On the formulation, performance and design choices of Cost-Curve Occupancy Grids for stereo-vision based 3D reconstruction

Martim Brandão, Ricardo Ferreira, Kenji Hashimoto, José Santos-Victor and Atsuo Takanishi

Abstract— **We present a grid-based 3D reconstruction method which integrates all costs given by stereo vision into what we call a Cost-Curve Occupancy Grid (CCOG). Occupancy probabilities of grid cells are estimated in a Bayesian formulation, from the likelihood of stereo cost measurements taken at all distance hypotheses. This is accomplished with only a small set of probabilistic assumptions which we discuss in the paper. We quantitatively characterize the method's performance under different conditions of both image noise and number of used stereo pairs, compared also to traditional algorithms. We complement the study by giving insights on design choices of CCOGs such as likelihood model, window size of the cost function and use of a hole filling method. Experiments were made on a real-world outdoors dataset with ground-truth data.**

I. INTRODUCTION

Occupancy grids [1] are a major tool for robot mapping and planning (e.g. [2], [3], [4]). This framework models the environment as a grid of cells which can be either occupied or free, with an occupancy probability computed from measurements of object distances integrated over time. In stereo vision, distances to objects are estimated at each image pixel by computing a cost for each distance hypothesis. This cost-curve provides a confidence on each distance hypothesis which could be further used to accumulate occupancy evidence over time. However, it is common practice in both occupancy grids [2], [5], [6], [7], [8] and stereo fusion algorithms [9] to integrate only the measurements taken at least-cost distance. In some works, the value of the minimum of the cost-curve is used to discard low confidence matches [2], [9], [8], but other entries of the cost-curve are not used. A notable exception is for example [10], where a low number of best hypotheses are kept, but still the whole probability distribution over distance that stereo provides is not fully integrated into a probabilistic framework.

In [11] we proposed a new kind of occupancy grid algorithm, to which we here call Cost-Curve Occupancy Grids (CCOGs). Such algorithms integrate all distance hypotheses of stereo over time from a direct measurement model of the those costs. The method then relied on a few strict assumptions which we alleviate in this paper. Also, performance of occupancy grids both in CCOGs and in methods considering only least-cost distance has not yet been fully characterized in terms of robustness to image noise, number of stereo pairs used for the reconstruction, or design choices such as stereo model and hole filling. In this paper we approach all these issues with experiments on real-world outdoor datasets.

This paper has the following contributions:

1) A Bayesian occupancy grid formulation for stereo-vision using its whole cost-curve and direct models of stereo cost measurement. We call it Cost-Curve Occupancy Grid (CCOG). With respect to [11], we relax a number of probabilistic assumptions and describe how we integrate these CCOGs over time using Bayes filters, as well as how reconstruction holes are dealt with.

2) We thoroughly characterize performance of CCOGs on an outdoor dataset both in terms of the number of stereo pairs used for reconstruction and robustness to image noise.

3) We provide insights regarding the design choices associated with CCOGs such as stereo model choice, cost function window size and hole filling method. Their influence to reconstruction performance is quantified.

II. RELATED WORK

In applications of occupancy grids to stereo vision, several algorithms have been proposed where measurements of distance are taken from the least-cost distance of each pixel [2], [5], [6], [7], [8]. For example, Andert [6] proposed an algorithm where occupancy is computed from least-cost distance assuming a model of measurement uncertainty which increases with distance itself. Similar models of stereo have long been introduced, some of them accounting also for distance to image edges [3]. Some works have dealt with the possibility that the minimum cost does not correspond to true disparity by filtering out pixels where a smoothness constraint was not verified [2], or by assuming a certain prior probability of such errors occurring [7]. An elaborate model recently proposed by Pfeiffer et al. [8] used a likelihood function for correspondence errors of the minimum cost by training on human-labeled datasets.

On the multi-view stereo literature it is more common for multiple distance hypotheses to be considered, however at the cost of strong computational requirements and GPU implementations. In [12], for example, each distance hypothesis accumulates an L1 norm of all pixel intensity errors

*This study was conducted as part of the Research Institute for Science and Engineering, Waseda University, and Humanoid Robotics Institute, Waseda University. It was also supported in part by JSPS KAKENHI (Grant Number: 24360099 and 25220005), and Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation, JSPS, Japan.

M. Brandão is with the Graduate School of Advanced Science and Engineering, Waseda University, 41-304, 17 Kikui-cho, Shinjuku-ku, Tokyo 162-0044, JAPAN. contact@takanishi.mech.waseda.ac.jp

K. Hashimoto is with the Faculty of Science and Engineering, Waseda University.

A. Takanishi is with the Department of Modern Mechanical Engineering, Waseda University; and the director of the Humanoid Robotics Institute (HRI), Waseda University.

R. Ferreira and J. Santos-Victor are with the Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Portugal.

falling in that 3D point from several images. An energy minimization algorithm is then run such that both the total error and a regularization term are minimized. Merrell et al. [9] fuse several stereo pairs into one 3D reconstruction based on geometric constraints and a Gaussian model of cost likelihood. This likelihood is nevertheless used only as a confidence for the least-cost distances. The work was then extended by Hu et al. [10] in order to use the 3 highest likelihood distance hypotheses. Distance is there estimated as the likelihood-weighted average of the distances obtained from several stereo pairs. When compared to such multi-view stereo methods, CCOGs are both computationally inexpensive, probabilistically defined and integrate the whole probability distribution of distance given by the cost-curve.

III. COST-CURVE OCCUPANCY GRIDS

A. Definition

Consider a sensor ray r as the line defined by the origin of the sensor and a point in the world to which distance is measured. The point which is measured is called the target of that ray. In laser sensors a ray exists for each direction at which the laser measures a distance, while in stereo vision a ray exists for each image pixel. We define Cost-Curve Occupancy Grids (CCOGs) as 1-dimensional grids aligned with a sensor ray. The grid is divided into N cells and each cell i can in be in one of two states: occupied O_i or free $\overline{O}_i$. The specificity of CCOGs is that they are designed for sensors which provide a quantitative cost $E_i^{(r)}$ of assigning the target of ray r to each cell i. Stereo vision is such a sensor since it matches a pixel in one image to another in a second image, measuring a cost for each match hypothesis. The curve $E^{(r)} = E_1^{(r)}, ..., E_N^{(r)}$ is called the cost-curve of stereo matching and the number $d = N - i$ is called disparity.

The objective of CCOGs is to compute the probability of occupation of each cell i given the cost-curve $E^{(r)}$,

$$P(O_i|E^{(r)}) = P(O_i\overline{V}_i|E^{(r)}) + P(O_iV_i|E^{(r)}), \tag{1}$$

where V_i is short for the event $\overline{O}_1 \cap \overline{O}_2 ... \cap \overline{O}_{i-1}$ and represents visibility of cell i. Using Bayes' rule, the first term in equation (1) can be rewritten as $P(O_i\overline{V}_i|E^{(r)}) = P(O_i|\overline{V}_iE^{(r)})P(\overline{V}_i|E^{(r)})$, whereas the second term can be rewritten as $P(O_iV_i|E^{(r)}) = P(O_i|V_iE^{(r)})P(V_i|E^{(r)})$.

As we showed in [11], $P(V_i|E^{(r)})$ can be computed by recursively applying the definition of conditional probability,

$$\begin{aligned} P(V_i|E^{(r)}) &= P(V_{i-1}\overline{O}_{i-1}|E^{(r)}) \\ &= P(\overline{O}_{i-1}|V_{i-1}E^{(r)})P(V_{i-1}|E^{(r)}) \\ &= ... = \prod_{j=1...i-1} P(\overline{O}_j|V_jE^{(r)}). \end{aligned} \tag{2}$$

On the other hand, $P(O_i|V_iE^{(r)})$ is given by

$$P(O_i|V_iE^{(r)}) = \frac{p(E^{(r)}|O_iV_i)P(O_iV_i)}{P(V_i|E^{(r)})p(E^{(r)})}, \tag{3}$$

where $P(O_iV_i)$ is a prior on world geometry. The denominator of (3) can also be computed recursively as

$$\begin{aligned} &P(V_i|E^{(r)})p(E^{(r)}) = \\ &= P(O_iV_i|E^{(r)})p(E^{(r)}) + P(\overline{O}_iV_i|E^{(r)})p(E^{(r)}) \\ &= p(E^{(r)}|O_iV_i)P(O_iV_i) + P(V_{i+1}|E^{(r)})p(E^{(r)}) \\ &= ... = \sum_{j=i...N} p(E^{(r)}|O_jV_j)P(O_jV_j), \end{aligned} \tag{4}$$

where we assume that $P(V_{N+1}|E^{(r)}) = 0$, as we will explain next. From this equation it is now possible to estimate $P(O_i|V_iE^{(r)})$ without assuming any strict probabilistic assumptions that we originally proposed in [11] (i.e. independence of V_i and $E^{(r)}$ and $\sum P(O_i|V_iE^{(r)}) = 1$). Before we continue with the formulation of CCOGs, we now clarify the whole set of assumptions we make in this work:

- A target exists for any ray r, or in other words, there exists at least one occupied cell along r. Thus $P(V_{N+1}) = 0$ and $P(V_{N+1}|E^{(r)}) = 0$;
- The target is equally probable to be at any of the cells along a ray r. Thus $P(O_iV_i) = 1/N \ \forall_i$;
- The cost-curve can give no information about occupancy on invisible cells $\overline{V}_i$. Thus $P(O_i|\overline{V}_iE^{(r)}) = P(O_i|\overline{V}_i)$, which corresponds to a prior on world geometry. In our work we model this prior as a constant 0.5 for all i, so that occupied and free cells are equally probable. Thus $P(O_i|\overline{V_i}) = 0.5 \ \forall_i$;
- Costs along a cost-curve are independent from each other. $p(E^{(r)}) = p(E_1^{(r)}...E_N^{(r)}) = \prod_{j=1}^{N} p(E_j^{(r)})$
- Occupancy or visibility on a cell i gives no information on a cost $E_k^{(r)}$ for $k \neq i$. Thus $p(E_k^{(r)}|O_iV_i) = p(E_k^{(r)}) \ \forall_{k \neq i}$;

From (3), (4) and the second assumption follows that

$$P(O_i|V_iE^{(r)}) = \frac{p(E^{(r)}|O_iV_i)}{\sum_{j=i}^{N} p(E^{(r)}|O_jV_j)}, \tag{5}$$

and finally, according to the last 2 hypotheses,

$$p(E^{(r)}|O_iV_i) = p(E_i^{(r)}|O_iV_i)\prod_{k \neq i} p(E_k^{(r)}). \tag{6}$$

The function $p(E_k^{(r)})$ can either be measured directly from sensor measurements or assumed to be uniform by design. That case further simplifies equation (5) to:

$$P(O_i|V_iE^{(r)}) = \frac{p(E_i^{(r)}|O_iV_i)}{\sum_{j=i}^{N} p(E_j^{(r)}|O_jV_j)}. \tag{7}$$

B. Traditional occupancy grids as a special case of CCOGs

In traditional occupancy grids, a single metric distance to a target is directly or indirectly measured [1]. Since no other information is available, the real distance to the target is modeled as a normal distribution around the measured distance. Uncertainty on the measurement is modeled using the distribution's variance. Such range sensors can thus be seen as a special case of cost-curve sensors, but where a single cost is measured. If a target is measured to be at cell k, then in our formulation $E_k^{(r)}$ is minimum and $E_i^{(r)}\forall_{i\neq k}$ are equal and maximum. In traditional range-measurement occupancy grids we then have $p(E^{(r)}|O_iV_i) \propto exp\left(-\frac{(i-k)^2}{2\sigma^2_{range}}\right)$, to which all equations we just defined apply. Such models in computer vision are referred to as "winner-take-all" (WTA) models, where the distance with minimum cost is selected and the rest of the cost-curve discarded. We include this model in our evaluation as well and defined it in Section IV.

C. Grid projection and hole filling

We define R_t as the set of all sensor rays r measured at time t, thus providing a set of 1-dimensional CCOGs $P(O_i|E^{(r)})$. These occupancy probabilities are projected onto a global grid G fixed on the environment according to a transformation between sensor and environment coordinate frames at that instant of time: T_t. Our objective is then to compute $P(G_{XYZ}|R_tT_t)$, where G_{XYZ} is the event that the cell at 3D coordinates (X, Y, Z) is occupied. For a sparse sensor such as stereo vision, the resolution of CCOGs will most likely be different from that of the G since sampling along a sensor ray is not uniform in metric space. Several sensor rays might then intersect the same global grid cell and several cells might not be intersected by any ray. The usual practice when several rays intersect the same G_{XYZ} is to take the maximum occupancy probability of measurements projecting in that cell. On the other hand, holes in the reconstruction can either be left empty by attributing a prior $P(O) = 0.5$, or filled by interpolation of the nearest neighbor measurements. In this work we compare 2 design choices:

- No hole filling (sparse projection). $P(G_{XYZ}|R_tT_t) = P(O) = 0.5$, for all (X, Y, Z) where no sensor measurement is projected.
- Hole filling by nearest neighbor attribution. $P(G_{XYZ}|R_tT_t) = P(O_n|E^{(s)})$, for all (X, Y, Z) where no sensor measurement is projected. n and s in this case define the nearest cell in the set of sensor rays R_t. Such an approach thus corresponds to a prior on a piecewise smooth world geometry: closely located cells are likely to have similar occupancy probabilities.

D. Time filtering

Assuming each cell (X, Y, Z) to have a static state, a binary Bayes filter can be used to compute the probability $P(G_{XYZ}|R_{0:t}T_{0:t})$, where $R_{0:t} = R_0, R_1, ..., R_t$ and $T_{0:t} = T_0, T_1, ..., T_t$. To avoid numerical problems near probabilities 1 and 0, we use a log odds version of the filter as in [13]:

$$l_{t,XYZ} = l_{t-1,XYZ} + log\frac{P(G_{XYZ}|R_tT_t)}{1 - P(G_{XYZ}|R_tT_t)} - l_0, \quad (8)$$

where l_0 and $l_{t,XYZ}$ are defined as

$$l_{t,XYZ} = log\frac{P(G_{XYZ}|R_{0:t}T_{0:t})}{1 - P(G_{XYZ}|R_{0:t}T_{0:t})}, \quad (9)$$

$$l_0 = \frac{P(O)}{1 - P(O)}. \quad (10)$$

IV. STEREO MODELS FOR CCOGs

Consider two images I_1 and I_2, aligned along an x axis. In stereo vision, the cost-curve E of assigning $I_2(x, y)$ to $I_1(x + d, y)$ is computed for each pixel (x, y) and $d = N - i$ is called disparity. The cost-curve's global minimum, which occurs at $i_{min} = N - d_{min}$, is here referred to as E_{min}. Common cost functions for E are the Sum of Squared Differences (SSD), Sum of Absolute Differences, different forms of Correlation and others [14]. In this work we use the SSD since it is used by the first likelihood model.

A. Merrell's model

Merrel et al. [9] proposed the following stereo model:

$$p(E_i|O_iV_i) \propto exp\left(-\frac{(E_i - E_{min})^2}{2\sigma^2_{Mer}}\right), \quad (11)$$

where E_i is a cost function value (e.g. Sum of Squared Differences) and σ^2_{Mer} is a parameter which depends on image noise. In this paper and similarly to [11], we compute the maximum likelihood estimate of σ^2_{Mer} as the variance of all E_{min} in each stereo pair $\hat{\sigma}^2_{Mer} = Var(E_{min})$.

B. Matthies' model

This model was originally proposed in [15] by Matthies and Okutomi. It is defined as

$$p(E_i|O_iV_i) \propto exp\left(-\frac{E_i^TE_i}{2\sigma^2_{px}}\right), \quad (12)$$

where E_i is a vector containing the pixel differences inside the support window at cell i (note that $E_i^TE_i$ is thus a Sum of Squared Differences cost function). Parameter σ^2_{px} represents pixel intensity noise variance. In [11] we proposed to compute the maximum likelihood estimate $\hat{\sigma}^2_{px} = Var(I_2(x, y) - I_1(x + d_{min}, y))$ from all pixels (x, y) in each stereo pair.

C. Winner-take-all model

A winner-take-all (WTA) model can be defined as

$$p(E|O_iV_i) \propto exp\left(-\frac{(i - i_{min})^2}{2\sigma^2_i}\right). \quad (13)$$

In practice, in stereo-vision applications, σ^2_i is often assumed to be dependent solely on cell size [7] and $\sigma^2_i = 0.5^2$. Equation (13) is thus often approximated by

$$p(E|O_iV_i) = \begin{cases} 1, & \text{if } i = i_{min} \\ 0, & \text{otherwise} \end{cases}, \quad (14)$$

which is the model we used in our experiments.

V. EXPERIMENTAL EVALUATION

We evaluated the performance of the proposed CCOGs by comparing the final $P(G_{XYZ}|R_{0:t}T_{0:t})$ obtained on a stereo-vision sequence to its ground-truth occupancy grid. The dataset used was the KITTI dataset of outdoor stereo images [16] which is publicly available. Specifically, we used the residential area dataset "2011_09_26_drive_0079", where a stereo-vision equipped car is driven on a static environment

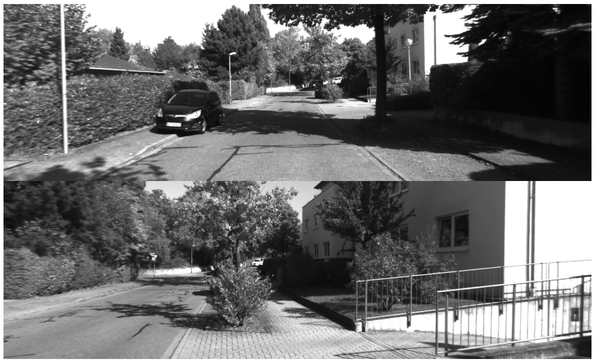

Fig. 1. Frames $t = 0$ and 50 of the used dataset: publicly available KITTI residential area dataset "2011_09_26_drive_0079".

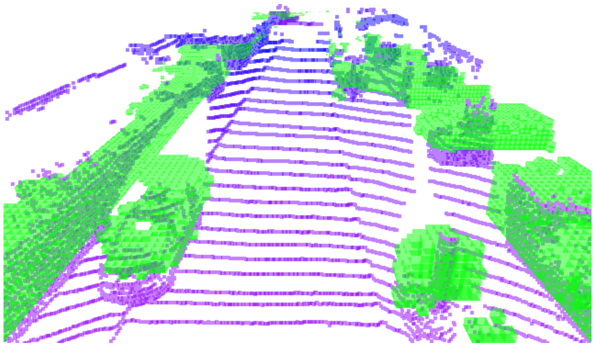

Fig. 2. Ground-truth grid G^* obtained from the laser-rangerfinder data. Cells are marked in green over the laser point data. Cells at the ground level were discarded for better visualization.

(i.e. no moving obstacles). Two frames of the image sequence are shown in Figure 1. The dataset contains a sequence of stereo pairs synchronized with laser-rangefinder measurements and localization data. We suppose pixel intensity noise σ^2 of the images is dependent only on the cameras and thus estimated it from the average variance of pixel intensity on the no-car-movement dataset "2011_09_28_drive_0043". Our noise estimate was $\sigma^2 = 13$.

We estimated a ground-truth grid G^* from the laser-rangefinder data by considering cells that were occupied with point data in more than 10 frames as occupied, and the rest as free. Cells were set as cubes of 0.20 meters each side. We show G^* in Figure 2 as green-colored squares over the laser data. Each grid G was compared with G^* by counting the following quantities:

- The number of true positives tp, i.e. the number of cells that satisfy $P(G_{XYZ}|R_{0:t}T_{0:t}) > 0.5$ and G^*_{XYZ};
- The number of false positives fp, i.e. the number of cells that satisfy $P(G_{XYZ}|R_{0:t}T_{0:t}) > 0.5$ and $\overline{G}^*_{XYZ}$;
- The number of true negatives tn, i.e. the number of cells that satisfy $P(\overline{G}_{XYZ}|R_{0:t}T_{0:t}) > 0.5$ and $\overline{G}^*_{XYZ}$
- The number of false negatives fn, i.e. the number of cells that satisfy $P(\overline{G}_{XYZ}|R_{0:t}T_{0:t}) > 0.5$ and G^*_{XYZ}.

We focused our evaluation on two criteria:

- Precision of G. $precision = \frac{tp}{tp+fp}$
- Recall of G. $recall = \frac{tp}{tp+fn}$

TABLE I
PRECISION (P) AND RECALL (R) IMPROVEMENT OBTAINED WITH NN HOLE FILLING WHEN COMPARED TO A SPARSE PROJECTION

	P_{start}	P_{end}	R_{start}	R_{end}
Mean improvement	-0.0891	0.0878	0.089	0.120
Std-dev improvement	0.109	0.026	0.033	0.037

Using these criteria, we experimentally quantify the influence of the following design choices on CCOG performance:

A. Hole filling

We measured the precision and recall ratios at the start of the image sequence ($t = 0$) and after integration of 20 stereo pairs (t = 0,5,10,...,95), for all stereo models and different cost function window sizes. When compared to a sparse grid, both ratios improved with the use of a dense nearest neighbor hole filling strategy. In Table I we show the average improvements of precision P computed as $(P_{nn} - P_{sparse})$ and recall R as $(R_{nn} - R_{sparse})$. Hole filling by setting occupancy values to the nearest stereo measurement basically corresponds to a prior on continuous world geometry. Even if certain false-positive problems arise from such approach at the first frames (i.e. there is low precision at the start due to a bad reconstruction far away from the camera), our data indicates that it can nevertheless lead to an important increase in both precision and recall over time. As more measurements are filtered both precision and recall of the map increase more in the case of hole filling (see P_{end}, R_{end} in Table I). Such hole filling approaches are also common in multi-view stereo literature [9], [10] where piece-wise linear disparity maps are assumed. Our experiments also indicate this assumption to have a positive impact to reconstruction results.

B. Stereo model

In Figure 2 we show the ground-truth grid G^*, where cells at the floor level were discarded for better visualization. Figure 3 shows the obtained reconstruction using CCOGs with Merrell's and Matthies' models (cost function window size 13x13, nearest neighbor hole filling, after integration of all images t = 0,5,10,...,95). Merrell's model lead to a more reliable reconstruction with a low number of false-positives, even though objects were reconstructed mainly on textured regions. It is clear that, for example, the car on the lower left corner of the image is only reconstructed on visual edges (the corners of the car). Salient obstacles such as the tree on the lower right corner and bushes and fence on the upper right were, however, either partly or fully reconstructed. False-positives are shown in brown color and are very scarcely located, often close to real objects in G^*. Matthies' model, on the other hand, has good reconstruction on low-textured objects (e.g. the wall on the right, a higher portion of the car) however at the cost of a high false-positive ratio.

Using a WTA model (i.e. discarding all costs except the minimum), on the other hand, leads to an even higher number of false positives, which shows the advantage of exploiting

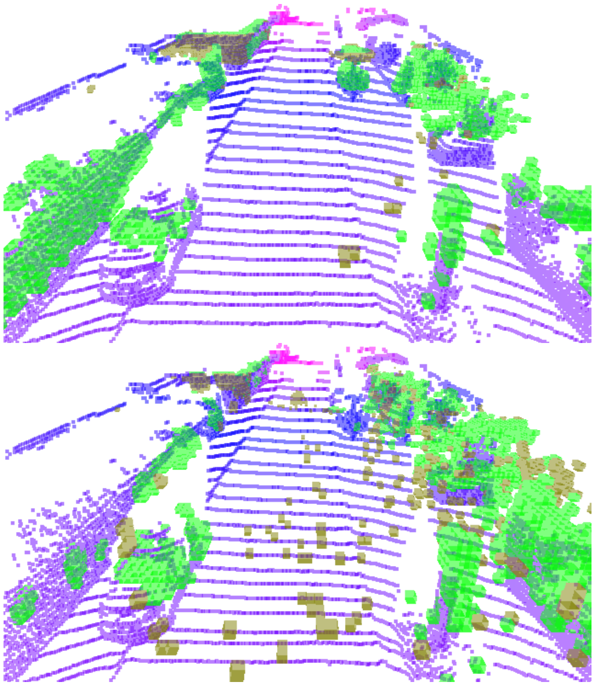

Fig. 3. Top: Merrell's model. Bottom: Matthies' model. Green squares represent true-positives (i.e. cells correctly classified as occupied), brown squares represent false-positives (i.e. cells incorrectly classified as occupied).

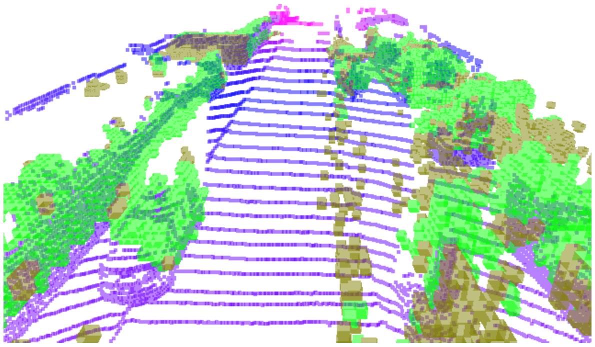

Fig. 4. Grid obtained using a traditional winner-take-all (WTA) model which discards the whole cost-curve except its minimum. Green squares represent true-positives, brown squares represent false-positives.

the cost-curve of stereo in CCOGs. The grid obtained with WTA is shown in Figure 4. In Figure 5 we show the obtained precision and recall ratios over time for the 3 stereo models, using nearest neighbor hole filling. Each dot represents a new stereo pair being used to update the grid, corresponding to frames t = 0,5,10,...,95. Both precision and recall increased considerably over time. Precision increased with cost function window size, while recall slightly decreased. An explanation for the lower recall of larger window sizes is that such cost-curves have less distinctive minima, which after the normalization in equation (7) leads to low occupancy probabilities. Stereo models including the whole cost-curve (i.e. Matthies, Merrell's) noticeably increased precision of the grids with respect to a WTA model, thus showing the advantages of CCOGs and whole cost-curve approaches to mapping in general. We obtained precision above 0.8 for large window sizes, while recall was around 50%. Merrell's model had the highest precision and second highest recall, while Matthies' model is better than WTA for smaller window sizes. The WTA model lead to the lowest precision grids and high recall, which indicates a grid where the number of false positives is dangerously high (as seen in Figure 4).

TABLE II

MAXIMUM ACCEPTABLE IMAGE NOISE VARIANCE σ^2 FOR DESIRED GRID PRECISION

		Mat	Mer	WTA
SSD 5x5	Min. precision 0.65	–	–	–
	Min. precision 0.75	–	–	–
	Min. precision 0.85	–	–	–
SSD 9x9	Min. precision 0.65	25	83	15
	Min. precision 0.75	14	43	–
	Min. precision 0.85	–	–	–
SSD 13x13	Min. precision 0.65	15	177	83
	Min. precision 0.75	13	83	14
	Min. precision 0.85	–	43	–

We also analyzed the influence of image noise in the performance of CCOGs. In Figure 6 we show results taken after integration of all images, but where different levels of Gaussian noise were added to the stereo pairs. Taking into consideration the original noise estimate of $\sigma^2 = 13$, the resulting pixel intensity noise variance levels tested were σ^2 = 13,14,15,18,25,43,83,177 and 397, where pixel intensity is in the range $[0; 255]$. Although larger cost function window sizes lead to higher robustness to noise, the performance of the occupancy grids quickly deteriorates in all cases. Matthies' model's recall deteriorates for 13x13 windows, as also seen in Figure 5. This is due to a very low parameter estimate (single pixel noise) when compared to the cost function value. Merrell's model lead to slightly slower precision deterioration than other models. For instance, on a 13x13 window size and if the minimum desired precision is 0.75, the maximum allowed image noise is 13, 83 and 14, for Matthies', Merrell's and WTA models respectively. In Table II we present the maximum allowed image noise for different values of minimum precision.

Finally we estimated the relationship between grid precision and noise variance. After fitting polynomial, exponential and power functions to the $precision(noise)$ data we found a function of the power of noise, $precision(\sigma^2) = a{*}(\sigma^2)^b +$ c, to be the best fit in all models (SSE $\in [0.0003; 0.004]$).

VI. CONCLUSIONS

We introduced the CCOG, a grid-based 3D reconstruction method that estimates occupancy in sequences of stereo pairs. The key feature of the method is that occupancy is computed not from the least-cost estimate of distance given by stereo, but from the likelihoods of all costs along the cost-curve.

We showed that our approach leads to reconstructions with higher precision and more robustness to image noise than those obtained by only considering least-cost distance hypotheses. We also evaluated the performance of CCOGs under different design choices. We showed that hole filling with a nearest neighbor strategy, despite the strong geometry

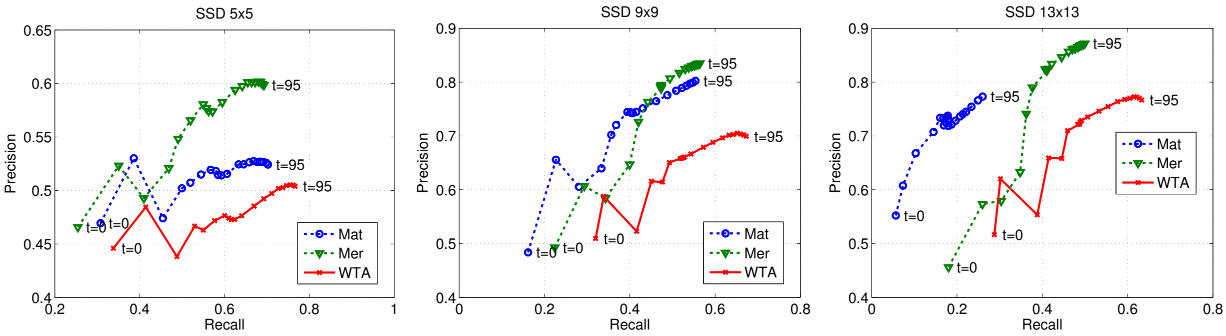

Fig. 5. Precision and recall ratios obtained over time, from $t = 0$ to $t = 95$, using different stereo models.

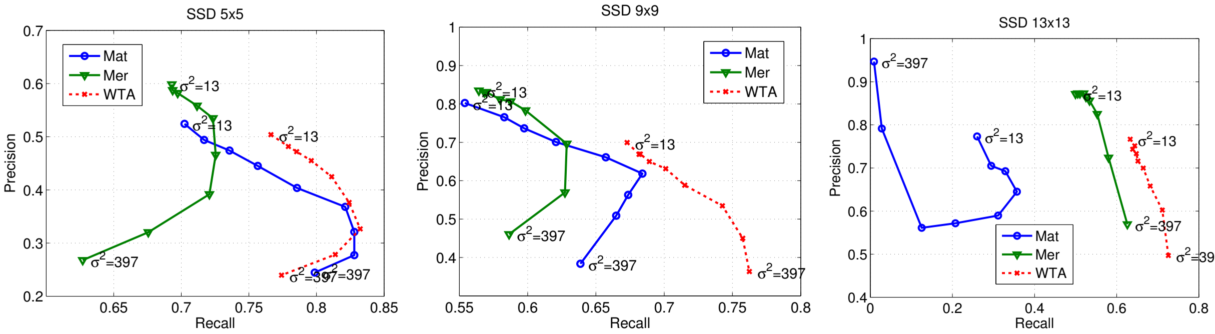

Fig. 6. Precision and recall obtained after integration of all images (t = 0,5,...,95), for different values of image noise σ^2 = 13,14,15,18,25,43,83,177,397.

prior it assumes, can lead to grids with not only higher recall but higher precision as well. Using cost functions with larger window sizes naturally lead to better results, while the stereo confidence model proposed by Merrell et al. in [9] provided the most reliable reconstructions and noise robustness. Each model seems to be better for different image texture conditions, which makes research into combining different confidence models relevant. Precision of CCOGs was observed to be a power function of image noise.

References

[1] A. Elfes, “Sonar-based real-world mapping and navigation,” vol. 3, no. 3, pp. 249–265, 1987.

[2] D. Murray and J. J. Little, “Using real-time stereo vision for mobile robot navigation,” *Autonomous Robots*, vol. 8, no. 2, pp. 161–171, 2000.

[3] L. Matthies and A. Elfes, “Integration of sonar and stereo range data using a grid-based representation,” *1988 IEEE International Conference on Robotics and Automation*, pp. 727–733, 1988.

[4] S. Thrun, “A probabilistic online mapping algorithm for teams of mobile robots,” *International Journal of Robotics Research*, vol. 20, p. 2001, 2001.

[5] H. Badino, U. Franke, and R. Mester, “Free space computation using stochastic occupancy grids and dynamic programming,” in *Workshop on Dynamical Vision, ICCV*, October 2007.

[6] F. Andert, “Drawing stereo disparity images into occupancy grids: measurement model and fast implementation,” in *IEEE International Conference on Intelligent Robots and Systems*, 2009, pp. 5191–5197.

[7] M. Perrollaz, A. Spalanzani, and D. Aubert, “Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection,” *2010 IEEE Intelligent Vehicles Symposium*, pp. 313–318, June 2010.

[8] D. Pfeiffer, S. Gehrig, and N. Schneider, “Exploiting the power of stereo confidences,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, pp. 297–304.

[9] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nistér, and M. Pollefeys, “Real-time visibility-based fusion of depth maps,” in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[10] X. Hu and P. Mordohai, “Least Commitment, Viewpoint-Based, Multi-view Stereo,” in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, Oct 2012, pp. 531–538.

[11] M. Brandao, R. Ferreira, K. Hashimoto, J. Santos-Victor, and A. Takanishi, “Integrating the whole cost-curve of stereo into occupancy grids,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4681–4686.

[12] R. A. Newcombe, S. Lovegrove, and A. Davison, “Dtam: Dense tracking and mapping in real-time,” in *2011 IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 2320–2327.

[13] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[14] H. Hirschmüller and D. Scharstein, “Evaluation of stereo matching costs on images with radiometric differences.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582–99, Sept. 2009.

[15] L. Matthies and M. Okutomi, “A Bayesian foundation for active stereo vision,” *Proc. SPIE Sensor Fusion II: Human and Machine Strategies*, pp. 1–13, 1989.

[16] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.