

FREE SPACE COMPUTATION FROM STOCHASTIC OCCUPANCY GRIDS BASED ON ICONIC KALMAN FILTERED DISPARITY MAPS

Carsten Højlund, Thomas B. Moeslund, Claus B. Madsen
Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark
{ch, tbm, cbm}@imi.aau.dk

Mohan M. Trivedi
Computer Vision and Robotics Research Laboratory, University of California, San Diego, U.S.A.
mtrivedi@ucsd.edu

Keywords: Free space, Stereo vision, Kalman filtering, Stochastic occupancy grid.

Abstract: This paper presents a method for determining the free space in a scene as viewed by a vehicle-mounted camera. Using disparity maps from a stereo camera and known camera motion, the disparity maps are first filtered by an iconic Kalman filter, operating on each pixel individually, thereby reducing variance and increasing the density of the filtered disparity map. Then, a stochastic occupancy grid is calculated from the filtered disparity map, providing a top-down view of the scene where the uncertainty of disparity measurements are taken into account. These occupancy grids are segmented to indicate a maximum depth free of obstacles, enabling the marking of free space in the accompanying intensity image. The test shows successful marking of free space in the evaluated scenarios in addition to significant improvement in disparity map quality.

1 INTRODUCTION

Autonomous robots in homes and factories are becoming ever more prevalent. These robots perform automatic analysis of their surroundings to determine their position and where they can move, and ideally to determine what their surroundings consist of and who might be present. This makes it natural to include cameras for functions such as object recognition, face recognition, and so forth.

To reduce the number of components on board, it would be beneficial to use the image data for navigation as well. One important aspect of navigation is free space. This needs to be established on-the-fly for the robot to be able to handle dynamic environments, such as doors opening or outdoor environments. Determining the depth of a scene is difficult based on 2D images, but easier with 3D data, leading to a choice of stereo images. It is important to have accurate depth information to reliably determine the free space.

The focus of this paper is to improve the 3D data used for determining free space, obtainable from e.g. monocular camera (Li et al., 2008), stereo camera (Murray and Little, 2000), and omni-directional cameras (Trivedi et al., 2007).

1.1 This Work

In this work we will use stereo cameras to simultaneously provide images and depth. We will not assume forward pointing cameras, as seen in e.g. (Badino et al., 2007), but rather allow for arbitrary camera poses. The stereo images and initial depth estimation are provided by a TYZX stereo camera with a 22 cm baseline (TYZX, 2009). The depth estimation is refined by applying ego-motion to a previous frame thereby transforming it to correspond to the new viewpoint, and subsequently merging the transformed frame and the current frame, creating a more accurate estimation. The refined depth estimation is used to determine the free space around the camera.

2 DEPTH ESTIMATION

Disparity is the measure of difference in position of a feature in one image compared to another image. This can be used to infer the depth of the feature. Several methods of estimating disparity exist and the crucial point in this work is to have an accurate measurement to reliably determine free space. Knowledge of the noise in the estimation process can significantly

improve the system. In our case the distribution of noise has been found to be Gaussian in image space and hence the Kalman filter (Kalman, 1960) can be applied.

2.1 Iconic Kalman Filter

The Kalman filter is a recursive estimation method that efficiently estimates and predicts the state of a discretized, dynamic system from a set of potentially noisy measurements such that the mean square error is minimized, giving rise to a statistically optimal estimation. Only the previous measurement and the current measurement are needed.

The general Kalman filter equations are not directly applicable to a stream of images without some additional considerations. If an $N \times M$ disparity map was treated as a single input to one Kalman filter the measurement covariance matrix \mathbf{R} would consist of $(N \times M)^2$ values, which quickly becomes a prohibitively large number. To make an implementation feasible, each pixel is assumed independent and consequently assigned a separate Kalman filter. A pixel-wise Kalman filter is said to be *iconic* (Matthies et al., 1989).

The filter is applied to disparities, meaning it is the disparity value that is estimated. The pixel coordinates are still used when applying the motion model, but not Kalman filtered. The state vector is therefore just a scalar, i.e. $\mathbf{x}_t = x_t = d_t$.

2.1.1 State, Measurement and Motion Models

The measurement is the disparity value supplied by the stereo camera. When the camera moves, the previous disparity map will in general not represent the new camera viewpoint. Since depth is available for each pixel they can be moved individually according to how the camera was moved, effectively predicting the scene based on the previous disparity map and the ego-motion information of the camera yielding two disparity maps depicting the same scene from the same view point; a prediction and a measurement.

The ego-motion in this context is linear in world space but since image formation is non-linear, the motion becomes non-linear in image space. The standard Kalman filter assumes a linear system which means it cannot handle the application of ego-motion when working in image space. Therefore, the state prediction is handled outside of the Kalman filter by triangulating the input into world space, applying ego-motion as $\mathbf{X}_t = \mathbf{M}_R \mathbf{X}_{t-1} + \mathbf{M}_T$ (where \mathbf{M}_T is translation and \mathbf{M}_R is rotation), and then back-projecting the result into image space. The Kalman filter therefore only handles the state variance.

Support for arbitrary camera poses is accomplished by including a transformation of the ego-motion information to the coordinate system of the camera, given as ω_T and ω_R for translation and rotation, respectively.

The complete motion model will be referred to as the function $\Phi(\cdot)$:

$$x_{t|t-1} = \Phi(x_{t-1|t-1}, \mathbf{M}_T, \mathbf{M}_R, \omega_T, \omega_R) \quad (1)$$

2.1.2 Noise Models

The process and measurement variance are both scalars and assumed to be constant, i.e. $\mathbf{Q} = q_d$ and $\mathbf{R} = r_d$. The process variance q_d reflects the confidence in the predictions and is set to a small number to filter out moving objects (e.g. 0.001). The measurement variance r_d represents the noise in the disparity values and is determined by measuring the precision of the stereo camera. The Kalman gain K is effectively a weighting between the prediction and the measurement based on the relative values of q_d and r_d .

The state variance \mathbf{P} is also a scalar, hence $\mathbf{P} = p$. It is initially set to r_d , and from there on maintained by the Kalman filter.

2.1.3 Predict Phase and Update Phase

The finished iconic Kalman filter equations become:

Predict

$$x_{t|t-1} = \Phi(x_{t-1|t-1}, \mathbf{M}_T, \mathbf{M}_R, \omega_T, \omega_R) \quad (2)$$

$$P_{t|t-1} = P_{t-1|t-1} + q_d \quad (3)$$

Update

$$K_t = P_{t|t-1} (P_{t|t-1} + r_d)^{-1} \quad (4)$$

$$x_{t|t} = x_{t|t-1} + K_t (y_t - x_{t|t-1}) \quad (5)$$

$$P_{t|t} = (1 - K_t) P_{t|t-1} \quad (6)$$

where x is the estimated state, $\Phi(\cdot)$ is the motion model, $\mathbf{M}_T/\mathbf{M}_R$ is the ego-motion as translation/rotation matrices, ω_T/ω_R is the translation/rotation of ego-motion to camera coordinate system, \mathbf{P} is the state variance, q_d is the process variance, y is the measurement variable, K is the Kalman gain, and r_d is the measurement variance.

2.1.4 Update Phase Merging

After the prediction phase has completed and a new disparity map has been calculated, the update phase begins. For each pixel, one of three scenarios occur:

1) No measurement exists for the position of the predicted pixel. The disparity present is simply kept as-is. A disparity is only carried over a certain number of consecutive frames before being discarded to prevent stale pixels from degrading the result.

2) No predictions exists for the position of the measured pixel, and the disparity is kept as-is.

3) Both exist. To filter out noise, a measurement is checked against the prediction it is set to be merged with by applying the Mahalanobis 3-sigma test (Mahalanobis, 1936). If a measurement is within three standard deviations (three times the square root of r_d) of the mean (the prediction) in a given Gaussian distribution, it is merged with the prediction by the Kalman filter. If it fails, it replaces the prediction only if the prediction is young (to prevent spurious pixels from keeping correct measurements out) or deemed stale (to filter out pixels that have drifted).

3 FREE SPACE ESTIMATION

3.1 Stochastic Occupancy Grid

With a disparity map (and consequently a reconstruction of the world points), the free space where a vehicle can navigate without collision can be computed by way of an occupancy grid (Elfes, 1989). An occupancy grid can be thought of as a top-down view of the scene, with lateral position along the horizontal axis and depth along the vertical axis. In a stochastic occupancy grid (Badino et al., 2007) the intensity of each cell denotes the likelihood that a world point is at the lateral position and depth represented by the cell. A world point can therefore occupy more than one cell. How large an area the world point affects depends on the variance (noise) associated with the point. There exists various occupancy grid representations. In this paper, the polar occupancy grid is chosen.

3.2 Free Space

Free space is determined with a combination of the occupancy grid and the underlying disparity map.

3.2.1 Occupancy Grid Segmentation

Each cell in an occupancy grid contains a value representing the probability of that cell being occupied. The grid is segmented by traversing it from near to far, stopping when the value of a cell exceeds a threshold. The depth of that cell is the maximum depth likely to be free of obstacles.

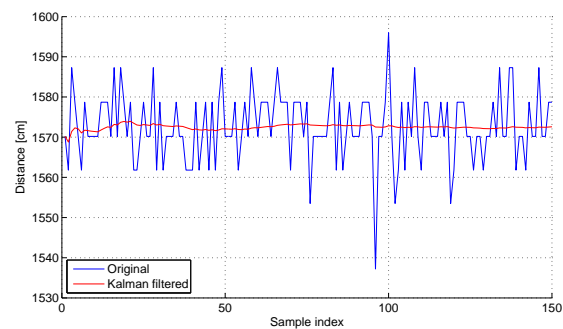


Figure 1: Kalman filtered vs. non-filtered depth values for a single pixel in a static scene, $q_d = 0.001$.

3.2.2 Free Space Marking

The occupancy grid is a direct top-down view of the intensity image. If no occupied cell is found in a particular column, i.e. the direction of a ray from the camera's focal point, all points in the real scene above or below that ray are part of the free space up to the vertical resolution of the occupancy grid.

4 EVALUATION

To evaluate the methods, a number of data sets have been collected. Presented here is mainly one captured on an electric vehicle with the camera at an angle of 65° relative to the direction of travel, which is forward.

4.1 State and Variance

To determine the effect on the state (the depth) of a pixel in a stationary scene with a non-moving camera, the first 150 frames of a single pixel is compared to its Kalman filtered counterpart. Figure 1 shows this for a depth of approximately 16 m. On average, the variance was reduced from 52.14 to 4.10.

4.2 Density

The prediction property of the Kalman filter leads to denser disparity maps, due to the accumulation of measurements over time. Pixels which are predicted for a given number of frames without being updated with a new measurement are discarded to ensure stale pixels do not degrade the occupancy grid. Note how areas occluded in the raw disparity map (Figure 2a) have a predicted depth in the Kalman filtered map (Figure 2d), leading to a denser disparity map than without ego-motion.

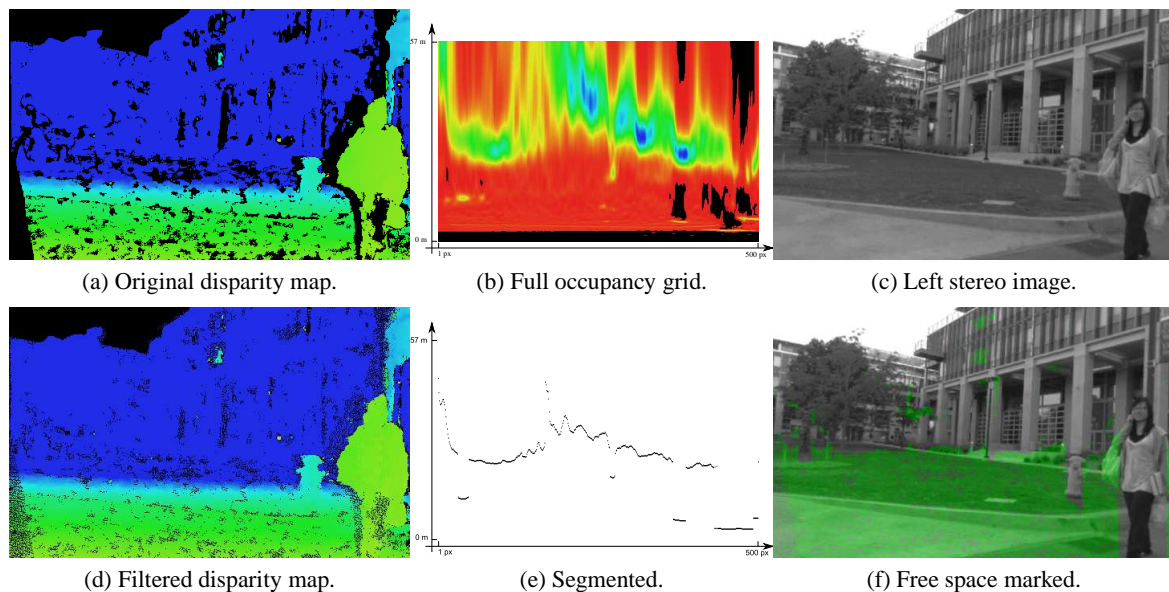


Figure 2: The steps undergone from stereo images to free space. They take place in the following order: (c) + right stereo image \rightarrow (a) \rightarrow (d) \rightarrow (b) \rightarrow (e) \rightarrow (f). In the last step, the depth of each pixel in (c) is looked up in (d), and if it is below the line in (e) it is regarded as free space and marked green in (f).

4.3 Occupancy Grid and Free Space

Figure 2b is an example of an occupancy grid and Figure 2e of the resulting segmentation. The range of depth represented by the occupancy grid is from 0 m to 57 m. Also shown is the scene with (Figure 2f) and without (Figure 2c) the free space marked.

5 CONCLUSIONS

We have shown how to estimate the free space in a scene with a stereo camera. Iconic Kalman filtering of the disparity maps increases density and reduces per-pixel variance (noise), leading to more accurate stochastic occupancy grids and ultimately yielding a more confident estimation of the occupancy of the scene. For a more detailed review see (Høilund, 2009).

REFERENCES

Badino, H., Franke, U., and Mester, R. (2007). Free space computation using stochastic occupancy grids and dynamic programming. *Workshop on Dynamical Vision, ICCV, Rio de Janeiro, Brazil*.

Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57.

Høilund, C. (2009). Free space computation from stochastic occupancy grids based on iconic Kalman filtered

disparity maps. Master’s thesis, Aalborg University, Laboratory of Computer Vision and Media Technology, Denmark.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D).

Li, M., Hong, B., Cai, Z., Piao, S., and Huang, Q. (2008). Novel indoor mobile robot navigation using monocular vision. *Engineering Applications of Artificial Intelligence*, 21(3).

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, volume 2.

Matthies, L., Kanade, T., and Szeliski, R. (1989). Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238.

Murray, D. and Little, J. (2000). Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2).

Trivedi, M. M., Gandhi, T., and McCall, J. (2007). Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety. *Intelligent Transportation Systems, IEEE Transactions on*, 8(1).

TYZX (2009). TYZX DeepSea Stereo Camera. [Online; accessed 20-November-09]. <http://www.tyzx.com/>.