5

3D Fusion of Stereo and Spectral Series Acquired With Camera Arrays

Ioana Gheţa[1], Michael Heizmann[2] and Jürgen Beyerer[3]
[1,3]*Karlsruhe Institute of Technology KIT*
[2,3]*Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB*
Germany

1. Introduction

One of the main requirements of industrial visual inspection is that the information acquisition is accomplished in real-time. Camera arrays are a promising solution since they offer the possibility of simultaneous image acquisition. Moreover, the acquisition parameters of the different cameras can be varied. Due to dropping prices of industrial cameras, a large number of cameras can be employed in a camera array for automated visual inspection.
The advantages offered by camera arrays come with a price. Simultaneously triggering the cameras results in obtaining image series that contain a stereo effect. If more acquisition parameters (e. g., focus, different spectral filters) are varied, the obtained image series are combined image series, i. e., the images differ in more than one effect. For example, if the cameras are equipped with spectral filters, the obtained image series are combined stereo and spectral series, i. e., the images differ due to both the stereo effect and the acquisition in different parts of the spectrum. However, the main advantage of such image series is that they contain different types of information gained simultaneously: in this case, it is spatial information due to the stereo effect and spectral information due to the use of spectral filters. The challenge consists in fusing the image series, since the different types of information in the combined image series cannot be evaluated separately.
The present chapter deals with different methods of fusing combined stereo and spectral images in order to obtain both spatial and spectral information. For obtaining the spatial information, region based image registration methods for the exploitation of the stereo effect are presented. The problem is modeled with energy functionals, which are minimized by state-of-the-art methods, e. g., dynamic programming or graph cuts. With the help of the obtained spatial information (in form of depth maps), the spectral information can be extracted and further employed, e. g., for material classification or an improved object detection.

1.1 State of the art

For a general theory of image registration techniques, the reader is referred to Modersitzki (2004). Practical examples for fusing image series acquired with camera arrays by means of energy functionals are given in Frese & Gheţa (2006); Gheţa et al. (2006); Gheţa, Frese, Heizmann & Beyerer (2007); Gheţa, Frese, Krüger, Saur, Heinze, Heizmann & Beyerer (2007).

The topic of fusing combined stereo and spectral series for obtaining depth information has not yet been approached in the literature. However, the registration of similar image series can be found in the domains of remote sensing (Lillesand et al. (2008)) and medical image processing (Bankman (2000)). The registration for satellite images usually means finding similar image structures in order to build image carpets. The registration of images for medical purposes generally means finding an elastic transformation between possibly multimodal images in order to determine how tissue or organs change in time. These registration methods for satellite images or for medical purposes are not applicable for industrial imaging, since they do not provide registration for each pixel.
One paper though deals with the problem of registering two stereo images having different gray values for corresponding pixels (Fookes et al. (2004)), without an explicit practical application. The images are not real spectral images, but have been obtained by uniformly modifying one image, e. g., through negation. The authors use entropy based registration. Entropy based methods are not successful for scene areas with little structure, and can therefore not be employed for industrial vision like in our case.
For the analysis of spectral series, various approaches are presented in the literature (Chang (2003)). An example for the fusion of spectral images with the purpose of false color representation is given in Tyo et al. (2003).

1.2 Camera array

For the acquisition of the images, a camera array with nine cameras is employed; see Fig. 1. The cameras are of the same type and have the same geometrical properties. For acquiring the spectral properties of the scene, the cameras are equipped with 50 nm wide spectral filters covering the visual and the near-infrared spectra (400-850 nm). The positions of the filters in the array are unimportant for the image evaluation, since the proposed approaches process all images in the same way; see Section 2. It is though to be mentioned that images of neighboring spectral intervals are more likely to have similar features than images taken at very different spectral wavelengths; see, e. g., Fig. 2. Previous to acquisition, the camera array is calibrated using an improved calibration method, which is based on Weng et al. (1992).

Fig. 1. Camera array used for the acquisition of combined image series. The cameras are equipped with spectral filters.

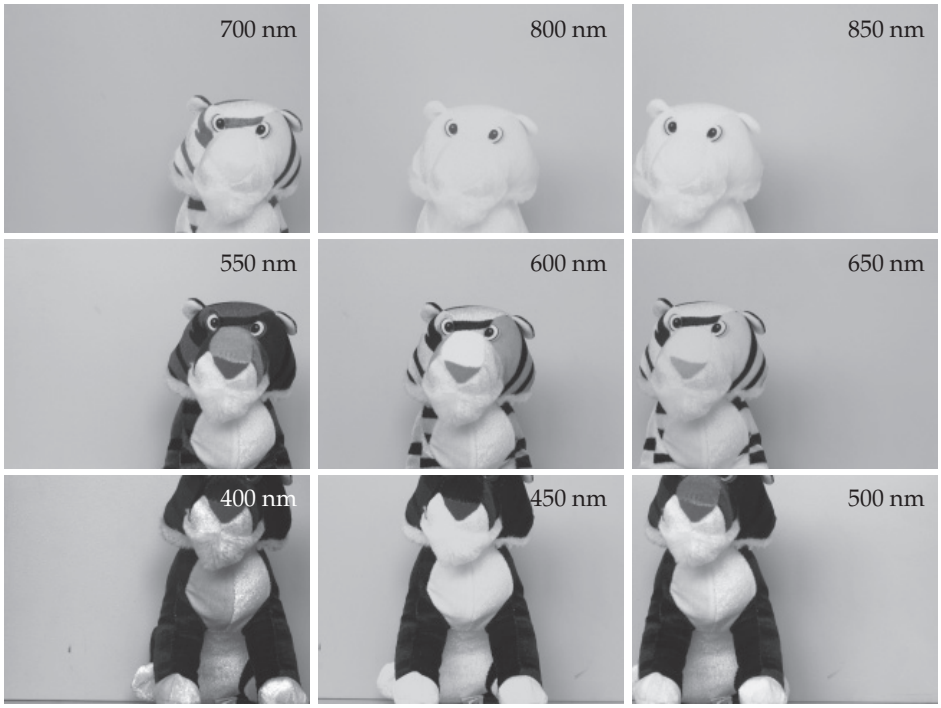


Fig. 2. Combined stereo and spectral series acquired with a camera array. The center wavelength of the filters used for acquisition is written in the upper right corner of each image. The observed scene is displayed in Fig. 3.

Fig. 3. Exemplary scene: an orange tiger with black and white stripes.

1.3 Notations and concepts

The acquired data, i. e., the image series, is denoted by $B := \{B_i,\ i = 1, \ldots, n\}$, where n is the number of images in the series. The images are interpreted as functions $B_i : \mathbb{R}^2 \to \mathbb{R}$, with

the gray value $B_i(\boldsymbol{u}_i)$ of the pixel $\boldsymbol{u}_i = (u_i, v_i)^{\mathrm{T}}$ in the i-th image of the series; u and v are the image coordinates.
The registration process requires in this case rectified and segmented images (Faugeras & Luong (2004); Gonzalez & Woods (2008); Hartley & Zisserman (2003)). For image segmentation, the watershed transform is used under consideration of some additional constraints: the parameters of the operation are chosen such that the images have similar numbers of regions and the sizes of the regions lie within a given range, thus too few segments and oversegmentation are avoided. The watershed transform yields that the images are completely segmented and the obtained regions are compact and disjoint (Gonzalez & Woods (2008)). Therefore, the regions build a partition of each image. The result of the image segmentation can be described using a function $R(\cdot)$, which assigns each pixel $\boldsymbol{u}$ to a region $\mathcal{R}^{\varpi}$ (described by the respective label ϖ):

$$R : \mathbb{R}^2 \to \mathcal{Q}\,, \quad R(\boldsymbol{u}) = \varpi\,, \tag{1}$$

where $\mathcal{Q}$ is the set of labels for the regions in an image. A *region* can thus be defined as a set of points having the same label:

$$\mathcal{R}^{\varpi} := \{\boldsymbol{u} | R(\boldsymbol{u}) = \varpi\} \tag{2}$$

with $\varpi \in \mathcal{Q}$ and $|\mathcal{Q}|$ the number of regions of a given image.
A region $\mathcal{R}$ consists of its interior and its boundary:

$$\mathcal{R} := \mathcal{K}_{\mathcal{R}} \cup \mathcal{R}^{\circ}\,, \tag{3}$$

where the boundary $\mathcal{K}_{\mathcal{R}}$ is defined as the set of pixels located on the border. The boundary $\mathcal{K}_{\mathcal{R}}$ is thus a subset of the region $\mathcal{R}$ and the intersection of the boundaries of two different regions is the empty set. The boundaries of the regions in an image can be determined using standard edge detection operators (Gonzalez & Woods (2008)). In this case morphological operators are used. Like pixels, regions belonging to the image i are marked with $\mathcal{R}_i$.
For the registration we also employ (image) *areas* $\mathcal{B}$. They are defined as a subset of the partition of the image and consist of neighboring regions:

$$\mathcal{B} := \bigcup_{r \in \mathcal{J}} \mathcal{R}^r = \bigcup_{r \in \mathcal{J}} \mathcal{K}_{\mathcal{R}^r} \cup \bigcup_{r \in \mathcal{J}} \mathcal{R}^{r\circ}\,. \tag{4}$$

$\mathcal{J}$ is the set of all indices of those neighboring regions $\mathcal{R}^r$ (of the same image) belonging to the area $\mathcal{B}$. The set of all areas of an image is equivalent to the power set of the set of all regions. The main difference between regions and areas is that regions build a partition of the image, while areas do not, since areas may overlap.
A common concept used for marking corresponding pixels when registering images is *disparity*. The disparity is the distance in coordinates of corresponding pixels in stereo images (Faugeras & Luong (2004); Hartley & Zisserman (2003)). Considering that the image series acquired with the camera array contain more than two images to be registered, a more generic concept than disparity must be used, which allows the labeling of correspondences between regions and areas in all images of the series (Boykov & Kolmogorov (2004)). To this purpose, a *labeling function* $s(\cdot)$ is defined, which assigns the same label α to all pixels of corresponding regions and areas:

$$s(\boldsymbol{u}_i) : \mathbb{R}^2 \to \mathcal{L}\,, \quad s(\boldsymbol{u}_i) = \alpha \;\forall\, \boldsymbol{u}_i \in \mathcal{R}_i\,, \quad \text{if}$$
$$\mathcal{R}_i \leftrightarrow \mathcal{R}_1 \wedge \mathcal{R}_1 \leftrightarrow \mathcal{R}_2 \left(\Leftrightarrow (\bar{u}_2, \bar{v}_2)^{\mathrm{T}} = (\bar{u}_1 + \alpha, \bar{v}_1)^{\mathrm{T}}\right), \quad (5)$$

where $\bar{\boldsymbol{u}} = (\bar{u}, \bar{v})^{\mathrm{T}}$ is a characteristic point of a region, e. g., in this case, the center of gravity. The last equivalence is only valid for horizontally rectified images; the definition for vertically rectified images is similar. The labels are defined using the disparity relation between any two images of the series. The labeling function $s(\cdot)$ for areas is used similarly.

2. Region based stereo fusion

The main challenges to deal with when registering stereo and spectral series are clearly displayed in Fig. 2:

- Corresponding points (points mapping the same 3D scene point in different images) have different gray values. For example, the stripes of the tiger are not visible in near infrared (NIR, 800 nm and above).
- Neighboring regions have different contrasts in different images of the series. For example, the regions forming the nose of the tiger have similar gray values in the image at 400 nm, but considerably different values at 600 nm.

Considering these challenges, a pixel based registration (Scharstein & Szeliski (2002); Seitz et al. (2006)) of the images is not appropriate. The methods for the region based registration can be divided into three categories:

- The straightforward way is to search for 1:1 region correspondences (or 1:0, if a region does not have a correspondent) in two segmented images using features of the regions, e. g., size or shape (Wang & Zheng (2008)).
- The second possible way is to determine 1:N ($N > 0$) correspondences, i. e., allowing a region in one image to correspond to more than one region in another image. It is reasonable that these regions must be connected.
- The third way uses the images without the need of segmentation. The registration is done, e. g., by means of entropy methods, which compare image areas without using gray values, but by analyzing the information content.

We propose two different region based registration methods for evaluating the stereo effect and determining spatial information (Gheţa et al. (2010; 2008)), according to the first two possibilities. The use of entropy based methods is not appropriate in this case, due to the fact that the scenes may not be entirely structured. As already mentioned, the images of the combined series are rectified and segmented prior to evaluation (Faugeras & Luong (2004); Gonzalez & Woods (2008)). A segmentation example of the image series of Fig. 2 is given in Fig. 4.

2.1 Registration based only on regions

In this case, 1:1 (i. e., two regions correspond) or 1:0 (a region has no correspondent) correspondences are determined. Therefore, an energy functional comprising two terms is defined and minimized:

$$E(s, B) := E_{\mathrm{d}}(s, B) + E_{\mathrm{n}}(s, B) \to \min\,. \quad (6)$$

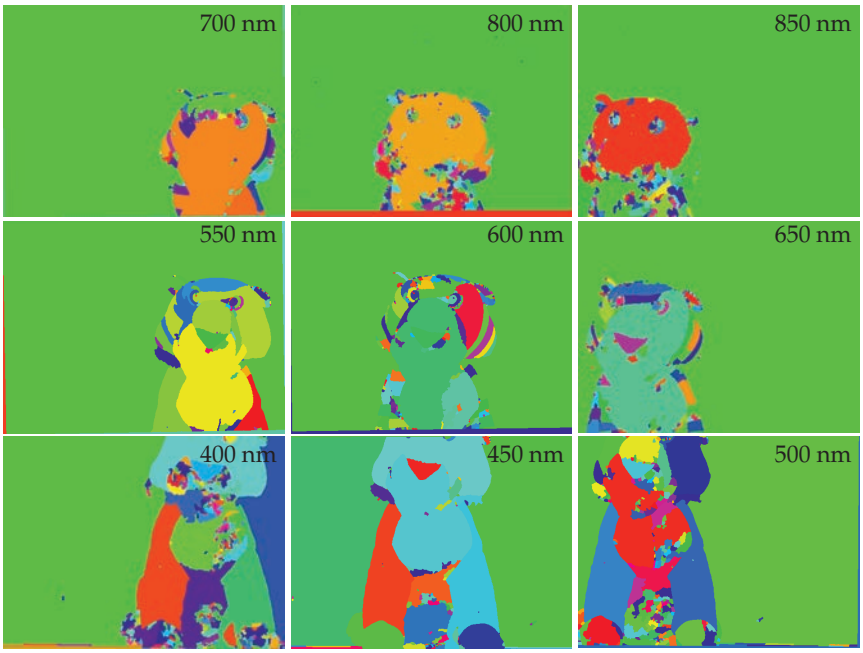


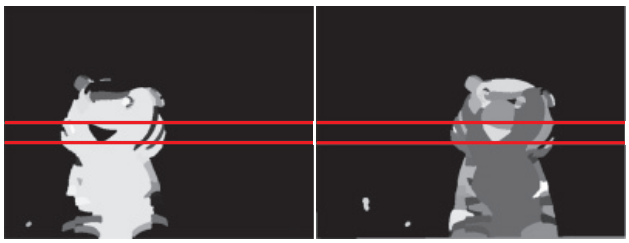
Fig. 4. False color representation of segmented images of the series in Fig. 2. The colors are just labels for the pixel assignment to regions, without marking corresponding regions.

The purpose of the minimization is to find an optimal function $s(\cdot)$ as defined in Eq. (5). The first term in Eq. (6) is the data term, which ensures the consistency of the result with the measured data, i. e., the difference between features of corresponding regions should be small. The second term defines additional constraints, e. g., neighboring constraints.
Since the images are rectified, the search for correspondences can take place without loss of generalization in an epipolar area, which is defined through two epipolar lines; see the example in Fig. 5. That way, the optimization is less time consuming.

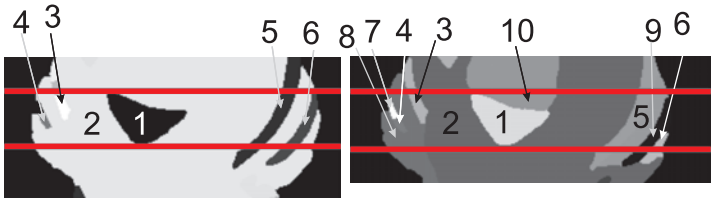
Data term
The data term $E_\mathrm{d}(s, B)$ of this approach evaluates the similarities between pairs of regions according to certain features, e. g., size, shape and position with regard to epipolar lines. It is defined as a sum over all image pairs in the series and all corresponding region pairs:

$$E_\mathrm{d}(s, B) := \sum_{\substack{(B_i, B_j) \\ i \neq j}} \sum_{\mathcal{R}_i \leftrightarrow \mathcal{R}_j} d_\mathrm{M}(\boldsymbol{m}_{\mathcal{R}_i}, \boldsymbol{m}_{\mathcal{R}_j}) \tag{7}$$

with $\boldsymbol{m}_\mathcal{R}$ as a feature vector characterizing the region $\mathcal{R}$ and $d_\mathrm{M}(\cdot, \cdot)$ as a function to measure the dissimilarity between two regions. The correspondences are defined over the function $s(\cdot)$; see Section 1.3.



(a) Two images of a combined stereo and spectral series;



(b) Example of epipolar areas defined by two epipolar lines.

Fig. 5. Search for correspondences in an epipolar area.

$\boldsymbol{m}_{\mathcal{R}}$ contains three features of the region $\mathcal{R}$ which are invariant to gray values: size $g_{\mathcal{R}}$, shape ${\boldsymbol{k}_{\mathcal{R}}}^{\mathrm{T}}$ (defined by the boundaries of the region) and position $p_{\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}$ on the epipolar line given by $\bar{\boldsymbol{u}}_{\mathcal{R}}$:

$$\boldsymbol{m}_{\mathcal{R}} := \left(g_{\mathcal{R}}, {\boldsymbol{k}_{\mathcal{R}}}^{\mathrm{T}}, p_{\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}\right)^{\mathrm{T}}. \tag{8}$$

The size of a region is given by its number of pixels:

$$g_{\mathcal{R}} := |\mathcal{R}|\,. \tag{9}$$

The shape of a region is determined by analyzing its boundaries. Considering an affiliation function

$$I_{\mathcal{R}}(\boldsymbol{u}) := \begin{cases} 1, & \boldsymbol{u} \in \mathcal{R} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

segments of the boundaries are determined according to their angular orientation: $\zeta_i := i \cdot \Delta\zeta$, $i \in \{0, 1, \ldots, N-1\}$, $N \cdot \Delta\zeta = 360°$. The boundary of a region can then be written as the union of its segments $\mathcal{K}_{\mathcal{R},\zeta_i}$, which are also interpreted as sets of pixels:

$$\mathcal{K}_{\mathcal{R}} := \bigcup_i \mathcal{K}_{\mathcal{R},\zeta_i}\,. \tag{11}$$

The feature vector ${\boldsymbol{k}_{\mathcal{R}}}^{\mathrm{T}}$ characterizing the boundary of a region $\mathcal{R}$ has the standardized number of points in each segment of the boundary as components:

$$\boldsymbol{k}_{\mathcal{R}} := \left(k_{\mathcal{R},0°}, \ldots, k_{\mathcal{R},\zeta_i}, \ldots\right)^{\mathrm{T}} = \left(|\mathcal{K}_{\mathcal{R},0°}|/|\mathcal{K}_{\mathcal{R}}|, \ldots, |\mathcal{K}_{\mathcal{R},\zeta_i}|/|\mathcal{K}_{\mathcal{R}}|, \ldots\right)^{\mathrm{T}}. \tag{12}$$

The position of the region with respect to the epipolar line is defined as the proportion of the number of points situated above $g_{A\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}$ and the number of points of the region situated below

$g_{\mathrm{B}\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}$ the epipolar line given by the reference pixel $\bar{\boldsymbol{u}}_{\mathcal{R}}$ (e. g., the epipolar line on which the center of gravity $\bar{\boldsymbol{u}}_{\mathcal{R}}$ is situated):

$$p_{\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}} := \frac{g_{\mathrm{A}\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}}{g_{\mathrm{B}\mathcal{R},\bar{\boldsymbol{u}}_{\mathcal{R}}}}. \tag{13}$$

The dissimilarity between two feature vectors is measured by means of the Manhattan metric, due to its capability of dealing with different scales (Marques de Sá (2001)):

$$d_{\mathrm{M}}(\boldsymbol{m}_{\mathcal{R}_i}, \boldsymbol{m}_{\mathcal{R}_j}) := |g_{\mathcal{R}_i} - g_{\mathcal{R}_j}| + \sum_q |k_{\mathcal{R}_i,\zeta_q} - k_{\mathcal{R}_j,\zeta_q}| + |p_{\mathcal{R}_i,\bar{\boldsymbol{u}}_{\mathcal{R}_i}} - p_{\mathcal{R}_j,\bar{\boldsymbol{u}}_{\mathcal{R}_j}}|\,. \tag{14}$$

Neighboring term

As regularization, a neighboring term is defined using a set $\mathcal{U}_{1ij}$ and a sequence $\mathcal{U}_{2ij}$. The role of the neighboring term is that the order of corresponding regions is maintained in the images. For example, if regions 6 in Fig. 5(b) left and 9 in Fig. 5(b) right correspond, then, region 6 in Fig. 5(b) right is not allowed to correspond to any region left of region 6 in Fig. 5(b) left. The set $\mathcal{U}_{1ij}$ comprises all correspondences between regions of two epipolar areas E_i and E_j in two images, found on basis of the data term:

$$\mathcal{U}_{1ij} = \{(\mathcal{R}_i^k, \mathcal{R}_j^l) | (\mathcal{R}_i^k, \mathcal{R}_j^l) \in E_i \times E_j \wedge \mathcal{R}_i^k \leftrightarrow \mathcal{R}_j^l\}\,. \tag{15}$$

The sequence $\mathcal{U}_{2ij}$ comprises all correspondences out of the set $\mathcal{U}_{1ij}$, i. e., $(\mathcal{R}_i^k, \mathcal{R}_j^l)$ and $(\mathcal{R}_i^m, \mathcal{R}_j^n) \in \mathcal{U}_{1ij}$, that satisfy the neighboring constraint: the indices $k > 0$ and $m > 0$ of regions in image i must be in the same order, i. e., $k < m$, as the indices l and n, i. e., $l < n$, of their corresponding regions in image j :

$$\mathcal{U}_{2ij} = \{\ldots, (\mathcal{R}_i^k, \mathcal{R}_j^l), (\mathcal{R}_i^m, \mathcal{R}_j^n), \ldots | (\mathcal{R}_i^k, \mathcal{R}_j^l), (\mathcal{R}_i^m, \mathcal{R}_j^n) \in \mathcal{U}_{1ij} \wedge k < m \wedge l < n\}\,. \tag{16}$$

The neighboring term is therefore defined as a sum over all image pairs in the series and over pairs of epipolar areas:

$$E_{\mathrm{n}}(s, B) = \sum_{(B_i,B_j)\in\mathcal{I}} \sum_{(E_i,E_j)} \lambda\, |\mathcal{U}_{1ij} \setminus \mathcal{U}_{2ij}|\,, \tag{17}$$

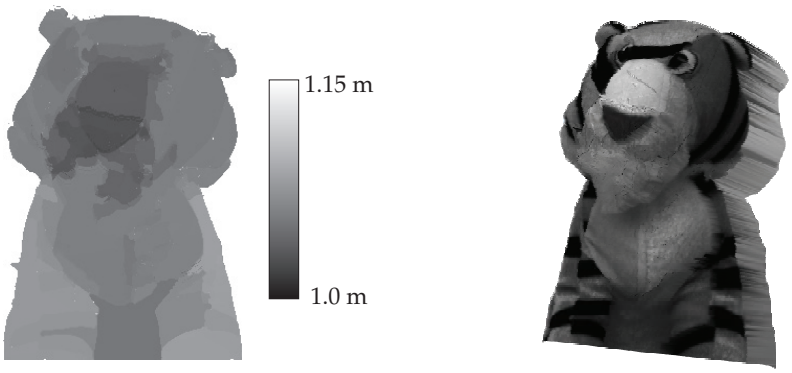
where $\lambda > 0$ is a constant, assuring that the energy increases, if the neighboring constraint is not satisfied.

Energy minimization using dynamic programming

In the case of registration based only on regions, the minimization of the energy functional is performed by means of dynamic programming (Bellman (1957); Bertsekas (2005)). Since the problem of finding region correspondences based on their features is deterministic and finite, and the costs for each step are additive, the employment of dynamic programming assures the optimal result. Moreover, the neighboring constraint is implicitly guaranteed through the employment of the forward dynamic programming algorithm.

The results of the registration based only on regions for the example of Fig. 2 are presented in Fig. 6 in form of a depth map and a textured $2\frac{1}{2}$D reconstruction. Their quality is high, though, the obtained depth maps might not be dense. This is due to the fact that depth values can only be computed for those regions for which correspondences have been found. Due to

the spectral component of the image series, the segmentation results are different within the images (i. e., the regions have different shapes and sizes), such that correspondences might not be found for all regions. In these cases, gaps may occur in depth maps.



(a) Depth map; (b) Textured $2\frac{1}{2}$D reconstruction.

Fig. 6. Depth map and reconstruction resulting from the registration based only on regions of the image series shown in Fig. 2.

2.2 Registration based on regions and areas

In this context, an (image) area is defined as a union of several regions. The registration based on regions and areas considers that a region may correspond to a union of regions, i. e., an area; see Section 1.3. This way, segmentation differences are accounted for. The registration problem is then modeled by the minimization of the following energy functional:

$$E(s, B) := E_{\mathrm{d}}(s, B) + \gamma_{\mathrm{s}} E_{\mathrm{s}}(s, B) \to \min . \tag{18}$$

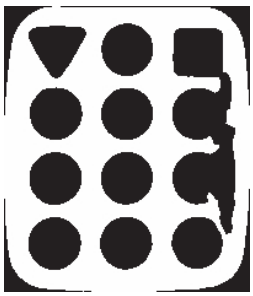
The data term $E_{\mathrm{d}}(s, B)$ measures the dissimilarity between a region and an area mainly based on their contours. For regularization, a smoothness term $E_{\mathrm{s}}(s, B)$ with an appropriate weighting factor γ_{s} is employed (Bleyer & Gelautz (2007)). The assumption modeled by the smoothness term is similar to the one usually made by pixel based registration: neighboring regions having similar gray values are likely to be mappings of the same 3D plane in space, and thus, they should have similar depth values.

Data term

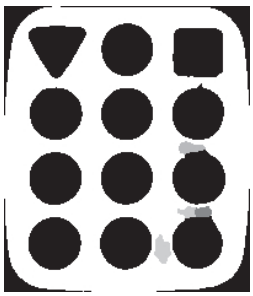
For each region $\mathcal{R}_i$ in image i, a feature $m_{\mathcal{R}_i,j}(s)$ is computed with respect to an area $\mathcal{B}_j$ in image j and a certain label s. The feature compares mainly the conformity of the boundaries of the region $\mathcal{R}_i$ and the area $\mathcal{B}_j$. Therefore, it can be interpreted as a distance measure for the correspondence between the region $\mathcal{R}_i$ and the respective area $\mathcal{B}_j$.
Following, the computation of the feature $m_{\mathcal{R}_i,j}(s)$ is explained based on an example. The first step consists in the definition and evaluation of four sets containing the pixels of the region $\mathcal{R}_i$ and their correspondences on the boundaries or in the interior of $\mathcal{B}_j$. As an example, two segmented images are shown in Figs. 7(a) and 7(b). The exemplary region $\mathcal{R}_i$ and area $\mathcal{B}_j$ are displayed in Figs. 7(c) and 7(d), respectively. Their boundaries are shown in Figs. 7(e) and 7(f).

(a) Segmented image $B_i(\boldsymbol{u}_i)$;

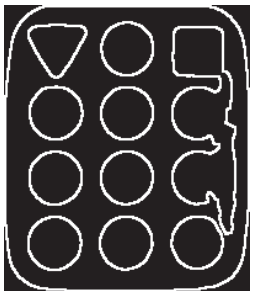
(b) Segmented image $B_j(\boldsymbol{u}_j)$;



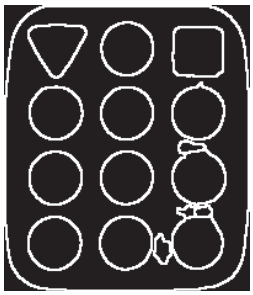
(c) Selected region $\mathcal{R}_i$ in image $B_i(\boldsymbol{u}_i)$;



(d) Selected area $\mathcal{B}_j$ in image $B_j(\boldsymbol{u}_j)$, different regions have different gray values (except black);



(e) $\mathcal{K}_{\mathcal{R}_i}$: Boundary of the region $\mathcal{R}_i$;



(f) $\bigcup_r \mathcal{K}_{\mathcal{R}_j^r}$: Boundaries of the area $\mathcal{B}_j$.

Fig. 7. Example of regions, areas and their boundaries for computing the sets of Eqs. (19) through (22).

The four sets used for the computation of the feature $m_{\mathcal{R}_i,j}(s)$ are defined as follows:

(a) The corresponding pixels $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are situated both on boundaries of regions; see Fig. 8(a):

$$\mathcal{M}_{\mathrm{a},\mathcal{R}_i,j}(s) = \{\boldsymbol{u}_i | \boldsymbol{u}_i \in \mathcal{K}_{\mathcal{R}_i} \wedge \exists\, r \in \mathcal{J} : \boldsymbol{u}_j \in \mathcal{K}_{\mathcal{R}_j^r} \wedge s(\boldsymbol{u}_i) = s(\boldsymbol{u}_j)\}\,. \tag{19}$$

(b) The pixel $\boldsymbol{u}_i$ is situated on the boundary of $\mathcal{R}_i$, while its corresponding pixel $\boldsymbol{u}_j$ is situated in the interior of the area $\mathcal{B}_j$ in image j; see Fig. 8(b):

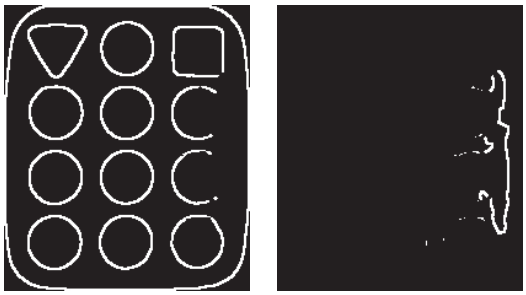
$$\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s) = \{\boldsymbol{u}_i | \boldsymbol{u}_i \in \mathcal{K}_{\mathcal{R}_i} \wedge \exists\, r \in \mathcal{J} : \boldsymbol{u}_j \in \mathcal{R}_j^{r\circ} \wedge s(\boldsymbol{u}_i) = s(\boldsymbol{u}_j)\}\,. \tag{20}$$

(c) The pixel $\boldsymbol{u}_i$ is situated in the interior of the region $\mathcal{R}_i$, while its corresponding pixel $\boldsymbol{u}_j$ is situated on a boundary; see Fig. 8(c):

$$\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s) = \{\boldsymbol{u}_i | \boldsymbol{u}_i \in \mathcal{R}_i^{\circ} \wedge \exists\, r \in \mathcal{J} : \boldsymbol{u}_j \in \mathcal{K}_{\mathcal{R}_j^r} \wedge s(\boldsymbol{u}_i) = s(\boldsymbol{u}_j)\}\,. \tag{21}$$

(d) Both corresponding pixels $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are situated in the interior of the region $\mathcal{R}_i$ and the area $\mathcal{B}_j$, respectively; see Fig. 8(d):

$$\mathcal{M}_{\mathrm{d},\mathcal{R}_i,j}(s) = \{\boldsymbol{u}_i | \boldsymbol{u}_i \in \mathcal{R}_i^{\circ} \wedge \exists\, r \in \mathcal{J} : \boldsymbol{u}_j \in \mathcal{R}_j^{r\circ} \wedge s(\boldsymbol{u}_i) = s(\boldsymbol{u}_j)\}\,. \tag{22}$$



(a) $\mathcal{M}_{\mathrm{a},\mathcal{R}_i,j}(s)$; (b) $\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s)$;

(c) $\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s)$; (d) $\mathcal{M}_{\mathrm{d},\mathcal{R}_i,j}(s)$.

Fig. 8. Sets according to Eqs. (19) through (22) for the example shown in Fig. 7.

The four sets defined in Eqs. (19) through (22) are shown in Figs. 8(a) through 8(d). The first set $\mathcal{M}_{\mathrm{a},\mathcal{R}_i,j}(s)$ in Eq. (19) (see Fig. 8(a)) can be interpreted as the intersection of the boundary of the region $\mathcal{R}_i$ (see Fig. 7(e)) and those of the area $\mathcal{B}_j$ (see Fig. 7(f)).
The second set $\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s)$ of Eq. (20), shown in Fig. 8(b), corresponds to the intersection of the boundary of the region $\mathcal{R}_i$ (see Fig. 7(e)) and the interior of the area $\mathcal{B}_j$ (see Fig. 7(d)).
The third set $\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s)$ of Eq. (21), shown in Fig. 8(c), is equivalent to the intersection of the interior of the region $\mathcal{R}_i$ (see Fig. 7(c)) with the boundaries of the area $\mathcal{B}_j$ (see Fig. 7(f)).

Fig. 8(d) displays the fourth set $\mathcal{M}_{\mathrm{d},\mathcal{R}_i,j}(s)$ (see Eq. (22)), which is the intersection of the interior of the region $\mathcal{R}_i$ (see Fig. 7(c)) and the interior of the area $\mathcal{B}_j$ (see Fig. 7(d)).
The cardinal numbers of the four sets are used to build the feature $m_{\mathcal{R}_i,j}(s)$ characterizing the region $\mathcal{R}_i$ with regard to the area $\mathcal{B}_j$:

$$\begin{aligned} m_{\mathcal{R}_i,j}(s) :=& \frac{|\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s)|}{|\mathcal{M}_{\mathrm{a},\mathcal{R}_i,j}(s)| + |\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s)|} + \gamma \frac{|\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s)|}{|\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s)| + |\mathcal{M}_{\mathrm{d},\mathcal{R}_i,j}(s)|} \\ =& \frac{|\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s)|}{|\mathcal{K}_{\mathcal{R}_i}|} + \gamma \frac{|\mathcal{M}_{\mathrm{c},\mathcal{R}_i,j}(s)|}{|\mathcal{R}_i^{\circ}|} \end{aligned} \tag{23}$$

with $\gamma > 0$ as a weighting factor.
The first term of Eq. (23) evaluates the share of pixels on the boundary of region $\mathcal{R}_i$ having no correspondences on any boundary of the area $\mathcal{B}_j$. The term models the preference for correspondences that match the boundaries of the region $\mathcal{R}_i$ and the area $\mathcal{B}_j$.
The second term evaluates the share of pixels situated in the interior of the region $\mathcal{R}_i$ with correspondences on any boundary of the area $\mathcal{B}_j$. In this way, the possibility of a 1:N assignment is modeled, i. e., correspondences between the region $\mathcal{R}_i$ and more than one region $\mathcal{R}_j^r$ in the image j (forming the area $\mathcal{B}_j$). However, the value of the term is higher, if such 1:N correspondences occur, compared to the case of 1:1 correspondences. Therefore, 1:N correspondences are admitted, but 1:1 correspondences are favored.
The first term of Eq. (23) reaches its minimum (equal to zero), if $\mathcal{M}_{\mathrm{b},\mathcal{R}_i,j}(s) = \emptyset$. In such a case, the boundaries of the region $\mathcal{R}_i$ and those of the area $\mathcal{B}_j$ match perfectly. In any other case, the value of the term varies between zero and one. The second term has its minimum (equal to zero), when the area $\mathcal{B}_j$ consists of only one region, i. e., all pixels in the interior of the region $\mathcal{R}_i$ have correspondences in the interior of the area $\mathcal{B}_j$ (a 1:1 correspondence exists between the region $\mathcal{R}_i$ and the area $\mathcal{B}_j$).
To conclude, it can be said that the smaller the value of feature $m_{\mathcal{R}_i,j}(s)$ in Eq. (23) is, the more similar the region $\mathcal{R}_i$ and the area $\mathcal{B}_j$ are. Modeling the data term is straightforward by summing the values of the feature $m_{\mathcal{R}_i,j}(s)$ over all regions in the images of the series:

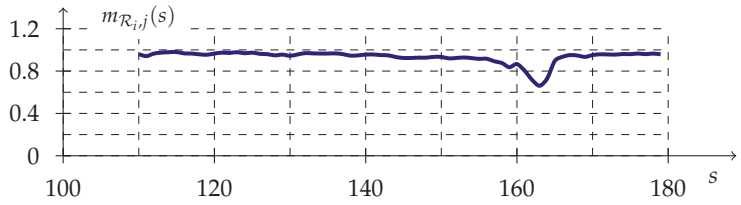
$$E_{\mathrm{d}}(s, B) := \sum_{(B_i,B_j)\in\mathcal{I}} \sum_{\mathcal{R}_i} m_{\mathcal{R}_i,j}(s) \tag{24}$$

with $\mathcal{I} := \{(B_i, B_j)|i \neq j\}$ being the set of image pairs.
For regions $\mathcal{R}_i$ with the feature $m_{\mathcal{R}_i,j}(s)$ taking the constant value one for almost all labels s (e. g., small regions), it is hard to determine the corresponding area $\mathcal{B}_j$. Such regions have no distinct minimum for the feature $m_{\mathcal{R}_i,j}(s)$ over the value range of s (see Fig. 9). To solve this problem, a second energy term is proposed, which takes the role of a smoothness term regularizing the problem.

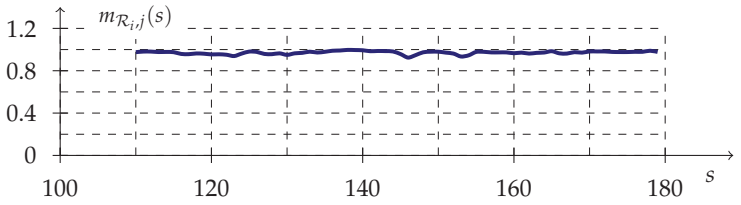
Smoothness term

For regularization, an energy term $E_{\mathrm{s}}(s, B)$ evaluating the neighborhood relations between pairs of corresponding regions is defined. The idea is that neighboring regions with similar gray values are assumed to be part of the same 3D plane, i. e., they should be assign similar labels s. This assumption is like the smoothness assumption made in the case of pixel based registration: pixels having similar gray values are likely to belong to the same plane and should therefore be labeled similarly. Moreover, by means of the smoothness term, regions for which the data term is approximately one over the entire value range of the label s (e. g., small

(a) Segmented image pair;



(b) Values of the feature $m_{\mathcal{R}_i,j}(s)$ for the horizontally striped region in the image of Fig. 9(a) right over the value range of the label s;



(c) Value of the feature $m_{\mathcal{R}_i,j}(s)$ for the vertically striped region in the image of Fig. 9(a) right over the value range of the label s.

Fig. 9. Value of the feature $m_{\mathcal{R}_i,j}(s)$ for regions in the image of Fig. 9(a) right over the value range of the label s, i. e., for different areas of the image in Fig. 9(a) left.

regions) receive the label of one of their neighbors. In order to determine which neighbor is appropriate, the mean gray values of the regions are computed and compared.
The smoothness term thus penalizes neighboring regions that have similar mean gray values and different labels s. The proposed model is similar to that in (Bleyer & Gelautz (2007)):

$$E_{\mathrm{s}}(s,B) := \sum_{B_i} \sum_{\mathcal{R}_i^k} \sum_{\mathcal{R}_i^l \in \mathcal{N}_{\mathrm{R}}(\mathcal{R}_i^k)} \left(1 - \delta_{s(\mathcal{R}_i^k)}^{s(\mathcal{R}_i^l)}\right) \cdot f_1(\mathcal{R}_i^k, \mathcal{R}_i^l) \cdot f_2(\mathcal{R}_i^k, \mathcal{R}_i^l)\,, \tag{25}$$

where $\mathcal{N}_{\mathrm{R}}(\mathcal{R}_i^k)$ is the set of neighboring regions $\mathcal{R}_i^l$ of the region $\mathcal{R}_i^k$.
Function $f_1(\mathcal{R}_i^k, \mathcal{R}_i^l)$ computes the length of the common boundary of two directly neighboring regions $\mathcal{R}_i^k$ and $\mathcal{R}_i^l$ with regard to the length of the boundary of the smaller

region:

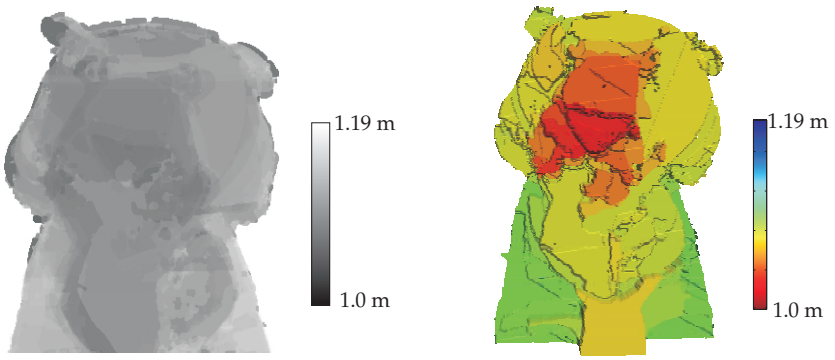
$$f_1(\mathcal{R}_i^k, \mathcal{R}_i^l) := \frac{\left|\{\boldsymbol{u}_i^k \in \mathcal{K}_{\mathcal{R}_i^k} |\ \exists \boldsymbol{u}_i^l \in \mathcal{K}_{\mathcal{R}_i^l} \wedge \boldsymbol{u}_i^l \in \mathcal{N}_\mathrm{P}(\boldsymbol{u}_i^k)\}\right|}{\min(|\mathcal{K}_{\mathcal{R}_i^k}|, |\mathcal{K}_{\mathcal{R}_i^l}|)}, \tag{26}$$

where $\mathcal{N}_\mathrm{P}(\boldsymbol{u}_i^k)$ is the set of neighboring pixels $\boldsymbol{u}_i^l$ of pixel $\boldsymbol{u}_i^k$. $f_1(\mathcal{R}_i^k, \mathcal{R}_i^l)$ takes high values for small regions, which consequently increases the value of the smoothness term $E_\mathrm{s}(s, B)$.
Function $f_2(\mathcal{R}_i^k, \mathcal{R}_i^l)$ evaluates the differences in the mean gray values of the two neighboring regions $\mathcal{R}_i^k$ and $\mathcal{R}_i^l$:

$$f_2(\mathcal{R}_i^k, \mathcal{R}_i^l) := \left(1 - \frac{\min(|\bar{g}_{\mathcal{R}_i^k} - \bar{g}_{\mathcal{R}_i^l}|, K)}{K}\right) \cdot \gamma + (1 - \gamma) \tag{27}$$

with $0 < \gamma < 1$. $\bar{g}_{\mathcal{R}_i^k}$ is the mean gray value of the region $\mathcal{R}_i^k$:

$$\bar{g}_{\mathcal{R}_i^k} := \frac{1}{|\mathcal{R}_i^k|} \sum_{\boldsymbol{u}_i^k \in \mathcal{R}_i^k} B_i(\boldsymbol{u}_i^k)\,. \tag{28}$$



(a) Depth map; (b) $2\frac{1}{2}$D reconstruction.

Fig. 10. Results of the registration based on regions and areas.

Energy minimization using graph cuts

For the minimization of the energy functional in the case of the registration based on regions and areas, a state-of-the-art graph-cuts algorithm (Boykov et al. (2001)) is employed with minor adaptations. For example, the construction of the graph is modified such that the nodes of the graph are not pixels, but regions. The results of the registration based on regions and areas of the image series in Fig. 2 are presented in Fig. 10. The resulting depth maps using this registration approach not only have a high quality, but they are also dense, since the smoothness term insures that for each region, a depth value is determined.
The depth maps obtained by evaluating the stereo information of combined stereo and spectral series can be further used for spectral fusion.

Fig. 11. Scene of an orange model robot holding an orange log of wood and an orange background.

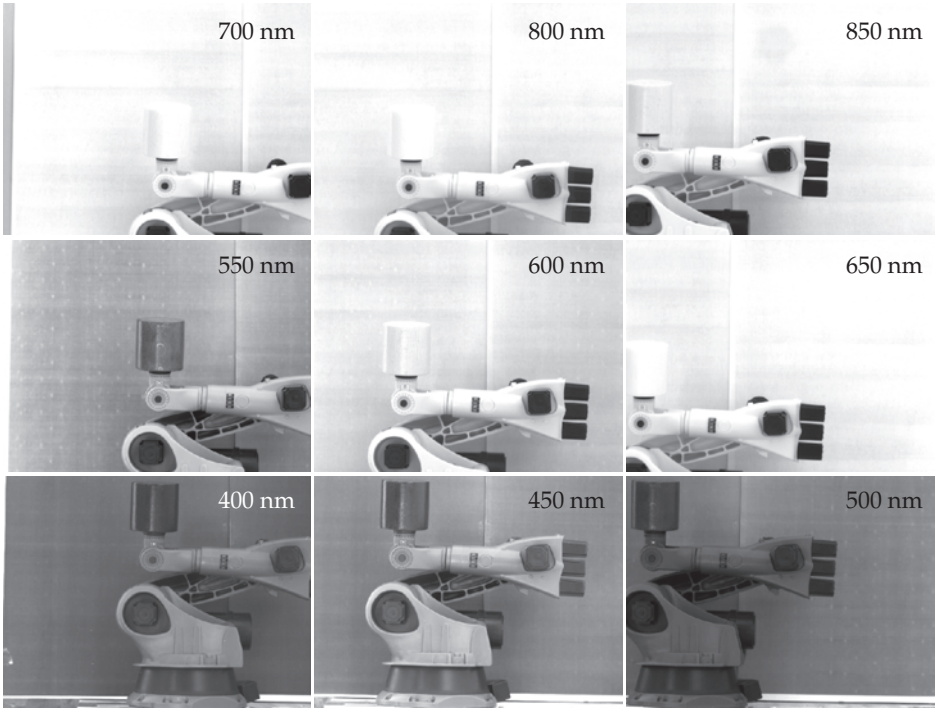


Fig. 12. Combined stereo and spectral series of the scene depicted in Fig.11.

3. Spectral fusion

For determining the spectral properties of a scene on the basis of the combined image series, the depth map obtained with the previously described registration methods is employed. An exemplary scene contains an orange model robot holding an orange log of wood and an

orange background; see Fig. 11. From this scene, an image series has been acquired with a camera array equipped with spectral filters; see Fig. 12.

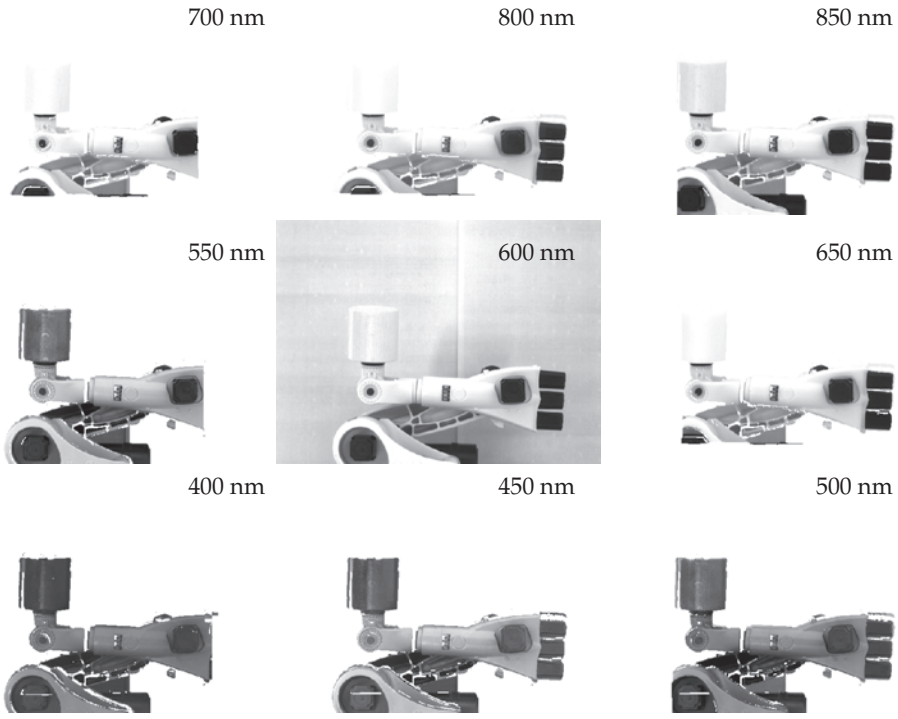


Fig. 13. Warped images of the series of Fig. 12 into the coordinate system of the middle camera of the array.

By means of image warping, all images are transferred into a common coordinate system, e. g., of one of the cameras (here we chose the middle camera of the array); see Fig. 13. In the case of stereo and spectral series, this leads to a pure spectral series, i. e., the stereo effect is eliminated. The spectral series contains the spectral characteristics for each mapped 3D scene point according to the employed spectral filters. The obtained intensity values can be interpreted as spectral features and can be further used, e. g., for material classification or for improving object detection. As an example, a common approach for analyzing spectral features is given in Tyo et al. (2003). The first step of this analysis consists in a dimension reduction for the spectral feature vectors, followed by a false color representation. The dimension reduction is done by employing the Principal Component Analysis (PCA, Gonzalez & Woods (2008)); the transformed images $(P_1, \ldots, P_n)$ are presented in Fig. 14. The components with the highest information content, i. e., corresponding to the highest eigenvalues, are chosen for the next step consisting in a false color representation (Tyo et al. (2003)). For a meaningful representation, the first three components are chosen according to their information content.

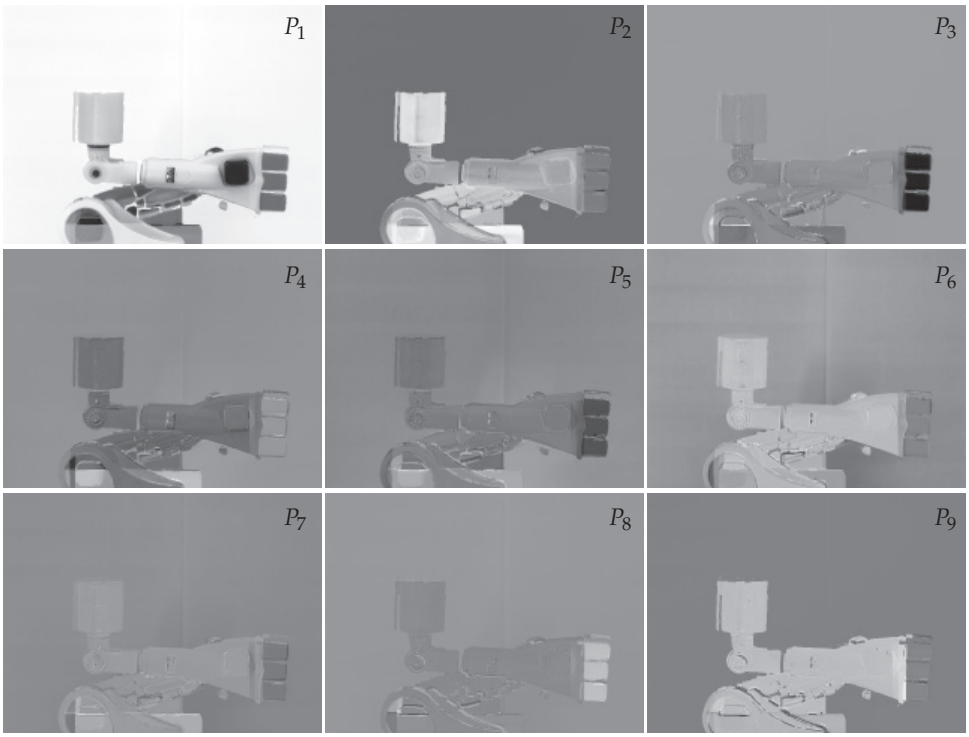


Fig. 14. Standardized principal components of the images of Fig. 13.

These images can be interpreted as RGB (red, green, blue) channels, but they can also be preferably transformed into the HSV (hue, saturation,value) color space:

$$\begin{aligned} H &: \mathbb{R}^2 \to \mathbb{R}\,, \quad H(\boldsymbol{u}) := \arctan\left(\frac{P_3(\boldsymbol{u})}{P_2(\boldsymbol{u})}\right), \\ S &: \mathbb{R}^2 \to \mathbb{R}\,, \quad S(\boldsymbol{u}) := \frac{\sqrt{P_2(\boldsymbol{u})^2 + P_3(\boldsymbol{u})^2}}{P_1(\boldsymbol{u})}\,, \\ V &: \mathbb{R}^2 \to \mathbb{R}\,, \quad V(\boldsymbol{u}) := P_1(\boldsymbol{u})\,. \end{aligned} \tag{29}$$

The advantages of using the HSV color space for false color representation are:

- The color representation is very similar to the way humans perceive color, i. e., three channels are considered: the achromatic channel and two orthogonal color channels. The achromatic channel, i. e., V in the HSV space, represents the intensity. The color channels are represented in the HSV color space by the components H and S.
- Data analysis in the HSV color space is simplified through the intuitively understandable meaning of its components.

The connection between the two color spaces is visually shown in Fig. 15. Figure 16 displays the H, S and V components obtained for the given example. The computed false color representation of the model robot scene is presented in Fig. 17.

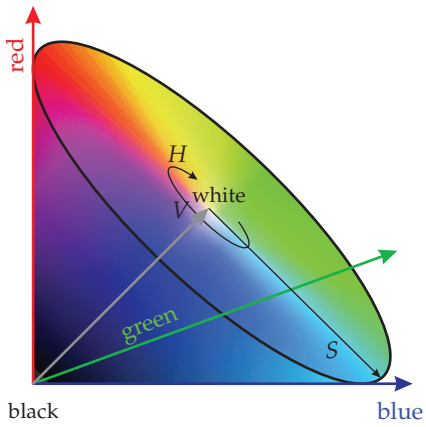


Fig. 15. Connection between RGB and HSV color spaces.

Fig. 16. H, S and V components as a result of the transformation of the first three principal components.

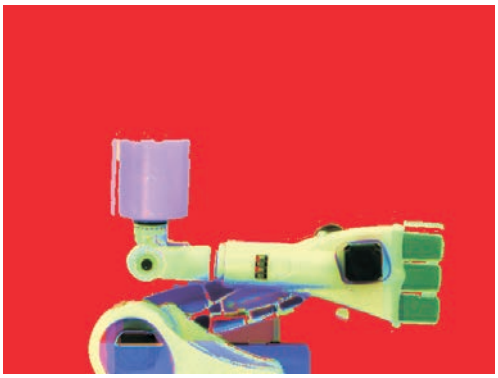


Fig. 17. False color representation for the scene of Fig. 11.

4. Conclusions

The challenges when fusing combined stereo and spectral image series are that corresponding pixels have different gray values and that neighboring image regions have different contrasts in different images. Therefore, common pixel and gray value based registration techniques cannot be applied. This chapter proposes two novel region based image registration approaches for combined stereo and spectral series. Prior to registration, the images are rectified and segmented by means of the watershed transform. The first approach, based only on regions, determines 1:1 region correspondences by measuring the dissimilarity between the feature vectors characterizing the regions. Thus, a gray value invariance is achieved. For regularization, a neighboring constraint is imposed. The second approach, based on regions and areas, searches for 1:N correspondences in images by mainly comparing the boundaries of the regions. The main advantage of this approach is that it can handle differences in the segmentation of the images, e. g., due to the use of spectral filters. As regularization, a smoothness constraint is enforced in the case of the registration based on regions and areas. In both cases, the registration problem is modeled by energy functionals. The solutions are obtained through the minimization of the functionals by dynamic programming or graph cuts algorithms. The registration results are directly employed for computing depth maps.
By means of image warping based on the obtained depth information, the originally combined image series can be transformed into a pure spectral series. In this chapter, an example of the analysis of the spectral information is given in form of a false color representation. Further work will concentrate on the evaluation of the results and on finding an appropriate method for the fusion and analysis of the spectral information.

5. Acknowledgments

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.

6. References

Bankman, J. N. (ed.) (2000). *Handbook of Medical Imaging Processing and Analysis*, Academic Press.

Bellman, R. (1957). *Dynamic Programming*, Princeton University Press.

Bertsekas, D. P. (2005). *Dynamic Programming and Optimal Control*, Athena Scientific.

Bleyer, M. & Gelautz, M. (2007). Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions, *Signal Processing: Image Communication* 22(2): 127–143.

Boykov, Y. & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(9): 1124–1137.

Boykov, Y., Veksler, O. & Zabih, R. (2001). Fast approximate energy minimization via graph cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11): 1222–1239.

Chang, C.-I. (2003). *Hyperspectral imaging: techniques for spectral detection and classification*, Kluwer Academic.

Faugeras, O. & Luong, Q.-T. (2004). *The Geometry of Multiple Images*, MIT Press.

Fookes, C., Maeder, A., Sridharan, S. & Cook, J. (2004). Multi-spectral stereo image matching using mutual information, *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)*.

Frese, C. & Gheţa, I. (2006). Robust depth estimation by fusion of stereo and focus series gained with a camera array, *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 243–248.

Gheţa, I., Frese, C. & Heizmann, M. (2006). Fusion of combined stereo and focus series for depth estimation, *Informatik 2006 – Informatik für Menschen*, Vol. 1, pp. 359–363.

Gheţa, I., Frese, C., Heizmann, M. & Beyerer, J. (2007). A new approach for estimating depth by fusing stereo and defocus information, *Informatik 2007 – Informatik trifft Logistik*, pp. 26–31.

Gheţa, I., Frese, C., Krüger, W., Saur, G., Heinze, N., Heizmann, M. & Beyerer, J. (2007). Depth estimation from flight image series using multi-view along-track stereo, *Optical 3D Measurement Techniques, Zürich*, pp. 119–125.

Gheţa, I., Höfer, S., Heizmann, M. & Beyerer, J. (2010). A novel approach for the fusion of combined stereo and spectral series, *in* D. Fofi & K. Niel (eds), *Image Processing: Machine Vision Applications III, IS&T/SPIE Electronic Imaging, Proceedings of SPIE 7538*, Vol. 7538, San Jose, California.

Gheţa, I., Mathias, M., Heizmann, M. & Beyerer, J. (2008). Fusion of combined stereo and spectral series for obtaining 3d information, *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Proceedings of SPIE 6974*, Orlando, Florida.

Gonzalez, R. C. & Woods, R. E. (2008). *Digital Image Processing*, Prentice Hall.

Hartley, R. & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*, 2 edn, Cambridge University Press.

Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. (2008). *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc.

Marques de Sá, J. P. (2001). *Pattern Recognition: Concepts, Methods and Applications*, Springer Verlag.

Modersitzki, J. (2004). *Numerical Methods for Image Registration*, 1 edn, Oxford University Press.

Scharstein, D. & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision* 47(1–3): 7–42.

Seitz, S., Curless, B., Diebel, J., Scharstein, D. & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, New York, pp. 519–528.

Tyo, J. S., Konsolakis, A., Diersen, D. I. & Olsen, R. C. (2003). Principal-components-based display strategy for spectral imagery, *IEEE Transactions on Geoscience and Remote Sensing* 41(3): 708–718.

Wang, Z.-F. & Zheng, Z.-G. (2008). A region based stereo matching algorithm using cooperative optimization, *IEEE Conference on Computer Vision and Pattern Recognition, 2008*, pp. 1–8.

Weng, J., Cohen, P. & Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(10): 965–980.