

Fusion of Combined Stereo and Spectral Series for Obtaining 3D Information

Ioana Gheța^a, Markus Mathias^a, Michael Heizmann^b and Jürgen Beyerer^{a,b}

^a Lehrstuhl für Interaktive Echtzeitsysteme (IES),
Institut für Technische Informatik (ITEC), Universität Karlsruhe (TH),
Adenauerring 4, D-76131 Karlsruhe, Germany;

^b Fraunhofer-Institut für Informations- und Datenverarbeitung IITB,
Fraunhoferstraße 1, D-76131 Karlsruhe, Germany

ABSTRACT

This contribution presents a fusion method for spectral series with the main purpose of obtaining 3D information. The image series to be fused are combined stereo and spectral series gained with a camera array. Therefore, in order to register them, features that are invariant with respect to the varying gray values in the spectral images are extracted. The proposed approach is region based and uses characteristics like size, position and shape for registration. The regions are identified using the watershed transformation. The fusion problem is modeled using energy functionals that are to be optimized. They take into consideration the size, position, shape and correlation of the regions. Using the implemented algorithm, several scenes have been reconstructed. The experimental results show that the proposed method delivers reliable and accurate dense depth maps of combined stereo and spectral series.

Keywords: Image fusion, spectral series, 3D information, 3D reconstruction, combined image series

1. INTRODUCTION

In many practical cases of automated visual inspection, certain scene properties can only be obtained by acquisition and fusion of multiple images. For example, authentic spectral information can be obtained by fusing spectral series, or panoramic images by fusion of image series acquired with different camera positions. Considering the low price of sensors and the increasing computational capacity for image processing, the idea of using camera arrays instead of a single camera for image acquisition in such cases is straightforward. To evaluate the potential of camera arrays, a system consisting of nine cameras of the same type, arranged in form of a 3×3 matrix has been built, Figure 1. The cameras have been provided with interference filters that uniformly sample the visible and near infrared (NIR) spectrum (400 – 850 nm). The geometrical arrangement of the filters in the matrix has no basic scientific importance. However, the probability that the properties of the objects (such as texture, shape) are captured through two filters of neighboring wavelength is high. Therefore, the geometrical configuration of the filters is such that filters with neighboring transmission range are placed side by side, see Figure 6.



Figure 1. The camera array in combination with an industrial robot.

The acquired image series are combined stereo and spectral series; the stereo property is given by the different camera positions in the array. The scope of the project is a 3D reconstruction of the scene with additional spectral information for each point. The present contribution describes the first step: the 3D reconstruction.

Most algorithms for estimating dense depth maps are based on evaluating similar structures in the images, such as gray values or texture. The challenge of the problem in this case consists in the fact that such structures cannot be found reliably in the gained image series. The spectral filters cause objects to

Further author information: (Send correspondence to Ioana Gheța)
E-mail: ioana.gheta@ies.uni-karlsruhe.de, Telephone: +49 721 6085914

Multisensor, Multisource Information Fusion: Architectures, Algorithms,
and Applications 2008, edited by Belur V. Dasarathy, Proc. of SPIE Vol. 6974,
697406, (2008) · 0277-786X/08/\$18 · doi: 10.1117/12.778357

appear differently in each image. They take different gray values or the object structure might disappear, especially when printed texture is present. Therefore, for identifying correspondences, features invariant with respect to texture and gray values are needed. Such features may be based on homogeneous regions, i. e. a region that has a uniform gray value.

Further on, a region based approach to fuse combined stereo and spectral series is presented. The notations and the preprocessing steps are introduced in Section 2. For the image segmentation, the watershed transformation has been applied under consideration of some additional constraints, see Section 3.1. Correspondences in the segmented images are searched based on the size, shape and position of the regions, see Section 3.2. The problem is modeled with the help of energy functionals. The registered image series are fused successively in pairs, such that all images contribute to the final result. The fusion takes place on two levels: on the first one, images are fused to obtain disparity maps and on the second one, preliminary depth maps are fused to a final result, see Section 3.4. The result is a depth map that combines the information brought by all images in the series. Finally, an example is presented to illustrate the steps of the algorithm from the acquisition to the reconstruction of the scene, see Section 4.

1.1 Previous Work

Fusion of spectral series is present in the literature mostly in the domain of satellite image processing with the purpose of identifying certain regions, e. g. green areas or minerals.¹ The 3D reconstruction takes place under different constraints and therefore the algorithms cannot be applied in the case of close scenes, such as the present case.

Region based image processing is used mainly in three domains: medicine, surveillance and industrial optical inspection for 3D reconstruction of scenes with lack of structure. The purpose in medicine is to fuse images acquired with different sensors for a detailed identification of tissues. For a 3D reconstruction, measurement techniques delivering 3D data like CT (computer tomography) are employed.² In the surveillance domain, the purpose is to find and follow the same objects through an image sequence.³ In both cases (medicine and surveillance), one region of the first image is searched in the other images of the series.

In the present case, which belongs to the third domain, a complete registration of the images in the series is required. For this scope (registration of spectral series), only a few algorithms are present in the literature. They can be found under two keywords: region based and segment based registration, both used equivalently. Modalities for finding the regions are multiple and the basic algorithms are found in the standard literature.⁴

For the registration of stereo pairs, two techniques are encountered. The first one is to segment the first image into regions. In each of the regions, features (texture, edges) are defined, which are then searched in the entire second image for identifying the corresponding region.⁵⁻⁷ The obtained correspondences are either labeled with their disparity⁶ or a modeled 3D plane is fitted to them.^{5,7}

The second technique is to segment both images and then to find correspondences between the obtained regions. In this case, there are two types of algorithms. The first type uses only one segmentation, e. g. based on color as an initial guess for estimating depth, which can then be optimized by means of an energy functional between the first warped image and the second.⁸ The second type fuses segmentations obtained using different features, e. g. by means of color and pixel based disparity estimation,⁹ or by employing the same feature at different granularity levels.¹⁰ Neighborhood and father-child relations can be incorporated by means of dynamic trees.¹¹

It is to mention that the algorithms in the literature use regions as elements to be registered in order to obtain improved results for scenes without structure or to identify certain objects or patterns in different images. In the present contribution, the spectral component of the image series takes care that regions are homogeneous elements that define parts of objects. Therefore, the identified regions can be used for a complete registration of the images. This contribution proposes an innovative approach for fusing stereo and spectral image series to obtain depth information.

2. PREPROCESSING

Prior to the actual fusion of the combined image series, the calibration of the camera system is mandatory for estimating depth. In addition, a preprocessing step consisting of rectifying pairs of images in the series is helpful for the image registration and the subsequent fusion.

In general, the images are considered as functions: $B_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, with the gray value $B_i(u, v)$ of the pixel $\mathbf{p} = (u, v)^T$ in the i -th image of the series; u and v are the image coordinates. The image function is also used in a general form with a region R as a variable, e. g. $B(R)$ is the image signal for an entire region. \mathcal{P}_i is the set of pixels in the i -th image. Generally, the index indicates which image of the series is referred (e. g. \mathbf{p}_i is a pixel in the i -th image).

2.1 Calibration

First, a weak calibration (i. e. estimating the fundamental matrices) is required for the rectification of the images in the series and second, the projection matrices are required for estimating depth. The estimation of both matrices is done by using several recordings of a planar chessboard employed as calibration object. The fundamental matrix for an image pair can be determined by using the eight-point-algorithm,^{12,13} but also directly by using the projection matrices:¹³ $\mathbf{F}_{ji} = [\mathbf{e}_j]_{\times} \mathbf{P}_j \mathbf{P}_i^+$ with the epipole \mathbf{e} and the projection matrix \mathbf{P} . $[\mathbf{e}]_{\times}$ is the skew-symmetric matrix of \mathbf{e} (see e. g. [13]) and \mathbf{P}^+ is the pseudo-inverse of \mathbf{P} .

The projection matrices are obtained using a standard calibration algorithm, which has been employed for all camera pairs.¹⁴ Once the camera projection matrices are known, it is straightforward to reconstruct depth from point correspondences. This triangulation problem is equivalent to solving a system of linear equations.¹³

2.2 Image Rectification

As a first preprocessing step, the images are rectified by means of epipolar geometry. For any given pixel $\mathbf{p}_i \in \mathcal{P}_i$, its corresponding pixel $\mathbf{p}_j \in \mathcal{P}_j$ must lie on a certain epipolar line within the image B_j , see Figure 2. This can be described mathematically by the fundamental matrix \mathbf{F}_{ji} : for a pair of corresponding pixels $\mathbf{p}_i \leftrightarrow \mathbf{p}_j$, the following equation in homogeneous coordinates is satisfied: $\mathbf{p}_j^T \mathbf{F}_{ji} \mathbf{p}_i = 0$. The epipolar line corresponding to \mathbf{p}_i in the j -th image is given by $\mathbf{F}_{ji} \mathbf{p}_i$. The fundamental matrices are estimated as described in Section 2.1.

For describing corresponding pixels, the standard disparity notion for stereo image pairs is used. This defines disparity as a translation of the pixel position. If two pixels correspond ($\mathbf{p}_i \leftrightarrow \mathbf{p}_j$), then for two horizontally rectified images (see Figure 3(a)), the relation between them can be formulated as $\mathbf{p}_i = (u_i, v_i)^T \leftrightarrow (u_i + \alpha, v_i)^T = (u_j, v_j)^T = \mathbf{p}_j$. For two vertically rectified images (see Figure 3(b)), the relation is $\mathbf{p}_i = (u_i, v_i)^T \leftrightarrow (u_i, v_i + \gamma)^T = (u_j, v_j)^T = \mathbf{p}_j$. α and γ are the translations in horizontal and vertical direction, respectively. The notion of disparity and with it the rectification in both directions are required due to the matrix form of the camera array, since the images from cameras situated side by side and from those situated on top of each other must be fused.

Image rectification in the present case means projecting the images on a common plane parallel to the line between the optical centers of the cameras, also called the plane at infinity. For that, two homographies are computed with the help of the fundamental matrices between the two images. The resulted rectified images contain a minimum of distortion.^{12,13}

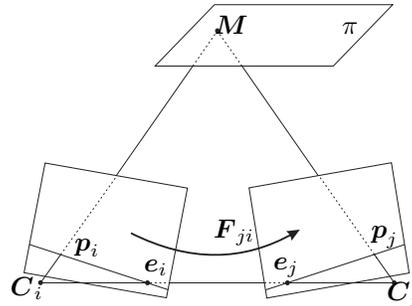


Figure 2. Image formation: \mathbf{p}_i and \mathbf{p}_j are projections of the same 3D point M and are therefore corresponding pixels. Mathematically, their relation can be described with the help of the fundamental matrix \mathbf{F}_{ji} . \mathbf{C} represents the camera optical center and \mathbf{e} is the epipole.

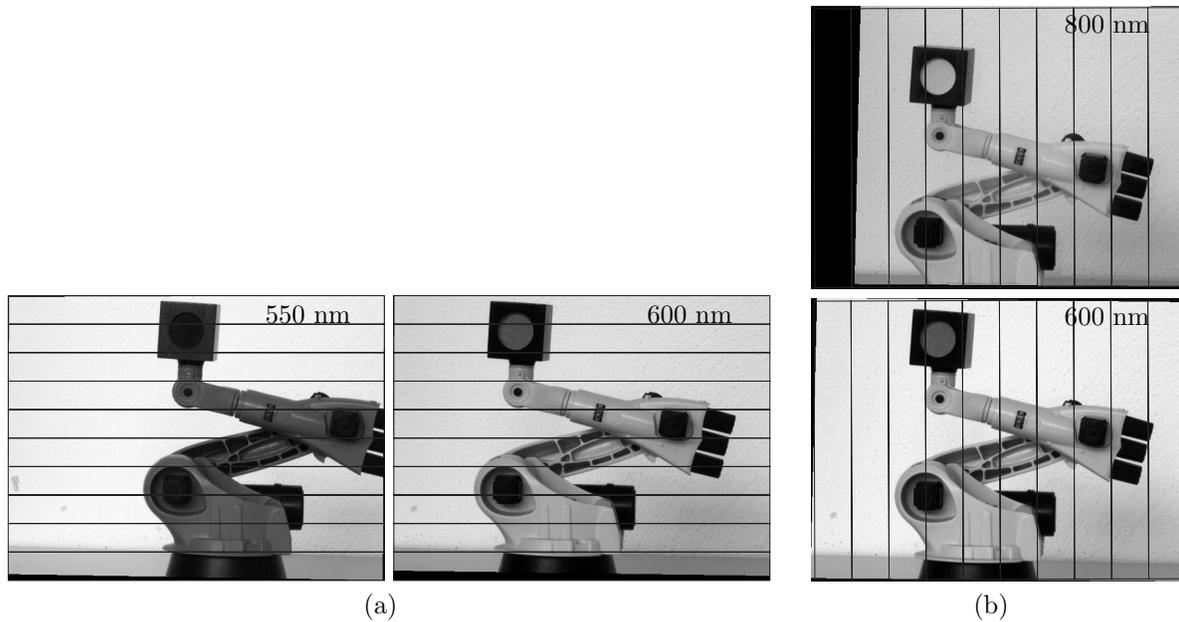


Figure 3. Horizontally (a) and vertically (b) rectified images.

3. IMAGE FUSION AND 3D RECONSTRUCTION

The fusion of the combined image series to estimate depth uses a region based approach. It consists of three steps: first, homogeneous regions are defined. Second, corresponding regions in different images are identified, and third, disparities of the regions are evaluated in order to estimate a depth map.

3.1 Image Segmentation

A good segmentation satisfies three requirements:⁴ first, the segmentation should be disjoint and cover the entire image, i. e. it should be a partition of the image; second, within each region, a given homogeneity criterion should be satisfied, and third, the homogeneity criterion should not be satisfied in between the regions.

For image segmentation in the present case, the watershed transformation has been chosen. It segments the images according to the requirements mentioned above: the segmentation is implicitly disjoint and complete, the homogeneity criterion, which requires a uniform gray value, is only satisfied within the regions. As input, the watershed transformation requires gradient images. Any gradient operator which weights the edges and lines according to the steepness of their gradient, e. g. the Sobel operator,⁴ can be used.

In order to avoid oversegmentation, the images have been smoothed using an anisotropic filter, which preserves edges in the images without loss of sharpness.¹⁵ The parameters of the watershed transformation are adapted such that the number and the size of the identified regions in each image generally correspond. Two problems can intervene: the regions may be so small that their large number leads locally to false correspondences (e.g. caused by noise), and a structure identified as a region in one image may appear segmented in a large number of regions in another image. Such cases make the estimation of correspondences hard and should therefore be avoided. For that purpose, a postprocessing step merges too small regions.

An example of segmented images of the series in Figure 6 is presented in Figure 7. The gray values are used as labels and have no significance regarding the correspondences between regions in different images.

3.2 Image Registration

The registration of two images B_r (reference image) and B_t (template image) can be defined as the search for a geometrical transformation $\varphi_{rt} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, such that the dissimilarity between the reference image

and the transformed template image is as small as possible. The dissimilarity is measured by a distance function $d: d(B_r, B_t \circ \varphi) \rightarrow \min$.

Image registration can be formulated as the optimization of an energy functional:

$$E(\varphi_{rt}) = D(\varphi_{rt}) + \beta S(\varphi_{rt}) = d(B_r, B_t \circ \varphi_{rt}) + \beta S(\varphi_{rt}) \rightarrow \min. \quad (1)$$

The purpose is to find a function φ_{rt} such that the energy functional is minimized. The first term in Equation (1) is the data term, which ensures the consistency of the result with the measured data, i. e. the distance function of corresponding features should be small. The rest of the terms, here cumulated in S , define additional constraints, e. g. smoothness; β is a weight. The reader is referred to [16] for a theory of image registration techniques. For practical examples of fusing image series acquired with a camera array by means of energy functionals, see [17–20].

For the region based registration, Equation (1) is adapted such that the features compared are regions:

$$E(\varphi) = D(\varphi) = \sum_{(B_r, B_t)} \sum_R d_R(B_r(R), (B_t \circ \varphi_{rt})(R)) \rightarrow \min. \quad (2)$$

The function $d_R(., .)$ measures the correspondence between regions in two images. In order to assess all image pairs, φ is the set of all geometrical transformations between the chosen image pairs (B_r, B_t) : $\varphi = \{\varphi_{12}, \varphi_{23}, \dots\}$. By summing over all pairs (B_r, B_t) , the entire image series is regarded. In this step, only neighboring image pairs are used, see Section 3.4. Since the compared regions are homogeneous (see Section 3.1), the smoothness term $S(\varphi)$ can be omitted without loss of information.

The similarity of the regions is compared using a multidimensional vector $\mathbf{m}(\cdot)$ of invariant features like size, position and shape:

$$E(\varphi) = D(\varphi) = \sum_{(B_r, B_t)} \sum_R d_m(\mathbf{m}(B_r(R)), \mathbf{m}((B_t \circ \varphi_{rt})(R))) \rightarrow \min. \quad (3)$$

In order to evaluate all invariant features within $\mathbf{m}(\cdot)$, which cannot be compared by using the same metric, the remaining data term $D(\varphi)$ in the energy functional $E(\varphi)$ is split into three components with the respective weights $\beta \geq 0$:

$$E(\varphi) = D(\varphi) = \beta_s D_s(\varphi) + \beta_p D_p(\varphi) + \beta_c D_c(\varphi) \rightarrow \min. \quad (4)$$

The first component $D_s(\varphi)$ takes care that the two regions have similar sizes $A(\cdot)$:

$$D_s(\varphi) = \sum_{(B_r, B_t)} \sum_{(R_r, R_t)} d_s(R_r, \varphi_{rt}(R_t)) = \sum_{(B_r, B_t)} \sum_{(R_r, R_t)} |A(R_r) - A(\varphi_{rt}(R_t))|. \quad (5)$$

The second term $D_p(\varphi)$ evaluates the size $o(R_r, R_t)$ of the overlapping area of regions in comparison to their position on the epipolar lines:

$$D_p(\varphi) = \sum_{(B_r, B_t)} \sum_{(R_r, R_t)} d_p(R_r, \varphi_{rt}(R_t)), \quad d_p(R_r, \varphi_{rt}(R_t)) \propto \frac{f(R_r, \varphi_{rt}(R_t))}{o(R_r, \varphi_{rt}(R_t))}. \quad (6)$$

Here, the function $f(., .)$ measures the geometrical distance between two regions with respect to the epipolar lines.

The third term $D_c(\varphi)$ evaluates the correlation between two regions, which implies the similarity of their shape:

$$D_c(\varphi) = \sum_{(B_r, B_t)} \sum_{(R_r, R_t)} d_c(R_r, \varphi_{rt}(R_t)) = - \sum_{(B_r, B_t)} \sum_{(R_r, R_t)} \max_{\tau} \{B_r(R_r) \cdot (B_t \circ \varphi_{rt})(R_t - \tau)\}. \quad (7)$$

Finding a minimal solution for the energy functional defined in Equation (3) means obtaining an optimal estimation for the set of geometrical transformations φ .

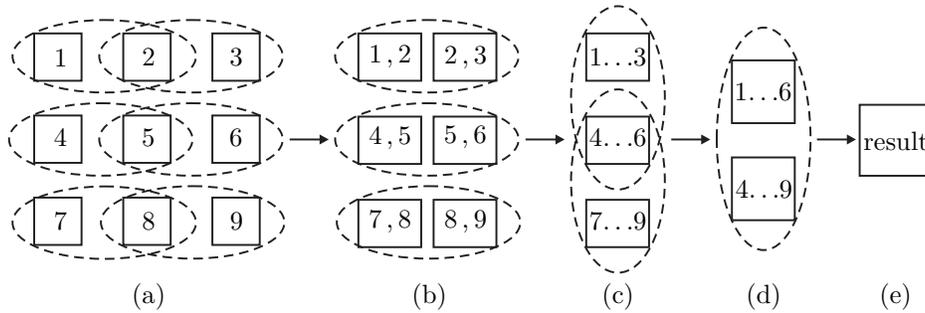


Figure 4. Fusion sequence of the image series: (a) Choice of the image pairs for the first fusion level. (b) First fusion level: computation of disparity and depth maps by fusing horizontally neighboring images. (c,d) Second fusion level: computation of depth maps by fusing vertically neighboring depth maps obtained from the first fusion level (b). (e) Final result of the fusion process.

3.3 Implementation

For finding an optimal solution for the energy functional in Equation (3), a sequential region matching approach is proposed. The first step is to build a matrix of region pairs that are highly probable to match. This is done by comparing the size, the position and the correlation of the regions:

$$E(R_r, R_t) = \min_{\varphi_{rt}} \{ \beta_s d_s(R_r, \varphi_{rt}(R_t)) + \beta_p d_p(R_r, \varphi_{rt}(R_t)) + \beta_c d_c(R_r, \varphi_{rt}(R_t)) \}. \quad (8)$$

The result is recorded in the matrix for each candidate pair.

The second step is to find the optimal global match of two images. This implies that each region in the first image should have exactly one corresponding region in the second image and vice versa, i. e. exactly one entry in each row and in each column should be considered for the global solution. This is done by selecting the entries leading to the minimum sum of energy components $E(R_r, R_t)$ under the constraint given above. To ensure an efficient solution, the implemented algorithm considers for each region in the first image only the best two matches in the second image.

3.4 Depth Estimation

The image fusion takes place on two abstraction levels both belonging to the level of features.²¹ The chosen fusion algorithm ensure that the information contained in the images is exhaustively processed and comprised in the result.



Figure 5. Experimental scene: robot holding a green log of wood (acquired with an RGB camera).

At first level, disparity maps are computed. This is done in each row of the camera array for each two neighboring images, Figure 4(a). Since the disparities for image pairs are not directly comparable (unlike the generalized disparity notion in [17–20]), the immediate fusion of two disparity maps on the abstraction level of disparities is not possible. In order to obtain a uniform measure, which enables the further fusion process, depth maps are computed from the disparity maps.

The fusion of depth maps takes place at a second level in two different steps: first, horizontally neighboring depth maps are fused (Figure 4(b)) and then, vertically neighboring ones (Figure 4(c,d)). The middle image of the series (camera 5) has been chosen as final perspective due to the fact that it highly probably sees most of the scene (as a consequence of its location in the array). The order of the image pairs to be fused has been chosen such that a minimum of fusion operations have to be made and

nevertheless, the information gathered is completely comprised in the final result.

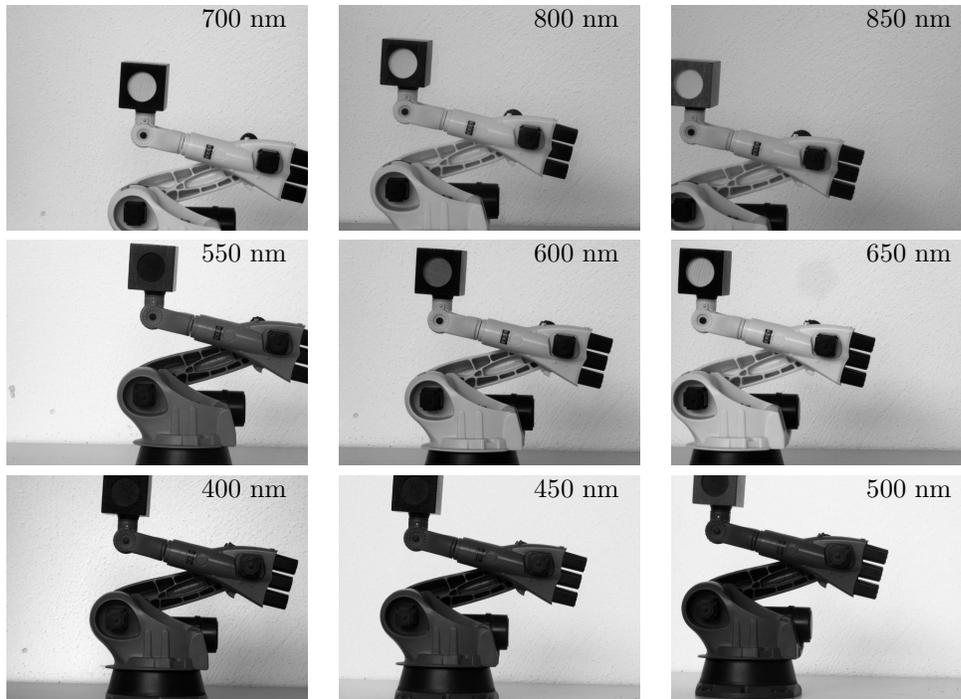


Figure 6. Image series acquired with the camera array. In the upper right corner of each image, the peak wavelength of the spectral filter is tagged.

4. RESULTS

The developed algorithm has been tested on different scenes. An exemplary scene contains a model robot holding a green log of wood, Figure 5. From this scene, an image series has been acquired, Figure 6. The region segmentation result is presented in Figure 7. The result of the first fusion step is depicted in Figure 8. Figure 9 presents the estimated depth map (left) and the reconstruction result (right).

The image series has been acquired under directional day light illumination. Therefore, certain structures (e.g. on the robot, in the middle of the image) are amplified or reduced in each image depending on the perspective of the camera and the shadows thrown. Since the watershed transformation assesses the steepness of the gradient, the contrast influences the segmentation result. In such cases, it can happen that parts of the objects in the scene are interpreted as belonging to the background (the green log of wood in Figure 7 third row, third column). Another example is the structure of the robot that is interpreted as one single region in certain images (Figure 6 first row, third column) and as many regions in others (Figure 6 third row, third column).

In consequence, not all regions are identified with the same shape and size in each image. Therefore, certain regions cannot be identified as corresponding, which leads to incomplete depth maps with missing details after the first fusion step (white holes in the depth maps provoked by the lack of depth values for that regions), see the intermediate results in Figure 8.

Nevertheless, the fusion of all depth maps results in a dense and correct depth map, see Figure 9. The pieces of information dispersed through the intermediate depth maps are gathered in the final result.

5. CONCLUSIONS

The present contribution proposes a method for fusing combined stereo and spectral series in order to estimate depth. The main problem is that the objects of the scene are imaged with different gray values due to the application of spectral filters. Therefore, regions have been chosen as invariant element for image registration. Based on their characteristics (size, position and shape), a registration of image pairs of the series is accomplished. First, the registered images are fused in pairs to obtain disparity maps,

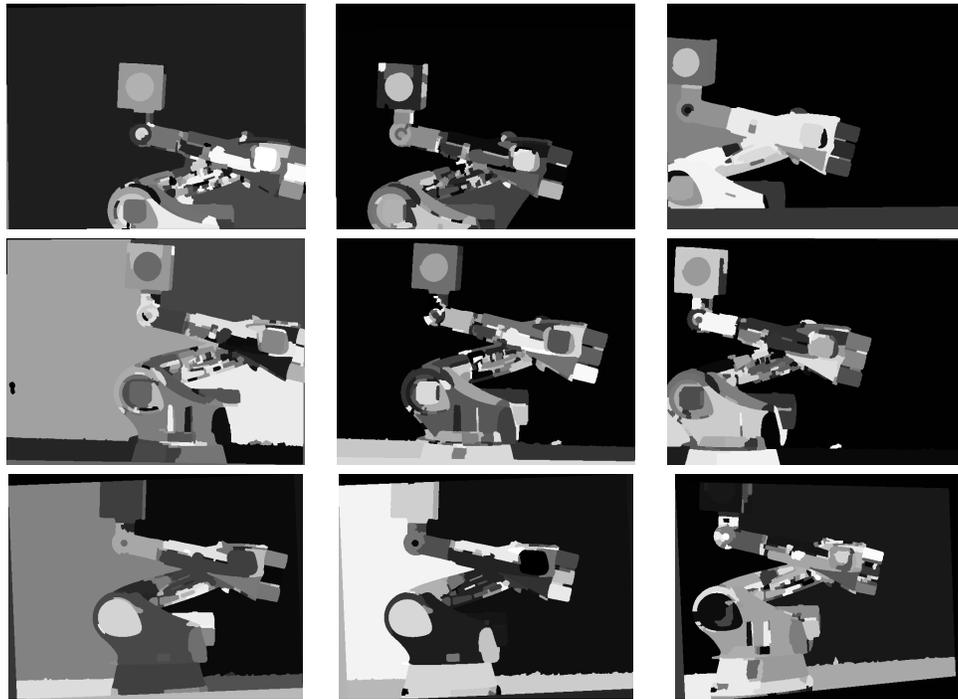


Figure 7. Identified regions in each image of the series in Figure 6. The gray values are used as labels for distinguishing neighboring regions in an image and have no significance regarding a possible correspondence.

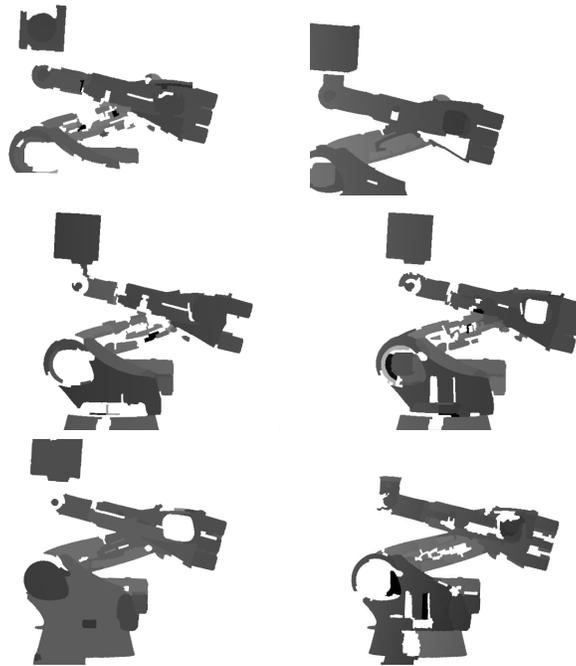


Figure 8. Depth maps corresponding to intermediate fusion results according to the scheme in Figure 4(b). Dark gray indicates regions near to the camera, while light gray indicates regions farther away.

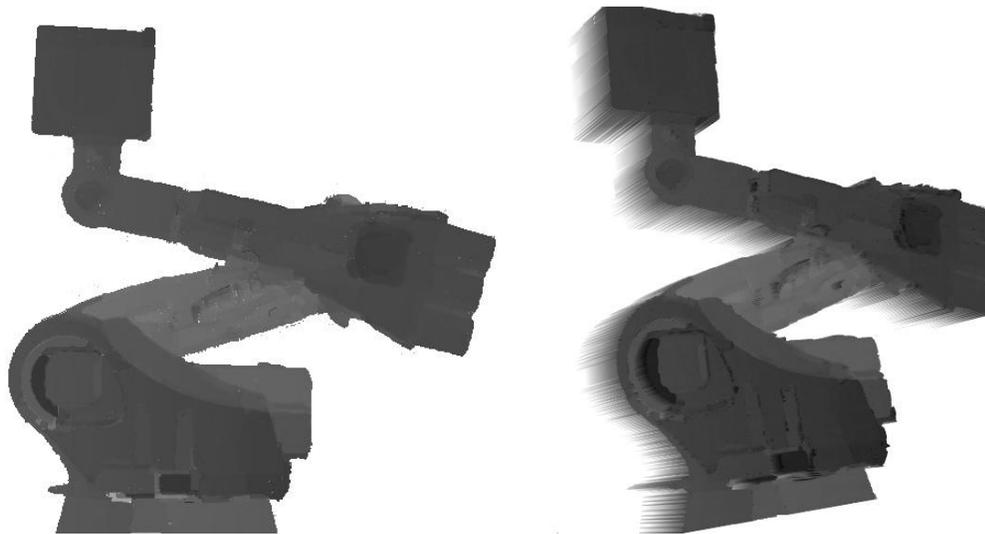


Figure 9. Left: Depth map according to the final fusion result. Right: 2 1/2 D reconstruction of the scene in Figure 5. Dark gray indicates regions near to the camera, while light gray indicates regions further away.

which are then used to compute depth maps. Due to the spectral properties of the scene and thrown shadows in the scene, regions may be identified differently in the images, causing false correspondences. This effect can be observed in the intermediate results, which contain gaps. Nevertheless, by considering all images of the series, a complete dense depth map is obtained.

Future work will take the project a step further and fuse more depth maps obtained from different viewing angles. A second step is to add the fused spectral information to the reconstruction. The main advantage of such a system is the recovery of both depth and detailed spectral information by just one acquisition of a combined image series. A few examples of practical applications of the final product are material classification or color visual inspection.

REFERENCES

1. J. S. Tyo, A. Konsolakis, D. I. Diersen, and R. C. Olsen, "Principal-components-based display strategy for spectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **41**(3), pp. 708–718, 2003.
2. J. Beutel and M. Sonka, eds., *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, SPIE Press Monograph Vol. PM80, 2000.
3. W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *SMC-C* **34**, pp. 334–352, August 2004.
4. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, Inc., 2002.
5. L. Hong and G. Chen, "Segment-based stereo matching using graph cuts," *cvpr* **01**, pp. 74–81, 2004.
6. Y. Wei and L. Quan, "Region-based progressive stereo matching," *cvpr* **01**, pp. 106–113, 2004.
7. M. Bleyer and M. Gelautz, "Graph-based surface reconstruction from stereo pairs using image segmentation," in *Proceedings of SPIE*, **5665**, pp. 288–299, 2005.
8. H. Tao, H. S. Sawhney, and R. Kumar, "A global matching framework for stereo computation," in *ICCV*, pp. 532–539, 2001.
9. X. Zhou and R. Wang, "Stereo matching based on color and disparity segmentation by belief propagation," *Optical Engineering* **46**, 2007.
10. L. Cohen, L. Vinet, P. T. Sander, and A. Gagalowicz, "Hierarchical regions based stereo matching," in *Proceeding of Computer Vision and Pattern Recognition*, pp. 416–421, 1989.

11. S. Todorovic and M. C. Nechyba, "Dynamic trees for unsupervised segmentation and matching of image regions," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, pp. 1762–1777, 2005.
12. O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, 2001.
13. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2 ed., 2003.
14. J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **14**(10), pp. 965–980, 1992.
15. B. Jähne, *Digitale Bildverarbeitung*, Springer Verlag, 4 ed., 1997.
16. J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, 1 ed., 2004.
17. C. Frese and I. Gheța, "Robust depth estimation by fusion of stereo and focus series gained with a camera array," in *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006.
18. I. Gheța, C. Frese, and M. Heizmann, "Fusion of combined stereo and focus series for depth estimation," *Informatik 2006 – Informatik für Menschen* **1**, pp. 359–363, 2006.
19. I. Gheța, C. Frese, W. Krüger, G. Saur, N. Heinze, M. Heizmann, and J. Beyerer, "Depth estimation from flight image series using multi-view along-track stereo," *Optical 3D Measurement Techniques, Zürich*, pp. 119–125, 2007.
20. I. Gheța, C. Frese, M. Heizmann, and J. Beyerer, "A new approach for estimating depth by fusing stereo and defocus information," *INFORMATIK 2007, Informatik trifft Logistik, 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Bremen*, 2007.
21. M. Heizmann and F. Puente León, "Fusion von bildsignalen," *Technisches Messen* **74**(3), pp. 130–138, 2007.