10

Non-Iterative Vision-Based Interpolation of 3D Laser Scans

Henrik Andreasson[1], Rudolph Triebel[2], and Achim Lilienthal[1]

[1] AASS, Dept. of Technology, Örebro University, Sweden
henrik.andreasson@tech.oru.se, achim.lilienthal@tech.oru.se
[2] Autonomous Systems Lab, ETH Zürich, Switzerland
rudolph.triebel@mavt.ethz.ch

Summary. 3D range sensors, particularly 3D laser range scanners, enjoy a rising popularity and are used nowadays for many different applications. The resolution 3D range sensors provide in the image plane is typically much lower than the resolution of a modern colour camera. In this chapter we focus on methods to derive a high-resolution depth image from a low-resolution 3D range sensor and a colour image. The main idea is to use colour similarity as an indication of depth similarity, based on the observation that depth discontinuities in the scene often correspond to colour or brightness changes in the camera image. We present five interpolation methods and compare them with an independently proposed method based on Markov random fields. The proposed algorithms are non-iterative and include a parameter-free vision-based interpolation method. In contrast to previous work, we present ground truth evaluation with real world data and analyse both indoor and outdoor data.

Keywords: 3D range sensor, laser range scanner, vision-based depth interpolation, 3D vision.

10.1 Introduction

3D range sensors are getting more and more common and are found in many different areas. A large research area deals with acquiring accurate and very dense 3D models, potential application domains include documenting cultural heritage [1], excavation sites and mapping of underground mines [2]. A lot of work has been done in which textural information obtained from a camera is added to the 3D data. For example, Sequeira et al. [3] present a system that creates textured 3D models of indoor environments using a 3D laser range sensor and a camera. Früh and Zakhor [4] generate photo-realistic 3D reconstructions from urban scenes by combining aerial images with textured 3D data acquired with a laser range scanner and a camera mounted on a vehicle.

H. Andreasson et al.: *Non-Iterative Vision-Based Interpolation of 3D Laser Scans*, Studies in Computational Intelligence (SCI) **76**, 83–90 (2007)
www.springerlink.com

Fig. 10.1. *Left*: Image intensities plotted with the resolution of the 3D scanner. The laser range readings were projected onto the right image and the closest pixel regions were set to the intensity of the projected pixel for better visualisation. *Middle*: Calibration board used for finding the external parameters of the camera, with a chess board texture and reflective tape (*grey border*) to locate the board in 3D using the remission/intensity values from the laser scanner. *Right*: Natural neighbours $\mathrm{R}_1 \ldots \mathrm{R}_5$ of R_i^*. The interpolated weight of each natural neighbour R_i is proportional to the size of the area which contains the points Voronoi cell and the cell generated by R_j^*. For example the nearest neighbour R_1 will have influence based upon the area of A_1

In most of the approaches that use a range scanner and a camera, the vision sensor is not actively used during the creation of the model. Instead vision data are only used in the last step to add texture to the extracted model. An exception is the work by Haala and Alshawabkeh [5], in which the camera is used to add line features detected in the images into the created model.

To add a feature obtained with a camera to the point cloud obtained with a laser range scanner, it is required to find the mapping of the 3D laser points onto pixel co-ordinates in the image. If the focus instead lies on using the camera as an active source of information which is considered in this chapter, the fusing part in addition addresses the question of how to estimate a 3D position for each (sub) pixel in the image. The resolution that the range sensor can provide is much lower than those obtained with a modern colour camera. This can be seen by comparing left of Fig. 10.1, created by assigning the intensity value of the projected laser point to its closest neighbours, with the corresponding colour image in middle of Fig. 10.1.

This chapter is a shortened version of [6].

10.2 Suggested Vision-Based Interpolation Approaches

The main idea is to interpolate low-resolution range data provided by a 3D laser range scanner under the assumption that depth discontinuities in the scene often correspond to colour or brightness changes in the camera image of the scene.

For the problem under consideration, a set of N laser range measurements $r_1 \dots r_N$ is given where each measurement $r_i = (\theta_i, \pi_i, r_i)$ contains a tilt angle θ_i, a pan angle π_i and a range reading r_i corresponding to 3D Euclidean coordinates (x_i, y_i, z_i).

The image data consists of a set of image pixels $\mathrm{P}_j = (X_j, Y_j, C_j)$, where X_j, Y_j are the pixel co-ordinates and $C_j = (C_j^1, C_j^2, C_j^3)$ is a three-channel colour value. By projecting a laser range measurement r_i onto the image plane, a projected laser range reading $\mathrm{R}_i = (X_i, Y_i, r_i, (C_i^1, C_i^2, C_i^3))$ is obtained, which associates a range reading r_i with the coordinates and the colour of an image pixel. An image showing the projected intensities can be seen in Fig. 10.1, where the closest pixel regions are set to the intensity of the projected pixel for better visualisation. The interpolation problem can now be stated for a given pixel P_j and a set of projected laser range readings R, as to estimate the interpolated range reading r_j^* as accurately as possible. Hence we denote an interpolated point $\mathrm{R}_j^* = (X_j, Y_j, r_j^*, C_j^1, C_j^2, C_j^3)$.

Five different interpolation techniques are described in this section and compared with the MRF approach described in Sect. 10.3.

10.2.1 Nearest Range Reading (NR)

Given a pixel P_j, the interpolated range reading r_j^* is assigned to the laser range reading r_i corresponding to the projected laser range reading R_i which has the highest likelihood p given as

$$p(\mathrm{P}_j, \mathrm{R}_i) \propto e^{-\frac{(X_j - X_i)^2 + (Y_j - Y_i)^2}{\sigma^2}}, \tag{10.1}$$

where σ is the point distribution variance. Hence, the range reading of the closest point (regarding pixel distance) will be selected.

10.2.2 Nearest Range Reading Considering Colour (NRC)

This method is an extension of the NR method using colour information in addition. Given a pixel P_j, the interpolated range reading r_j^* is assigned to the range value r_i of the projected laser range reading R_i which has the highest likelihood p given as

$$p(\mathrm{P}_j, \mathrm{R}_i) \propto e^{-\frac{(X_j - X_i)^2 + (Y_j - Y_i)^2}{\sigma_p^2} - \frac{||C_j - C_i||^2}{\sigma_c^2}}, \tag{10.2}$$

where σ_p and σ_c is the variance for the pixel point and the colour, respectively.

10.2.3 Multi-Linear Interpolation (MLI)

Given a set of projected laser range readings $\mathrm{R}_1 \dots \mathrm{R}_N$, a Voronoi diagram V is created by using their corresponding pixel co-ordinates $[X, Y]_{1 \dots N}$. The natural neighbours NN to an interpolated point R_j^* are the points in V, which

Fig. 10.2. From *left* to *right*: Depth image generated with the NR method. Depth image generated with the NRC method, small details are now visible, note that a depth image generated from a similar viewpoint as the laser range scanner makes it very difficult to see flaws of the interpolation algorithm. MLI method. LIC method

Voronoi cell would be affected if R_j^* is added to the Voronoi diagram, see Fig. 10.1. By inserting R_j^* we can obtain the areas $A_{1\ldots n}$ of the intersection between the Voronoi cell due to R_j^* and the Voronoi cell of R_i before inserting R_j^* and the area $A_{\mathrm{R}_j^*}$ as a normalisation factor. The weight of the natural neighbour R_i is calculated as

$$w_i(\mathrm{R}_j^*) = \frac{A_i}{A_{\mathrm{R}_j^*}}. \tag{10.3}$$

The interpolated range reading r_j^* is then calculated as

$$r_j^* = \sum_{i \in NN(\mathrm{R}_j^*)} w_i r_i. \tag{10.4}$$

This interpolation approach is linear [7]. One disadvantage is that nearest neighbourhood can only be calculated within the convex hull of the scan-points projected to the image. However, this is not considered as a problem since the convex hull encloses almost the whole image, see Fig. 10.2.

10.2.4 Multi-Linear Interpolation Considering Colour (LIC)

To fuse colour information with the MLI approach introduced in the previous subsection, the areas A_{R_i} and $A_{\mathrm{R}_j^*}$ are combined with colour weights $w_{1\ldots n}^c$ for each natural neighbour based on spatial distance in colour space.

Similar as in Sect. 10.2.2, a colour variance σ_c is used:

$$w_i^c(\mathrm{R}_j^*) = e^{-\frac{||C_i - C_j||^2}{\sigma_c^2}}. \tag{10.5}$$

The colour-based interpolated range reading estimation is then done with

$$r_j^* = \sum_{i \in NN(\mathrm{R}_j)} \frac{w_i w_i^c}{W^c} r_i \tag{10.6}$$

where $W^c = \sum_{i=1}^{n} w_i^c$ is used as a normalisation factor.

10.2.5 Parameter-Free Multi-Linear Interpolation Considering Colour (PLIC)

One major drawback of the methods presented so far and the approach presented in the related work section is that they depend on parameters such as σ_c, for example. To avoid the need to specify colour variances, the intersection area A_{R_i} defined in Sect. 10.2.3 is used to compute a colour variance estimate for each nearest neighbour point R_i as

$$\sigma_{c_i} = \frac{1}{n_i - 1} \sum_{j \in A_i} ||\mu_i - C_j||^2, \tag{10.7}$$

where $\mu_i = \frac{1}{n_i} \sum_{j \in A_i} C_j$ and n_i is the number of pixel points within the region A_i. σ_{c_i} is then used in (10.5).

This results in an adaptive adjustment of the weight of each point. In case of a large variance of the local surface texture, colour similarity will have less impact on the weight w_i.

10.3 Related Work

To our knowledge, the only work using vision for interpolation of 3D laser data is [8] where a Markov random field (MRF) framework is used.

The method works by iteratively minimising two constraints: ψ stating that the raw laser data and the surrounding estimated depths should be similar and ϕ stating that the depth estimates close to each other with a similar colour should also have similar depths.

$$\psi = \sum_{i \in N} k(r_i^* - r_i)^2, \tag{10.8}$$

where k is a constant and the sum runs over the set of N positions which contain a laser range reading r_i and r_i^* is the interpolated range reading for position i. The second constraint is given as

$$\phi = \sum_i \sum_{j \in NN(i)} e^{(-c||C_i - C_j||^2)} (r_i^* - r_j^*)^2, \tag{10.9}$$

where c is a constant, C is the pixel colour and $NN(i)$ are the neighbourhood pixels around position i. The function to be minimised is the sum $\psi + \phi$.

10.4 Evaluation

All datasets D were divided into two equally sized parts D_1 and D_2. One dataset, D_1, is used for interpolation and D_2 is used as the ground truth where each laser range measurement is projected to image co-ordinates. Hence for

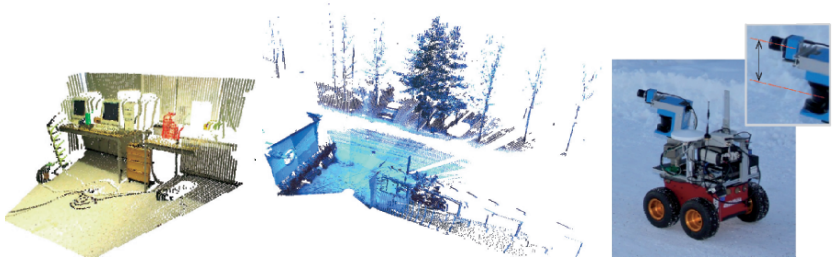

Fig. 10.3. *Left*: The third indoor evaluation scan, $Indoor_3$. *Middle*: Scans taken in winter time with some snow containing $Outdoor_1 - Outdoor_3$. *Right*: Our outdoor robot with the SICK LMS scanner and a colour CCD camera mounted on a pan tile unit from Amtec, that were used in the experiments. The close-up part shows the displacement between the camera and the laser which causes parallax errors

each ground truth point R_i we have the pixel positions $[X, Y]_i$ and the range r_i. The pixel position $[X, Y]_i$ is used as input to the interpolation algorithm and the range r_i is used as the ground truth. The performance of the interpolation algorithms is analysed based on the difference between the interpolated range r_i^* and the range r_i from the ground truth.

10.5 Experimental Setup

The scanner used is a 2D SICK LMS-200 mounted together with a 1 MegaPixel (1280×960) colour CCD camera on a pan-tilt unit from Amtec where the displacement between the optical axis is approx. 0.2 m. The scanner is located on our outdoor robot, see Fig. 10.3, a P3-AT from ActivMedia. The angular resolution of the laser scanner is 0.5°. Half of the readings were used as ground truth, so the resolution for the points used for interpolation is 1°.

10.6 Results

In all experiments the colour variance $\sigma_c = 0.05$ and the pixel distance variance $\sigma_d = 10\,\mathrm{mm}$ were used, which were found empirically. The parameters used within the MRF approach described in Sect. 10.3, where obtained by extensive empirical testing and were set to $k = 2$ and $C = 10$. The optimisation method used for this method was the conjugate gradient method described in [9] and the initial depths were estimated with the NR method. In all experiments the full resolution (1280×960) of the camera image was used.

All the interpolation algorithms described in this chapter were tested on real data consisting of three indoor and outdoor scans. The outdoor scans were taken in winter time with snow, which presents the additional challenge

Table 10.1. Results from $Indoor_1$, $Indoor_2$ and $Indoor_3$ datasets

	NR	NRC	MLI	LIC	PLIC	MRF
$Indoor_1$	0.065	0.054	0.052	0.048	0.049	0.048
$Indoor_2$	0.123	0.134	0.109	0.107	0.109	0.106
$Indoor_3$	0.088	0.072	0.067	0.060	0.060	0.067

Table 10.2. Results from $Outdoor_1$, $Outdoor_2$ and $Outdoor_3$ datasets

	NR	NRC	MLI	LIC	PLIC	MRF
$Outdoor_1$	0.067	0.068	0.056	0.059	0.054	0.054
$Outdoor_2$	0.219	0.294	0.235	0.322	0.275	0.218
$Outdoor_3$	0.526	0.584	0.522	0.574	0.500	0.498

that most of the points in the scene have very similar colours. The results are summarised in Tables 10.1 and 10.2, which show the mean error with respect to the ground truth.

For the indoor datasets, which comprise many planar structures, the lowest mean error was found with the multi-linear interpolation methods, particularly LIC and PLIC, and MRF interpolation. LIC and PLIC produced less (but larger) outliers.

With the outdoor data the results obtained were more diverse. For the dataset $Outdoor_1$, which contains some planar structures, a similar result as in the case of the indoor data was observed. For datasets with a very small portion of planar structures such as $Outdoor_2$ and $Outdoor_3$, the mean error was generally much higher and the MRF method performed slightly better compared to the multi-linear interpolation methods. This is likely due to the absence of planar surfaces and the strong similarity of the colours in the image recorded at winter time. It is noteworthy that in this case, the nearest neighbour interpolation method *without* considering colour (NR) performed as good as MRF. The interpolation accuracy of the parameter-free PLIC method was always better or comparable to the parameterised method LIC.

10.7 Conclusions

This chapter is concerned with methods to derive a high-resolution depth image from a low-resolution 3D range sensor and a colour image. We suggest five interpolation methods and compare them with an alternative method proposed by Diebel and Thrun [8]. In contrast to previous work, we present ground truth evaluation with simulated and real world data and analyse both indoor and outdoor data. The results of this evaluation do not allow to single out one particular interpolation method that provides a distinctly superior interpolation accuracy, indicating that the best interpolation method depends

on the content of the scene. Altogether, the MRF method proposed in [8] and the PLIC method proposed in this chapter provided the best interpolation performance. While providing basically the same level of interpolation accuracy as the MRF approach, the PLIC method has the advantage that it is a parameter-free and non-iterative method, i.e. that a certain processing time can be guaranteed. One advantage of the proposed methods is that depth estimates can be obtained without calculating a full depth image. For example if interpolation points are extracted in the image using a vision-based method (i.e. feature extraction), we can directly obtain a depth estimate for each feature.

References

1. M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3D scanning of large statues. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 131–144. ACM Press/ACM SIGGRAPH/ Addison Wesley Longman, 2000.
2. S. Thrun, D. Hähnel, D. Ferguson, M. Montemerlo, R. Triebel, W. Burgard, C. Baker, Z. Omohundro, S. Thayer, and W. Whittaker. A system for volumetric robotic mapping of abandoned mines. In *ICRA*, pages 4270–4275, 2003.
3. V. Sequeira, J. Goncalves, and M.I. Ribeiro. 3d reconstruction of indoor environments. In *Proc. of ICIP*, pages 405–408, Lausanne, Switzerland, 1996.
4. C. Früh and A. Zakhor. 3D model generation for cities using aerial photographs and ground level laser scans. In *CVPR*, pages 31–38, Hawaii, USA, 2001.
5. N. Haala and Y. Alshawabkeh. Application of photogrammetric techniques for heritage documentation. In *2nd Int. Conf. on Science & Technology in Archaeology & Conservation*, Amman, Jordan, 2003.
6. H. Andreasson, R. Triebel, and A. Lilienthal. Vision-based interpolation of 3d laser scans. In *Proc. of ICARA*, pages 469–474, 2006.
7. R. Sibson. *A brief description of natural neighbour interpolation*, pages 21–36. John Wiley & Sons, Chichester, 1981.
8. J. Diebel and S. Thrun. An application of markov random fields to range sensing. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 291–298. MIT Press, Cambridge, MA, 2006.
9. William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 2nd edition, 1992.