sensor is widely known and has been discussed in [3, 11, 10, 1]. Grover-et-al [2] perform fusion on low level features such as blobs detected in both radar and vision data. Fusion is performed in polar coordinates and based on angular position. A single radar image and a single night-vision image is used. A more common approach is to perform fusion at the target level. A radar system produces a list of targets with range and rough azimuth. The image is then tested for a vehicle [4] or pedestrian target by applying pattern recognition techniques in a location given by the azimuth value from the radar. Couet-el [1] describe a Bayesian framework for fusion based on target information (position and range) from multiple modalities. In this paper, however, it is assumed that range information is not available since the visual processing is monocular.

2 Radar Target Validation

The visual processing is based on a single forward facing camera mounted inside the host vehicle typically near the rear view mirror. The camera used in this work has VGA sensor (640×480) and a horizontal FOV of 47° . The visual processing is an independent multi-function system developed by MobilEye Ltd. that performs forward facing vehicle detection, vehicle cut-in, lanes following and pedestrian detection [9, 8, 6, 5, 7]. We divide the target validation task into two cases: (i) class-specific object,

and (ii) non-class-specific object. When a match between a radar and a vision target occurs we refer to this situation as a "class specific". According to our high-level approach such a match means that the target has been independently corroborated by the vision sensor as a known object type and thus the radar target is validated. When the radar target has no match with a vision target we refer to this situation as a non-class-specific situation. The latter case requires specialized processing combining visual motion analysis and shape and texture cues for validating the radar target. The measurements we collect for the validation process include:

1. Measure of target stability from the radar sensor. Ghost target tend to be less stable than rear longitudinally moving objects. However, unstable radar targets also arise from physical solid targets such as laterally moving vehicles.
2. Radar target is matched to the closest edge-defined image region. The image region is tracked over a number of frames for the purpose of collecting motion parallax cues — both direct and indirect.
3. The area of the radar target is matched against guard-rail templates and against general repetitive patterns (such as parked cars columns).
4. The radar target is matched against a Pole-like template.
5. Unstable radar targets are matched against laterally moving vehicle templates based on motion and wheel detectors.
6. Radar occlusion analysis is conducted to determine whether the radar target is occluded. The occlusion criterion requires that the target will overlap a closer vehicle, pedestrian or any other bounded target.

A radar target that fails all these tests is determined to be a ghost target. These tests are explained below in more detail.

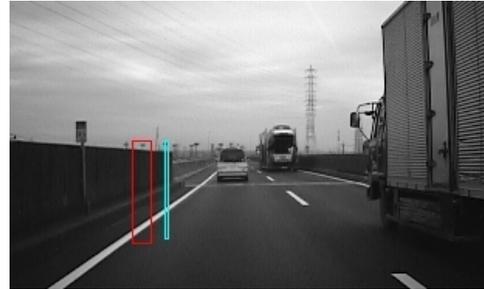
2.1 Class specific object

Certain classes of objects can be detected in the image on the basis of image information alone. For example cars, truck and motorcycles have a characteristic size and appearance so they can be reliably detected from the rear and the side if the whole vehicle is in view.

This step receives as input a list of radar targets and a list of vision detected vehicles. A radar targets that matches a vision target is considered a valid target. Matching is based on range, range rate and angular position. After a match is found, the radar range and vision angular position and extent are used to determine the possibility of collision. For example, the lateral position X can be computed using radar range Z_r , the horizontal image coordinate x and the camera focal length f : $X = \frac{xZ_r}{f}$.



(a)



(b)

Figure 1: *In (a) we see the target in first frame and the rectangle the system tries to track after. In (b) we see the result of the tracking and the position of the target in this frame. The difference in the position of the tracked rectangle and the target indicates the target is a ghost.*

2.2 Image stability

The main characteristic of radar ghosts arising from reflections is that they are not correlated to any feature in the world. Thus they appear to "float" in the image. In this test the image region around the radar target is tracked over multiple frames and the motion is compared to the motion of the radar target, taking into account the lateral accuracy of each sensor. An example is shown in figure 1. As mentioned above, an unstable radar target is not necessarily a ghost as instability may arise from laterally moving objects as well. We therefore test the region for the existence of a vehicle side as described in section 2.3.2.

2.3 Non-class specific object

The test whether an object is a non-class specific is based on motion cues. Motion cues have been used in [12] to determine if groups of horizontal lines are on the road surface or on an upright structure (e.g. rear of a vehicle). However since they do not have accurate range and range rate information from a radar they require more horizontal lines and must use second order effects. Instead, in the method described below any two lines or a single line and the FOE (focus of expansion) can be used to verify a radar/vision match.

Angle and range are used to locate a candidate region in

the image with strong edges. However, since the radar angle and vision range are not very precise these alone are not sufficient for target validation. To validate the target, this region is tracked over multiple frames and the scale change in the image is compared to the range rate of the radar target:

$$\left| \frac{w_1}{w_2} - \frac{Z_2}{Z_1} \right| < T \quad (1)$$

where w_1 and w_2 are the image dimensions of the region (width or height) in two frames and Z_1 and Z_2 are the corresponding radar ranges. T is a threshold. If there is a match the target is validated as a *solid* otherwise it is classified a *ghost*.

The change in image dimensions Δw due to change in relative distance ΔZ depends on the range, range rate and object width and also on the camera parameters:

$$\begin{aligned} \Delta w &= w_2 - w_1 = \frac{fW}{Z + \Delta Z} - \frac{fW}{Z} \\ &= -\frac{fW\Delta Z}{Z(Z + \Delta Z)} \approx -\frac{fW\Delta Z}{Z^2} \end{aligned} \quad (2)$$

Using subpixel tracking methods, the measurable change that can be resolved in real-world images is on the order of 0.25 pixel.

Typically we are interested in situations where $\frac{\Delta Z}{Z} < 0.1$. An object whose width is 1 meter and whose distance is 50 meter will appear in the image as 15 pixels wide and the value of Δw will be considerably larger than the noise. However if the object is very narrow, as in the case of a pole or lamppost, the change in w can be on the order of 0.1 pixel or less. In that case it is better to use the image motion of the pole (Δx) relative to the FOE to compute the scale change.

$$\begin{aligned} \Delta x &= x_2 - x_1 = \frac{fX}{Z + \Delta Z} - \frac{fX}{Z} \\ &= -\frac{fX\Delta Z}{Z(Z + \Delta Z)} \approx -\frac{fX\Delta Z}{Z^2} \end{aligned} \quad (3)$$

Using robust image stabilization techniques the FOE can be determined to within 1 pixel. This puts a constraint on how close the vertical edge can be to the FOE for it's motion to be detected reliably. The practical limit is 20 pixels.

In order to have a reliable scale tracking one must first bound the region of the obstacle in the image. Since the shape of the reflecting object is unknown we developed detection strategies for typical object types: wide objects with strong vertical edges (cars, dumpsters), wide objects with strong horizontal edges (sides of cars, walls), narrow objects with strong vertical edges (poles, lamppost), strong straight edges in the direction of the FOE (guard rails and walls parallel to the road).



(a)



(b)

Figure 2: (a) The radar target is on an edge inside object and a vertical edge is found on each side. In (b) the radar target is to the right and both detected edges are on the left of the radar target.

2.3.1 Vertical edges

The first attempt to bound the target area is by finding vertical edges around the target. Using the following method:

1. We assume the target is 1-3m wide and look for vertical edges with a lateral distance of 0.5 - 1.5m from the target. If two edges are found than these are the bounding edges of the object. See Figure 2a.
2. If only one edge was found we assume that the target is narrow and look for the closest edge within 1m from the first edge. See Figure 2b.

2.3.2 Horizontal edges

If we fail to bound the target with vertical edges we try to find horizontal edges that goes through radar target. The search region is determined by the radar range and the ground-plane constraint.

The typical target detected in this way are sides of vehicles (T-bone). See Figure 3. These are particularly important targets to recognize since not only is the consequence of hitting such a target very bad but also there are often moving targets that might be moving into or out of a collision course. Accurate target boundary location and target lateral speed estimation are critical. The radar alone lacks the angular accuracy and resolution to correctly compute these



Figure 3: *Horizontal edges detected on the radar target. A typical example of a T-bone vehicle.*

values but provides the vision system with accurate range which is used to convert angles to lateral position.

Since a target with strong horizontal lines is often a side of a vehicle and since these targets are of special importance, a detector based on pattern recognition is used to verify a T-bone target and to determine the extent of the vehicle side. This detector first detects wheel candidates by searching for circles in the image that correspond to wheels of 0.5m to 1.5m in diameter at the range given by the radar. Restricting the search in this way speeds up the computation and also reduces the likelihood of false detection.

After the wheels are detected the wheel base length is used to estimate the vehicle length. The vehicle is tracked in the image and the likelihood of collision can be estimated.

2.3.3 Pole like structures

The nature of the vertical bounded region we search for in section 2.3.1 is at least 1m wide. Poles are narrow and do not fit this model. In addition, the motion criteria relative to the FOE which we use on narrow structures is a weak criteria since it assumes a rigid world. Therefore a special search for poles has been implemented that puts some restrictions on what edges can be classified as 'pole like'. This is done by looking for two close and long vertical edges. These vertical edges must extend across the horizon line. They must be straight although not necessarily perfectly vertical. See Figure 4. The lowest point on the edge can be used to estimate the range to the pole candidate using the ground plane assumption: $Z_{est} = \frac{fh}{y}$ where Z_{est} is the radar range, f is the focal length, h is the camera height and y is the image position of the lowest edge point. Both lateral position and range estimate are used for matching. If the estimated range matches the radar range then we allow larger error of lateral position. After a candidate pole has been found it is tracked and the motion relative to the FOE is compared to the radar range rate.

2.3.4 Forward edges

Radar targets on the sides which we are not matched to horizontal or vertical regions are possibly due to structures

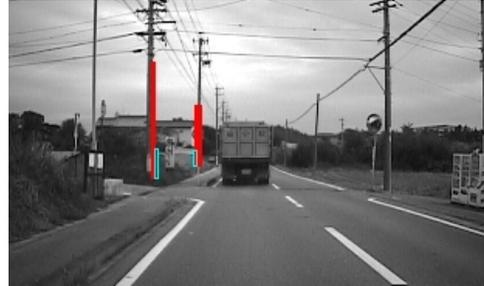


Figure 4: *Radar targets on poles.*



Figure 5: *Radar targets detected on the guard rail. In the image these targets appear as lines in the direction of the road vanishing point.*

parallel to the road edge such as guard rails. In the image these structures appear as lines passing through the vanishing point as determined by the vision based *lane detection system*. See Figure 5.

When dealing with fences and guard rails one must pay special attention to repetitive structures which are parallel to the road. These can create aliasing effects due to the finite sampling rate of the video camera and the radar. The result of these aliasing effects can be incorrect motion tracking. Sometime vertical lines on a fence or rail appear to be stationary or even moving inwards towards the FOE.

In order to detect the presence of repetitive structure we use the invariance of cross ratios under perspective projection. If a set of edges are due to a set of equally spaced structures (such as poles) in a line parallel to the road then any two such edges together with the location of the vanishing point can be used to predict the location of the other edges in the image. By comparing the predicted edge location and the actual edge location one can compute the likelihood of this to be a random occurrence or due to repetitive structure. Figure 6 shows an example of a fence with strong repetitive structure.

3 Results

The combined system was tested on data recorded from three hours driving in both highway and urban scenes. The



Figure 6: Structures along highways often have a repetitive structure which can cause problems to tracking algorithms due to aliasing. The cyan lines show repetitive structures that were detected using the projective invariant of the cross ratio.

radar target data was recorded together with the video data and vehicle speed. The recorded data was then analyzed offline. Radar targets that were reported as *mature* targets and that lasted for at least five radar cycles (0.5 seconds) were manually tagged as either solid, ghost or occluded. The total number of mature radar targets detected was 1165.

The system was then run offline on the recorded data with the system classifying the targets to either solid, ghost or undetermined. Table 3 compares the manual tag to the system classification.

System Manual Tag	Solid	Ghost	Occluded
Solid	1096	10	0
Ghost	20	13	0
Undetermined	16	0	10

4 Summary

We have presented a combined system that uses vision for radar target validation. The combined system can achieve high reliability target detection with low false rate in demanding situation such as cluttered urban environments.

The validation strategy is based on a high-level sensor modality approach which assumes that each sensor has a capacity to form independent target acquisition. Therefore matched targets are by definition validated and do not require further processing. We have broken down the validation process for unmatched targets into a number of steps combining motion analysis with shape and texture analysis. Much of the challenge in this process lies in the fact that typically the areas of interest are small or narrow and may not be sufficiently described by nearby distinct image

features. As a result, we have developed specialized detectors both for the motion analysis and for the shape analysis to overcome these challenges.

References

- [1] C. Coue, Th. Fraichard, P. Bessiere, E. Mazer Multi Sensor Data Fusion Using Bayesian Programming: an Automotive Application In *International Conference on Intelligent Robots and Systems*. 2002, EPFL, Switzerland.
- [2] R. Grover, G. Brooker, H. F. Durrant-Whyte A low level Fusion of Millimeter Wave radar and Night-vision Imaging for Enhanced Characterization of a Cluttered Environment In *Proc. 2001 Australian Conference on Robotics and Automation Sydney*, 14-15 Nov 2001
- [3] U. Hofmann, A. Rieder, E. D. Dickmanns Radar and Vision Data Fusion for Hybrid Adaptive Cruise Control on Highways. In *ICVS* pages 125-138, 2001
- [4] J. Laneurit, C. Blanc, R. Chapuis, L. Trassoudaine Multi-sensorial Data Fusion for Global Vehicle and Obstacles Absolute Positioning In *IEEE Intelligent Vehicles Symposium (IV2003)*, June 2003, Columbus, OH.
- [5] O. Mano, G. Stein, E. Dagan and A. Shashua. Forward Collision Warning with a Single Camera In *IEEE Intelligent Vehicles Symposium (IV2004)*, June. 2004, Parma Italy.
- [6] P. Reisman, O. Mano, S. Avidan and A. Shashua. Crowd Detection in Video Sequences In *IEEE Intelligent Vehicles Symposium (IV2004)*, June. 2004, Parma Italy.
- [7] A. Shashua, Y. Gdalyahu, G. Hayun and L. Mann. Pedestrian Detection for Driving Assistance Systems. In *IEEE Intelligent Vehicles Symposium (IV2004)*, June. 2004, Parma Italy.
- [8] G. P. Stein, O. Mano and A. Shashua. Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy In *IEEE Intelligent Vehicles Symposium (IV2003)*, June 2003, Columbus, OH.
- [9] G. P. Stein, O. Mano and A. Shashua. A Robust Method for Computing Vehicle Ego-motion In *IEEE Intelligent Vehicles Symposium (IV2000)*, Oct. 2000, Dearborn, MI.
- [10] K. Weiss, D. Stueker A. Kirchner Target Modeling and Dynamic Classification for Adaptive Sensor Data Fusion In *IEEE Intelligent Vehicles Symposium (IV2003)*, page 132 - 137, June 2003, Columbus, OH.
- [11] B. Steux Fade: A Vehicle Detection and Tracking System Featuring Monocular Color Vision and Radar Data Fusion In *IEEE Intelligent Vehicles Symposium (IV2002)*, 2002
- [12] R. Okada, Y. Taniguchi, K. Furokawa and K. Onoguchi Obstacle Detection Using Projective Invariants and vanishing Lines In *ICCV 2003*
- [13] G. R. Widman, W. A. Bauson and S. W. Alland Development of collision avoidance systems at Delphi Automotive Systems, In *Proceedings of the Int. Conf. Intelligent Vehicles*, pages 353 -358, 1998