On fusion of range and intensity information using Graph-Cut for planar patch segmentation

Olaf Kähler*, Erik Rodner
and Joachim Denzler

Friedrich-Schiller-University of Jena,
Faculty of Mathematics and Computer Science,
Chair for Computer Vision,
Ernst-Abbe-Platz 2, 07743 Jena, Germany
Fax: +49 3641 946372
E-mail: kaehler@informatik.uni-jena.de
E-mail: rodner@informatik.uni-jena.de
E-mail: denzler@informatik.uni-jena.de
*Corresponding author

Abstract: Planar patch detection aims at simplifying data from 3D imaging sensors to a more compact scene description. We propose a fusion of intensity and depth information using Graph-Cut methods for this problem. Different known algorithms are additionally evaluated on low-resolution high-frame rate image sequences and used as an initialisation for the Graph-Cut approach. In experiments, we show a significant improvement of the detected patch boundaries after the refinement with our method.

Keywords: graph cut; planar patches; segmentation.

Reference to this paper should be made as follows: Kähler, O., Rodner, E. and Denzler, J. (2008) 'On fusion of range and intensity information using Graph-Cut for planar patch segmentation', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 5, Nos. 3/4, pp.365–373.

Biographical notes: Olaf Kähler recieved the Diploma degree in Computer Science from University Erlangen in 2004. Since 2005, he has been working as a PhD student at the Chair for Computer Vision in Jena. His research interests focus on sensor data fusion, tracking, 3d computer vision and especially the structure-from-motion problem.

Erik Rodner received the Diploma degree in Computer Science with honours in 2007 from the University of Jena, Germany. He is currently a PhD student under supervision of Denzler at the Chair for Computer Vision, University of Jena. His research interests include discrete optimisation and object recognition with few training examples.

Joachim Denzler, born in 16 April 1967, got the degree 'Diplom-Informatiker', 'Dr.-Ing.' and 'Habilitation' from the University of Erlangen in the year 1992, 1997 and 2003, respectively. Currently, he holds a position of a full Professor for Computer Science and is the Head of the Chair for Computer Vision, Faculty of Mathematics and Informatics, Friedrich-Schiller-University of Jena. His research interests comprise active computer vision, object recognition and tracking, 3d reconstruction and plenoptic modelling as well as computer vision for autonomous systems. He is an author and a co-author of over 90 journal papers and technical articles. He is a member of the *IEEE, IEEE computer*

society, DAGM and *GI*. For his work on object tracking, plenoptic modelling, and active object recognition and state estimation he was awarded with DAGM best paper awards in 1996, 1999, and 2001, respectively.

1 Introduction

Finding a simple description of a scene observed with a camera is one of the main goals of computer vision. Acquiring 3D information in terms of depth images became possible due to recent advances in measurement technologies (Lange, 2000). However, efficient interpretation and simplification of this data is a current research topic. In typical man-made environments, planar features are abundant and provide useful features for scene interpretation. They have been used for robot navigation (Cobzas and Zhang, 2001), for coarse-registration of 3D point clouds (von Hansen, 2006), modelling of the environment (Gorges et al., 2004) and layered representation of video (Xiao and Shah, 2005) among many other applications.

As detecting and segmenting planar patches is typically used as preprocessing for further steps, an accurate and reliable method is needed. Especially for modelling-tasks, a detailed, pixel-wise segmentation is also important. We propose and compare methods for finding an accurate and simple description of the scene as a collection of planar regions. In a first step, different known methods of finding initial seed regions are employed. As a novelty, the detected regions are then refined and accurately segmented with a Graph-Cut technique. For both steps, information is taken from measured image intensities and from measured depth information.

In the intensity domain, the plane induced homographies between multiple images are of primary importance for the detection of planes (Gorges et al., 2004; Kähler and Denzler, 2007). However, this will only work, if a camera motion is present. In addition, homogeneity (Cobzas and Zhang, 2001) and edges (Xiao and Shah, 2005) can be cues on the exact outlines of planar patches, even with single still images. For the domain of depth images, no camera motion is needed. Typically, region-growing approaches (Hoover et al., 1996; Cobzas and Zhang, 2001) are employed as simple but effective methods. In this work, we use all of the above-mentioned methods to initialise a fine-segmentation based on Graph-Cut methods (Boykov, Veksler and Zabih, 2001; Kolmogorov and Zabin, 2004; Freedman and Drineas, 2005). Recently such methods became popular to efficiently solve discrete optimisation problems in computer vision. In addition, the detection of planar patches from motion information alone has been approached with such methods (Xiao and Shah, 2005). As the segmentation problem is formulated with an objective function, it is easy to incorporate different sources of information, like the depth and intensity information acquired with a 3D imaging sensor.

First, we will review methods for detecting the initial seed regions either from intensity images or from depth images. Next we will show, how a refinement of the detected planes can be achieved using Graph-Cut. In an experimental evaluation, we will compare the different approaches of detecting the planes in images from a current 3D imaging sensor, the Photonic Mixer Device (PMD) vision 19k camera. We will demonstrate the benefits of the refinement stage using the proposed Graph-Cut methods. A summary of our findings will conclude the paper.

2 Detection of seed regions

As a first step in our system, seed regions have to be detected in the images. Both the information contained in the intensity images and the measured depth information can be used for this purpose. In the following, we will shortly present methods for both of the two information sources.

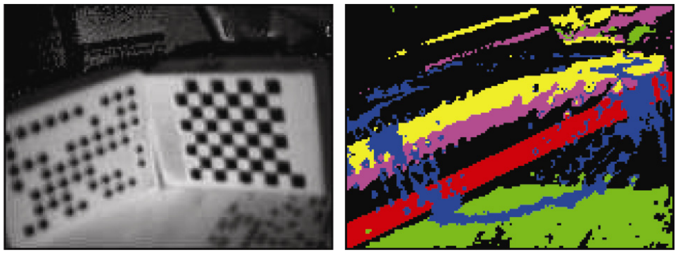
2.1 Intensities and homographies

A single intensity image alone cannot be used to infer on any 3D information. If a sequence of images with camera motion is available however, classical approaches like structure from motion can be applied. In particular, the detection of planar patches using homographies has been studied before (Gorges et al., 2004; Kähler and Denzler, 2007).

In these approaches, point correspondences are established between successive images of a sequence using standard techniques like Kanade-Lucas-Tomasi (KLT)-tracking (Shi and Tomasi, 1994). Despite the low resolution of current 3D imaging sensors, it is possible to track points through an image sequence, if the scene is sufficiently textured. Feature points on one common 3D plane satisfy a homography between two images of a sequence. Starting from a list of known point correspondences, the next step is hence to detect subsets of correspondences fulfiling common homographies. This can be accomplished e.g. with the Random Sample and Consensus (RANSAC)-algorithm.

Not all point sets satisfying a homography are necessarily related to a 3D plane. Two major problems are 'virtual' planes and pairs of images without a camera translation (Kähler and Denzler, 2007). Virtual planes are geometrically valid 3D planes, that do not correspond to any physically present scene plane, as shown in Figure 1. To avoid such planes in the detection steps and to achieve contiguous planar patches, it was proposed to modify the RANSAC approach not to select points randomly, but in a local neighbourhood (Gorges et al., 2004). As another major problem of intensity based detection of planar patches, such methods will only work for non-zero camera translations. Cases of purely rotating or still cameras can be detected automatically (Kähler and Denzler, 2007), but with depth information gathered by the sensor, it seems an unnecessary restriction to detect coplanarity only in cases of a moving camera. Hence, we will focus on extraction of planar patches from depth information in the next section.

Figure 1 Examples for 'virtual planes': points of the same colour in the right image reside on common 3D planes, which are not corresponding to any physical plane (see online version for colours)



2.2 Range images for detecting planes

Various algorithms (Hoover et al., 1996) have been proposed for classical range image segmentation into planar surfaces. Usually they work with single still images and only few take image intensities into account (Cobzas and Zhang, 2001). Altogether region growing approaches have proved as simple and efficient methods for this segmentation task.

In our work, we use a region growing algorithm basically as it is described in Cobzas and Zhang (2001) to get an initial estimate of planar patches from the depth images. Starting with random seed points in the images, pixels in the eight-neighbourhood are iteratively added to the regions, as long as the continuity criterion of coplanarity is satisfied. This criterion is evaluated by fitting a plane to the points currently contained in the region, and applying a threshold to the distance of new points to the estimated plane.

Let $x^{(3D)}$ denote the 3D position of the point mapped to pixel x in the image. For convenience, let the origin of the world coordinate system be identical with the optical centre of the camera. The normal $\mathbf{n}$ of plane P fitting to a set of pixels $\mathbf{x} \in S$ is then computed with the M-estimator according to:

$$\hat{n} = \arg\min_{\mathbf{n}} \sum_{\mathbf{x} \in s} \rho\left(\mathbf{n}^{\mathrm{T}} \mathbf{x}^{(3\mathrm{D})} - 1\right) \tag{1}$$

In our implementation, ρ is set to the Huber function and the computation is performed efficiently with an Iteratively Reweighted Least Squares (IRLS) approach. Note that with the chosen parameterisation of planes, it is not possible to represent planes with zero distance to the origin of the world coordinate system. However, as this origin coincides with the optical centre of the camera, such planes are not visible in the images at all. For each candidate pixel in the region growing algorithm, the distance $\delta(\mathbf{x})$ of the corresponding 3D point to the given plane P is then:

$$\delta(\mathbf{x}) = \frac{1}{\|\hat{n}\|} \left| \hat{n}^{\mathrm{T}} \mathbf{x}^{(3\mathrm{D})} - 1 \right|. \tag{2}$$

Points that are eight-connected to the region and have a distance below the threshold δ_{thresh} are considered to be coplanar with P and hence added to the region. After each step, a new plane normal is estimated, and the whole procedure is iterated, until no more points can be added to the plane.

3 Refining patch boundaries using Graph-Cut

As a novelty, we present a method for refining the detected seed regions with the goal of finding more accurate boundaries for the planar patches. Instead of using only local decisions, as in the region growing methods before, we use discrete optimisation with Graph-Cut (Boykov, Veksler and Zabih, 2001) to find the outlines of a planar patch in one step. Further, this allows to formulate a single objective function incorporating different sources of information, like the 3D data from the range images, edge information from the intensity images or even information from point correspondences and estimated homographies. First, we will formulate the segmentation problem mathematically and then present the objective functions used in the optimisation.

3.1 Problem formulation

Let R_P denote the set of all pixel coordinates being projections of plane *P*. With the previous detection steps, typically an approximation *S* of R_P was found. With noiseless depth information, R_P corresponds to the zero level set of the function δ from Equation (2). Due to noise and a potentially high amount of outliers, it is not possible to find an estimate of R_P by a local threshold decision in δ. Instead, the estimation is done using discrete optimisation in the form of Graph-Cut methods.

The problem is formulated as a binary labelling problem of the whole image *I*. A label $l(\mathbf{x})$ is assigned to each point $\mathbf{x} \in I$ with $l(\mathbf{x}) = 1$ if and only if $\mathbf{x} \in R_P$. With the use of discrete optimisation, the problem of finding an optimal labelling reduces to the minimisation of the objective function:

$$E(l) = \sum_{\mathbf{x}} \mathrm{D}(\mathbf{x}, l(\mathbf{x})) + \lambda \sum_{(\mathbf{x},\mathbf{y}) \in N} V(\mathbf{x},\mathbf{y}) T(l(\mathbf{x}) \neq l(\mathbf{y})), \tag{3}$$

where *N* is a neighbourhood, like the set of all eight-connected pixels. Further, D(**x**, l(**x**)) is the cost of assigning a given label to a single pixel independent from its neighbourhood, V (**x**, **y**) is a penalty imposed by a segmentation boundary between **x** and **y** (also called smoothness cost), and $T(\cdot)$ is 1, if its argument is true, and zero otherwise. Parameter λ controls the influence of neighboured pixels and therefore the 'smoothness' of the segmentation result.

This kind of energy function is known as the Generalised Potts Model (cf. Boykov, Veksler and Zabih, 2001) and the minimisation of the function can be mapped to the well-known 'Min-Cut'- problem in graph theory (Kolmogorov and Zabin, 2004), which can be solved in polynomial time.

3.2 Cost functions for the segmentation problem

To select data and smoothness functions, we will follow the suggestions of (Xiao and Shah, 2005). First, we will define the data cost function *D*, using the following construction to avoid problems with the image border and unavailable depth information:

$$D^{\text{plane}}(\mathbf{x}, l) = \begin{cases} 0 & \text{no depth information available at } \mathbf{x} \\ -\infty & 1 = 0 \wedge \mathbf{x} \text{ is on the image border} \\ \tilde{D}^{\text{plane}}(x, l) & \text{otherwise.} \end{cases} \tag{4}$$

The cost ∞ corresponds to a suitably large value, which ensures the fixation of the labels at the image border. Further, a sigmoidal function s(·, α, β) is chosen, with a phase transition defined by β and the length of the transition selected by α:

$$s(\mathbf{x}, \alpha, \beta) = \frac{1}{2} + \frac{\tan^{-1}(\alpha(\mathbf{x} - \beta))}{\pi} \tag{5}$$

$$\tilde{D}^{\text{plane}}(\mathbf{x}, l) = \begin{cases} s(\delta(\mathbf{x}), \alpha_p, \beta_p) & l = 1 \\ 1 - s(\delta(\mathbf{x}), \alpha_p, \beta_p) & l = 0. \end{cases} \tag{6}$$

The smoothness function $V(\mathbf{x}, \mathbf{y})$ controls the cost of the segmentation boundary, which should prefer discontinuities in the image (or in general arbitrary real valued functions defined on a grid). Hence, we define V with parameters γ>1.0 and κ_p:

$$V^{\text{plane}}(\mathbf{x},\mathbf{y}) = \begin{cases} \gamma & |\delta(\mathbf{x}) - \delta(\mathbf{y})| < \kappa_p \\ 1.0 & \text{otherwise.} \end{cases} \tag{7}$$

With the assumption of a strong contour forming the boundary of R_P in the intensity image I, we are able to penalise non-edges in I analogous to (7):

$$V^{\text{intensity}}(\mathbf{x},\mathbf{y}) = \begin{cases} \gamma & |I(\mathbf{x}) - I(\mathbf{y})| < \kappa_e \\ 1.0 & \text{otherwise.} \end{cases} \tag{8}$$

In conjunction with the data cost D^{plane}, these boundary costs lead to a segmentation combining depth and grey-value information. For both cases of the smoothness cost, we are now able to give an explicit form of the objective function (3):

$$E^{\text{depth}} = \sum D^{\text{plane}}(\cdot) + \lambda \sum V^{\text{plane}}(\cdot)T \tag{9}$$

$$E^{\text{intensity}} = \sum D^{\text{plane}}(\cdot) + \lambda \sum V^{\text{intensity}}(\cdot)T. \tag{10}$$

4 Experiments

To show their applicability, we experimentally evaluated the presented methods. First, we will present the data used for this evaluation, and then show comparative results for the various approaches from the previous sections.

4.1 Experimental method

For a quantitative evaluation, three image sequences with highly textured, planar calibration patterns were recorded with a ‘PMD vision 19k’ 3D camera. For these sequences, planar regions were manually labelled, such that a ground truth comparison is possible. Further sequences of more realistic environments and less artificial texture were used for a qualitative evaluation. For the ground truth comparison, we compute the symmetric set distance between an estimated region A and ground truth region B by

$$e(A,B) = \frac{|A \setminus B| + |B \setminus A|}{|A| + |B|}, \tag{11}$$

which is zero in the case of identical sets and 1 if A and B are disjoint. As the number of ground truth regions and detected regions is not necessarily the same, the best pair of ground truth regions and estimated regions in the sense of minimal e were compared at each time.

4.2 Compared approaches and results

For all approaches, some elementary preprocessing steps were performed on the raw data from the PMD MiniSDK. The intensity images were clipped to the range [0–255]. The depth measurements for pixels were discarded, if the measured amplitude of the reflected active illumination was below a threshold of 2.0. Further, a 3 × 3 median filter was applied to the measured depths.

Initial planar patches are estimated using the method of (Kähler and Denzler, 2007) revised in Section 2.1 and the region growing approach from Section 2.2. The refinement stage with Graph-Cut, was either based on depth information alone, or on both depth and intensity information, as presented in Section 3.2. For all experiments, the parameters from Table 1 were used. Examples for the initial detection stage and the effects of the different refinements are shown in Figure 2. The average differences between the ground truth regions and those, detected with different combinations of the presented approaches, are given in Table 2. These values were computed from the symmetric set differences, as explained in Section 4.1.

Figure 2 Example for the result of our plane detection algorithms: (left) initial plane region (centre) refining with depth information (right) refining with grey-value and depth information (see online version for colours)

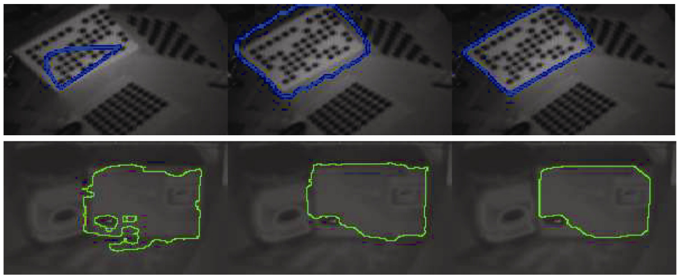
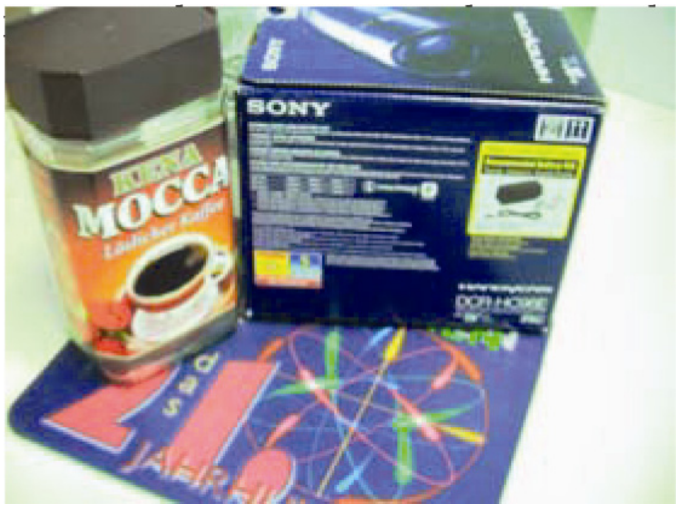


Figure 3 View of the scene used to generate the second example of Figure 2 (see online version for colours)



The quantitative evaluation of our method in Table 2 reveals that planar patches refined with Graph-Cut in the second stage always have a significant smaller difference to the ground truth plane, i.e. they more accurately correspond to them. Further, the use of edge information slightly reduces the segmentation error over using only 3D information.

Excerpts from the plane detection results in more natural environments are shown in Figure 2. Note that due to the low texture and coarse resolution of 160 × 120 pixels, point tracking and the methods based on homographies do not provide a useful initialisation for the plane segmentation in the shown sequence. Instead the region growing scheme was employed in these examples.

Table 1 Parameters

Parameter	*Value*	*Parameter*	*Value*
γ	3	β_p, κ_p	$0.03 \cong 3$cm
λ	1/2	δ_{thresh}	$0.01 \cong 1$cm
α_p	4.10^3	κ_e	8

Table 2 Average of *e* calculated for various scenes and presented methods

Method	*seq1*	*seq2*	*seq3*
Kähler and Denzler (2007) + Graph-Cut refining $E^{\text{intensity}}$	0.153	0.244	0.214
Kähler and Denzler (2007) + Graph-Cut refining E^{depth}	0.184	0.252	0.217
Kähler and Denzler (2007)	0.462	0.345	0.260
Region growing + Graph-Cut refining $E^{\text{intensity}}$	0.178	0.239	0.202
Region growing + Graph-Cut refining E^{depth}	0.213	0.249	0.205
Region growing	0.556	0.476	0.459

5 Conclusions and further work

We presented a novel approach to detect planes with 3D imaging sensors. Initialisation methods to detect planar patches were followed by a refinement step using a Graph-Cut approach. We were able to combine edge information from intensity images with depth information to enhance segmentation results.

The initial planar patches can be detected with point tracking and homographies in the intensity domain (Kähler and Denzler, 2007). This unnecessarily restricts the algorithms to a moving camera and a highly textured scene, however. Instead, also region growing approaches can be applied in the depth domain. Due to the noisy depth information, the local decisions of region growing are not sufficient for an accurate segmentation into planar regions. With the proposed Graph-Cut methods, a global optimality can be achieved. The additional use of intensity information further benefits the plane segmentation problem. In the presented experiments, we only used edge information from the intensities. The information gained from homographies is limited by the extremely low resolution of only 160 × 120 pixels of currently available 3D imaging sensors. As shown by Xiao and Shah (2005), intensity images of higher resolution are well suited for plane detection. Hence, it seems promising to further promote a combination of the different measurement modalities with higher resolution images.

Finally note, the presented binary labelling problem can also be extended to a simultaneous segmentation of all initial planar patches using the α-expansion algorithm (Boykov, Veksler and Zabih, 2001). The selection of energy functions for the Graph-Cut algorithm is quite difficult however, due to a lack of theoretical investigation on the properties of the segmentation results. In addition, more general models, such as general graphrepresentable functions with high order clique potentials (Freedman and Drineas, 2005), offer the possibility to incorporate more a *priori* knowledge or even skip an initialisation step.

References

Boykov, Y., Veksler, O. and Zabih, R. (2001) 'Fast approximate energy minimisation via graph cuts', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, pp.1222–1239.

Cobzas, D. and Zhang, H. (2001) 'Planar patch extraction with noisy depth data', Paper presented in the Proceedings the *Third International Conference on 3D Digital Imaging and Modelling*, (pp.240–245).

Freedman, D. and Drineas, P. (2005) 'Energy minimisation via graph cuts: Settling what is possible', Paper presented in the Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp.939–946), Washington, DC, USA.

Gorges, N., Hanheide, M., Christmas, W., Bauckhage, C., Sagerer, G. and Kittler, J. (2004) 'Mosaics from arbitrary stereo video sequences', *Lecture Notes in Computer Science* (Vol. 3175, pp.342–349), Heidelberg, Germany. 26th DAGM Symposium, Springer-Verlag.

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A.W. and Fisher, R.B. (1996) 'An experimental comparison of range image segmentation algorithms', *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 18, pp.673–689.

Kähler, O. and Denzler, J. (2007) 'Detecting coplanar feature points in handheld image sequences', Paper presented in the Proceedings of the *Conference on Computer Vision Theory and Applications (VISAPP 2007)* (Vol. 2, pp.447–452), Barcelona: INSTICC Press.

Kolmogorov, V. and Zabin, R. (2004) 'What energy functions can be minimised via graph cuts?', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, pp.147–159.

Lange, R. (2000) '3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology', *PhD thesis*, University of Siegen.

Shi, J. and Tomasi, C. (1994) 'Good features to track', *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pp.593–600.

von Hansen, W. (2006) 'Robust automatic marker-free registration of terrestrial scan data', Paper presented in the Proceedings of the *Photogrammetric Computer Vision 2006* (Vol. 36, pp.105–110), Bonn, Germany.

Xiao, J. and Shah, M. (2005) 'Motion layer extraction in the presence of occlusion using graph cuts', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, pp.1644–1659.