SEGMENTATION OF TERRESTRIAL LASER SCANNING DATA BY INTEGRATING RANGE AND IMAGE CONTENT

Shahar Barnea, Sagi Filin

Transportation and Geo-Information, Civil and Environmental Engineering Faculty, Technion – Israel Institute of Technology, Haifa, 32000, Israel - {barneas, filin}@technion.ac.il

Commission V, WG V/5

KEY WORDS: Segmentation, Terrestrial Laser Scanner, Point Cloud, Algorithms, Object Recognition

ABSTRACT:

Terrestrial laser scanning is becoming a standard technology for 3D modeling of complex scenes. Laser scans contain detailed geometric information, but still require interpretation of the data for making it useable for mapping purposes. A fundamental step in the transformation of the data into objects involves their segmentation into consistent units. These units should follow some predefined rules, and result in salient regions guided by the desire that the individual segments represent object or object-parts within the scene. Nonetheless, due to the scene complexity and the variety of objects in it, it is clear that a segmentation using only a single cue will not suffice. Considering the availability of additional data sources like the color channels, more information can be integrated in the data partitioning process and ultimately into the reconstruction scheme. We propose in this paper the segmentation of terrestrial laser scanning data by the integration of range and color content and by using multiple cues. This concept raises questions regarding the mode of their integration, and definition of the expected outcome. We show, that while individual segmentation based on given cues have their own limitations; their integration provide a more coherent partitioning that has better potential for further processing.

1. INTRODUCTION

Terrestrial laser scanners emerged in recent years as standard measuring technology for detailed 3D modeling of scenes. From a geometrical perspective, scanners provide rich and accurate information of the acquired scene. Additionally, with cameras becoming an integral part of modern scanners, the resulting radiometric information provides supplementary color content. The combination of direct geometric details and radiometric content offers excellent foundations for the extraction of objects in an autonomous manner.

Raw data (3D points and 2D RGB pixels) resulting from a single scan can reach tens of millions of elemental units. However, for common laser scanning applications, e.g., mapping, modeling, and object extraction that require high level of abstraction, this huge amount of data is hard to use. A fundamental step in the extraction of objects is the application a mid-level processing phase involving the grouping of pixels containing redundant information into segments. Essentially, each segment should form a collection of 3D points in which two conditions must be met, one is that the segment will maintain geometrical connectivity among all points constituting it; the second is that the feature value for the connected points will share similarity of some measure. Similarity can be geometrically based, radiometric based, or both. In addition, the basic units of each segment have to create a spatial continuation in the 3D sense. While segmentation of image content, and to some degree, of terrestrial point clouds, has been studied in the past, segmentation of the combined set has not been addressed by many so far. The motivation for pursuing this avenue is however clear and relates to the desire to benefit from the descriptive power of the rich radiometric content while being subjected to objects geometry and spatial connectivity in 3D space.

In general, segmentation concerns partitioning the data into disjoint salient regions usually under the assumption that individual segments tend to represent individual objects within the scene. Due to its important role, segmentation has been studied for years beginning from thresholding techniques (Otsu, 1979; Huang et al., 2005) and classic "region growing" based methods (e.g., Pal and Pal, 1993). Other methods propose converting the image into a feature space, and by doing so transforming the segmentation problem into a classification task. Carson et al. (2002) propose modeling the distribution of feature vectors as a mixture of Gaussians, with the model parameters being estimated using the expectation-maximization algorithm. Graph based approaches have been receiving growing attention. Using this scheme images are viewed as a graph in which each vertex represent a pixel (Shi and Malik, 2000; Felzenszwalb and Huttenlocher, 2004). The graph-view enables an intuitive representation of the segmentation problem as similarity between pixels can be assigned to the edges linking them. The challenge is then to find sets of vertices such that each has high connectivity value between its vertices and low connectivity to the rest of the graph. For a computational model for such segmentation, normalized cuts algorithm has been proposed (Shi and Malik, 2000). Sharon et al. (2000) make use of the multi-grid theory (Brandt, 1986) to solve efficiently the normalized-cut problem. A comprehensive review and test of some of the leading segmentation algorithms is provided in Estrada and Jepson (2005). Recent works, e.g., Russell et al. (2006), Roth and Ommer, (2006), Mian et al. (2006), and Alpert et al., (2007) demonstrated the application of segmentation processes for recognition tasks, showing promising results both in relation to object class recognition and to correct segmentation of the searched objects. Applications making use of segmentation as part of other tasks, have been reported for stereovision and image registration purposes (Bleyer and Gelautz, 2004; Klaus et al., 2006; Coiras et al., 2000).

Segmentation of laser scans offers a slightly different problem, as the data usually defines the geometric characterization of the scanned objects. Therefore, the interest is usually in the primitive extraction, and mostly in planar elements, e.g., Dold and Brenner (2006) for terrestrial scans and Vossleman and Dijkman, (2001) for aerial scans. For terrestrial scans Gorte (2007) presented a method for extracting planar faces using panoramic representation of the range data. Segmentation into a more general class of well-defined primitives, e.g., planes, cylinders, or spheres, is presented in Rabanni (2006). While being useful for reverse engineering practices it cannot be easily extended into general scenes.

Since most scenes are cluttered and contain entities of various shapes and forms, among which some are structured but others are not, approaching the segmentation problem by seeking consistency along a single cue is likely to provide partial results. Additionally, while some entities may be characterized by geometric properties, others are more distinguishable by their color content. Those realizations suggest that segmenting the data using multiple cues and integrating data source have the potential of providing richer descriptive information, and have better prospects for subsequent interpretation of the data. We present in this paper a segmentation model for terrestrial laser scanning data including range and image data while using multiple cues. We study how segments are defined when those sources should be merged together, how those sources should be integrated in a meaningful way, and ultimately how the added value of combining the individual sources can be brought into an integrated segmentation. Results of the proposed model show that better results than what is obtained by the individual segmentations can be achieved.

2. METHODOLOGY

The integration of different information sources requires securing their co-alignment, and association. The first aspect refers to establishing the relative transformation between the two sensors. The second suggests that in order to incorporate the interpretation of the two data sources, both have to refer to the same information unit. Considering the fact that images are a 2D projection of 3D space, whereas laser data is three dimensional, their mode of integration is not immediate.

2.1 Camera Scanner Co-alignment

The camera mounted on top of the scanner can be linked to the scanner body by finding the transformation between the two frames shown in Figure 1. Such relation involves three offset parameters and three angular parameters. This relation can also be formulated via the projection matrix **P**. With **P** a 3x4 matrix that represents the relation between world 3D point (X) and image 2D point (x) in homogeneous coordinates. Compared to the six standard boresighting pose parameters, the added parameters (five in all) will account to intrinsic camera parameters. The projection matrix can be formulated as follows:

$$x = \mathbf{K}\mathbf{R}\left[\mathbf{I} \mid -\mathbf{t}\right]X = \mathbf{P}X \qquad (1)$$

with

$$\mathbf{K} = \begin{bmatrix} f_x & s & x_0 \\ & f_y & y_0 \\ & & 1 \end{bmatrix}$$

f_x and f_y are the focal lengths in the x and y directions respectively, s is the skew value, x_0 and y_0 are the offsets with respect to the two image axes. **R** is the rotation matrix between the scanner and the camera frames (the red and the blue coordinate systems in the figure respectively) and **t** the translation vector (Hartley and Zisserman, 2003).

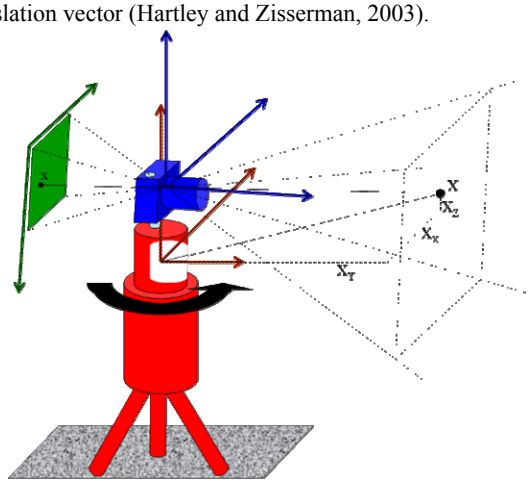

Figure 1. Reference frames of the scanning system with a mounted camera.

The projection matrix defines the image-to-scanner transformation and so allows linking the color content to the 3D laser points. While this transformation results in a loss of image content due to changes in resolution, it allows processing both information sources in a single reference frame and is therefore advantageous.

2.2 Data Representation

3D point clouds are difficult to process due to varying scale within the data, which leads to an uneven distribution of points in 3D space. To alleviate this problem we transform the data into a panoramic data representation. As the angular spacing in the ranging is fixed (defined by system specifications), regularity can be established when the data is transformed into a polar representation (Eq. (2))

$$(x, y, z)^T = (\rho\cos\theta\cos\varphi, \rho\cos\theta\sin\varphi, \rho\sin\theta)^T \qquad (2)$$

with x, y and z the Euclidian coordinates of a point, θ and φ are the latitudinal and longitudinal coordinates of the firing direction respectively, and ρ is the measured range. When transformed, the scan will form a panoramic range image in which ranges are "intensity" measures. Figure 2a shows range data in the form of an image where the x axis represents the φ value, $\varphi \in (0, 2\pi]$, and the y axis represents the θ value, $\theta \in (-\pi/4, \pi/4]$. The range image offers a compact, lossless, representation, but more importantly, makes data manipulations (e.g., derivative computation and convolution-like operations) simpler and easier to perform. Due to the convenience in data processing that this representation offers, all input channels are transformed into it.

2.3 Channel selection

As noted, different cues can be used to segment the data. These should feature attributes that can characterize the different elements of interest or supplement the information derived by other cues. For the segmentation, three cues are introduced. The first is the range content, namely the "intensity" value in the range panorama, the second is the surface normals, and the third

is the true color channel arriving from images acquired by the mounted camera. Notably additional (or different) cues can be formed and added.

Surface normals are computed by

$$\vec{N} = \frac{\vec{\nabla}_1 \times \vec{\nabla}_2}{\left\| \vec{\nabla}_1 \times \vec{\nabla}_2 \right\|} \tag{3}$$

with $\vec{\nabla}_1 = [dX_1, dY_1, dZ_1]^t, \vec{\nabla}_2 = [dX_2, dY_2, dZ_2]^t$. Differences are computed between neighboring pixels in the range panorama. The amplification of noise in the normal computation and the variations in scale across the scan affect the quality of the normal values in different levels, where noisier normals are expected close to the scanner. We reduce the noise effect by applying an adaptive Gaussian smoothing of the data as a function of the range. The physical window size, *D*, is set to a fixed value, which is then translated into an adaptive kernel size as a function of the range and scanner angular resolution Δ. The window size, *d*, in image space is given by Eq (4).

$$d(\rho) \cong \frac{D}{\rho \Delta} \tag{4}$$

The three individual channels can be seen in Figure 2. Figure 2a shows the range channel with the blue color indicating no-return regions that relate both to the sky and to specular points from which no return arrived. Figure 2b shows the normal directions (color coded) that are showing monotonicity on the ground and along the walls while exhibiting variations around trees and other non-flat or faceted objects. The consistency in the normal values is a result of the adaptive smoothing process. Figure 2c shows the projected color points on the range panorama as achieved via ray tracing. We note that due to some inaccuracies in the registration and the resolution of the laser data (compared to the image based one) some tree canopy points receive sky colors. To eliminate these artifacts from the segmentation, sky tones are masked and replaced by the closest darker tone. An alternative approach will segment the individual images in image space and then assemble them through the forward projection. In this setup, the assembly (and handling sky segments) will require treatment.

2.4 Segmentation

The transformation of the data into a panorama allows the use of common image segmentation procedures for segmenting the point-cloud. As a segmentation scheme, we use the Mean-Shift segmentation (Comaniciu and Meer, 2002), a scheme that was chosen due to its successful results with complex and cluttered images. Being a non-parametric model, it requires neither model parameters nor domain knowledge as inputs. The algorithm is controlled by only two dominant parameters: the sizes of spatial and the range dimensions of the kernel. The first affects the spatial neighborhood while the latter affects the permissible variability within the neighborhood. These two parameters are physical in a sense.

Generally, the mean-shift clustering, on which the segmentation process is based, is an iterative procedure, where each data point is "shifted" towards the centroids of it neighboring data points. The new value of the point is set as the mean, c_{j+1}, by

$$c_{j+1} = \frac{\sum_{s \in S(c_j)} w(c_j - s)s}{\sum_{s \in S(c_j)} w(c_j - s)} \tag{5}$$

with $w(\ )$ the weight attached to the vector *s* of the point, and *j* the iteration index number. Convergence is reached when the centroids is no longer updated. The segmentation algorithm itself is based on a derived filtering scheme beginning with feature vectors considered a cluster center. Using the update equation, an iterative convergence process into cluster centers is initialized. The pixel labels are set to the value of convergence. Then, neighboring regions sharing common values, up to the parameter defined for the range, are grouped together into a segment.

The application of the mean shift segmentation on the individual channels is shown in Figure 3. Figure 3a shows the segmentation based on the range, it shows that the patchy results appear in continuous regions where no meaningful separation can be identified. Nonetheless, elements like the tree stems or poles clearly stand out as individual segments. Figure 3b shows the results of the normal based segmentation. Contrary to the range based segmentation, the ground and the façades appear here as complete segments. Notice however the patchiness around unstructured elements as the trees, poles or the fountain in the front of the scene. Finally, Figure 3c shows that the color channel managed capturing some of the façades as complete objects, and vehicles (which are dominant in their color feature) were extracted. Generally, color exhibits sensitivity to illumination conditions and shadows, which can be noticed in the segmentation of the floor, in some of the walls and the fountain. Notice that poles and traffic signs, which are expected to be distinct with respect to their surroundings, were isolated in the color segmentation.

2.5 Integration scheme

When dealing with multi-cue based segmentation as in the present case, the main challenge is handling the different space partitioning of the different channels. As an example, the ground, which ideally would be extracted as a single segment, will have uniform values in the normals channel while having large variations in the distance channel (and also uneven intensity values in the true color channel). Therefore, our aim is not perform a segmentation that concatenates all channels into a single cube and performs the segmentation on the augmented feature vector. Such segmentation will be highly dimensional, computationally inefficient, and ultimately may lead to over-segmentation of the data.

Instead, the integration scheme we follow originates from the realization that the different channels exhibit different properties of the data. Consequently, they will provide "good" segments in some parts of the data and "noisy" ones in other parts. We segment therefore each channel independently (as the results in Figure 3 show) and then construct a segmentation that integrates them, by selecting the better segments from each channel. We note that in this scheme the addition of other channels can be accommodated without many modifications.

The approach we take for the integration assumes that quality segments exist in each channel, and so extracting the highest quality segments from the individual channels has the potential of providing a segmentation that feature the dominant phenomena in the scene, and thereby a meaningful partitioning.

Generally, our objective is to obtain segments that are uniform in their measured property, where optimally, all data units belonging to the segment will have similar attributes. Additionally, we aim for segments that are spatially significant and meaningful. As such, we wish to assemble large group of data units, preferably of significant size in object space. These segments should not lead however to under-segmentation.

In order to meat the need for significant grouping in object space, we set the score of a segment with respect to its 3D coverage. Due to varying scale within the scan the segment size in image-space cannot be represented by the number of pixels. 3D coverage, R, is therefore calculated via

$$R = \Delta \int_{s \in S} \rho(s)\, ds \approx \sum_{s \in S} \rho(s) \qquad (6)$$

The 3D coverage of the segment does not guarantee it correctness. As an example, it may happen that meaningless strips will be extracted in the range channel (see Figure 3a). In order to reduce the appearance and the influence wrong segments, we enforce uniformity standards that relate to the measured property. In the present case this variability is modeled using a preset threshold values on the within-segment dispersion.

The proposed model is applied as follows. First, the largest segment is selected from all channels, if the segment quality is satisfactory it is inserted into the integrated segmentation. All pixels relating to this segment are then subtracted from all channels and the isolated regions in the other channels are then regrouped and their attribute value is computed. Following, is the extraction of the next largest segment and the repetition of the process until reaching a segment whose size is smaller than a prescribed value and/or preset number of iterations. We note that due to the non-parametric nature of the mean-shift segmentation, re-segmenting the data between iterations has little effect.

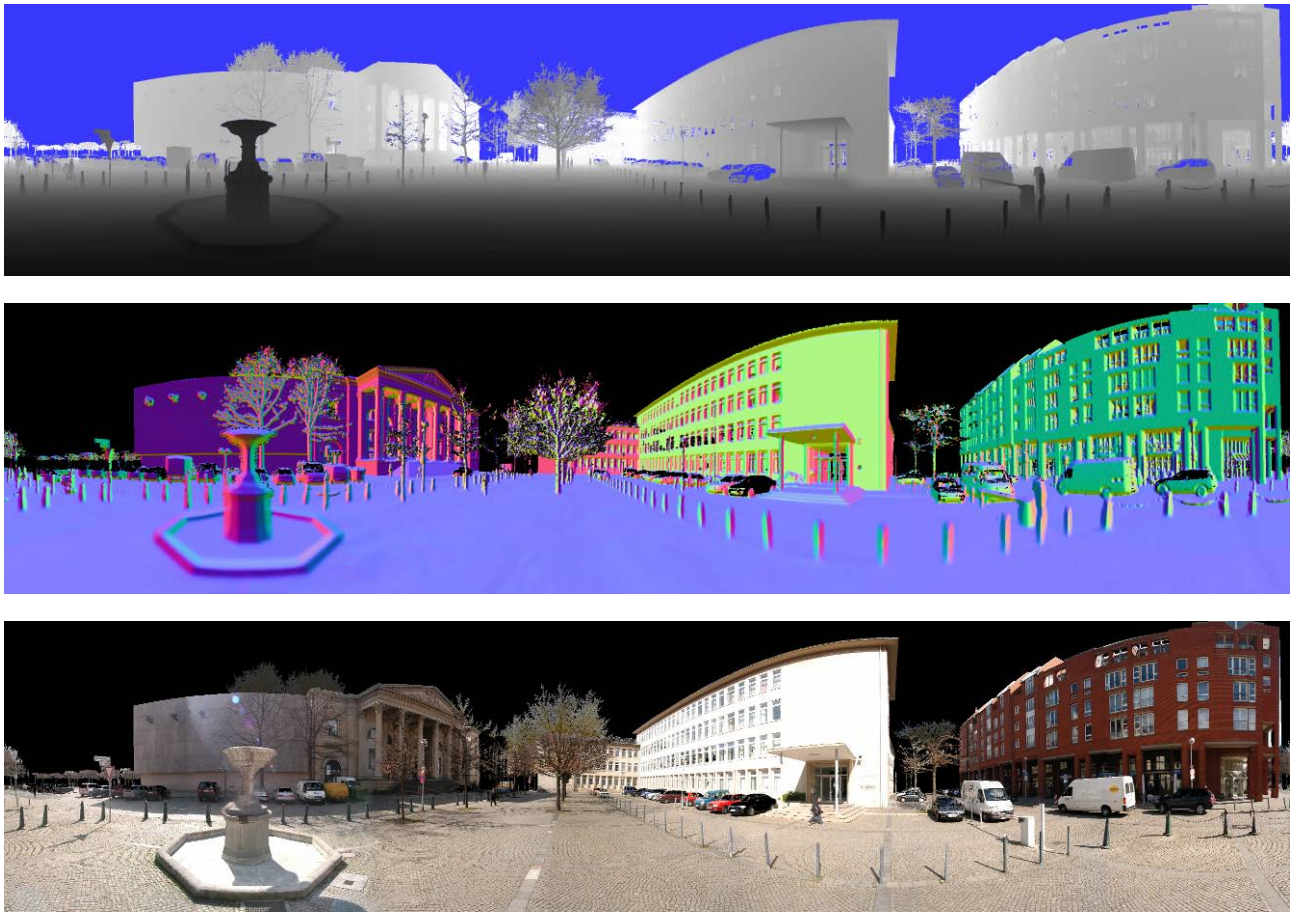

Figure 2. Polar representation of the individual cues used for the segmentation. The horizontal and vertical axes of the images represent the values of φ, θ respectively. (top) intensity values as distances ρ (bright=far), "no-return" and "no-reflectance" pixels are marked in blue, (middle) surface normals represented in different colored by their value, (bottom) color content as projected to the scanner system (see text).

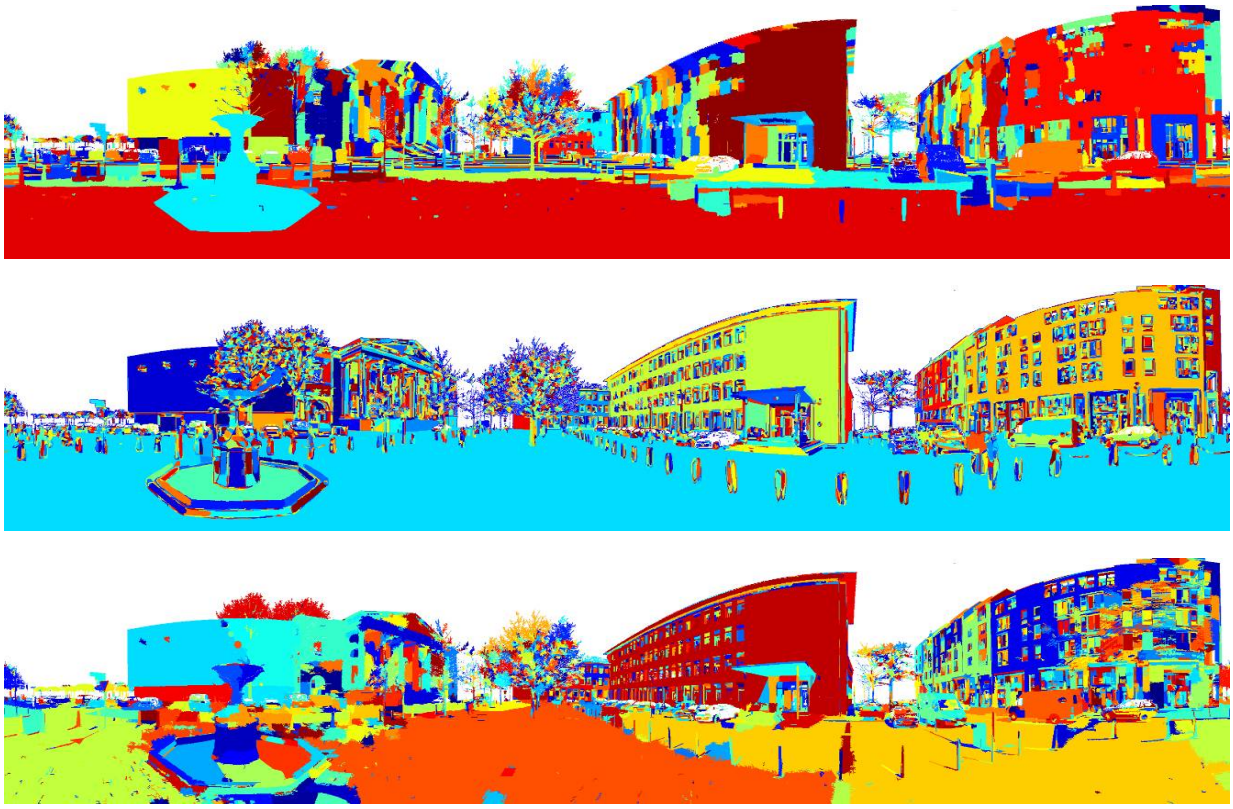

Figure 3. Results of the data segmentation using the mean-shift algorithm. (top) segmentation of the distance channel, (middle) segmentation of the surface normals channel, (bottom) segmentation of the true-color channel.

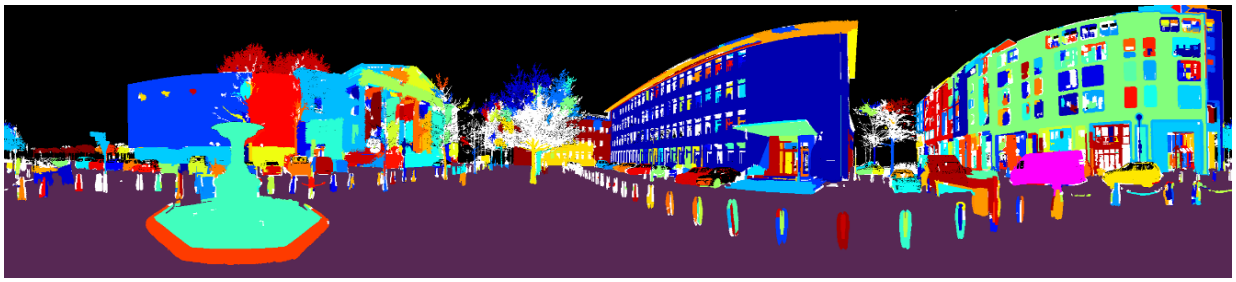

Figure 4. Results of the augmented segmentation based on the integration of the three channels.

3. RESULTS AND DISCUSSION

The integration of the segments into an augmented segmentation is presented in Figure 4. As the Figure shows the more dominant segments form the segmentation. The largest segment that was selected is the ground segment as derived from the normals channel. The second segment is blue façade (center building) which was extracted from color channel. Its extraction from that channel has to do with the strong and almost uniform intensity along it. The third segment is the green façade which was derived, again, from the normals segmentation. It is of interest to note that the third largest segment arrived from the range channel (relating also to that façade), but was discarded due to the variability. The fourth largest segment is the water fountain in the front of the scan. Notice how it did not exhibit any clear structure in neither the normals nor the color channels, and was extracted as an individual segment from the range channel. In relation to the fountain we also note that its border which was not segmented as a unique entity in any of the segmentations was extracted as a single entity due to the subtraction applied on the extracted segments from the individual channels. Additional noteworthy elements that were extracted by the augmented segmentation are vehicles which mostly were extracted as complete entities, and pole elements like the traffic sign at the left of the scan, or the streetlamp close to the building façade on the left.

The halos surrounding the poles at the center of the scan are ground points that were not segmented as such, as they were lying next to edge points (transition between objects). They were correctly separated from the pole in the range channel. Similar edge effects can be noticed in other parts like the cornice surrounding the blue façade building. Another feature of the segmentation is the partitioning of entities due to occlusion. This can be noticed in the wall on the leftmost building that was partitioned into three large segments. This partitioning is due to objects in front of it which hide parts of it.

While the overall segmentation features improvement to the basic ones, both features suggest that the segmentation along the proposed scheme can be further pursued in several directions. These include inclusion of additional cues that may feature other elements characterizing natural scenes, introducing merging schemes for segments based on connectivity and forms of similarity, and the analysis of occlusions as a means merge disconnected segments that belong in fact to the same object.

4. CONCLUDING REMARKS

The paper proposed an approach for the segmentation of terrestrial laser point clouds while assembling and integrating different data sources. The proposed model offers a general framework in the sense that it can utilize different features and can be customized according to application requirements. Overall, the results show that integration of different cues and information sources into a laser scanning segmentation has managed providing improved results in relation to each of the individual channels.

The model demonstrated that using an intuitive scheme for selecting the best segments from different segmentation maps provides satisfactory results. The solution for weighting the importance of different cues to the overall segmentation is modeled as a crisp decision favoring dominant segments in object space as long as they do not violate preset rules. Future work in this regards will pursue alternative weighting schemes for the data arriving from the individual channels.

ACKNWOLEDGEMENT

The authors would like to thank Dr. Claus Brenner for making the data used for our tests available.

REFERENCES

Alpert S., Galun M., Basri R., Brandt A., 2007. Image segmentation by probabilistic bottom-up aggregation and cue integration. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, June 2007.

Bleyer, M., and Gelautz, M., 2004. A layered stereo algorithm using image segmentation and global visibility constraints. in Proc. of ICIP 2004 **5**: 2997-3000.

Brandt A., 1986. Algebraic multigrid theory: the symmetric case. Appl. Math. Comput, **19** pp. 23–56.

Carson, C. Belongie, S. Greenspan, H. Malik, J. 2002. Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(8): 1026- 1038.

Coiras, E., Santamaría, J., Miravet, C., 2000. Segment-based registration technique for visual-infrared images. *Optical Engineering* **39**(1) pp. 282-289.

Comaniciu D., Meer. P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE transactions on PAMI, 24:603–19.

Dold, C., Brenner, C., 2006. Registration of Terrestrial Laser Scanning Data using Planar Patches and Image Data, in: H.-G. Maas, D. Schneider (Eds.), ISPRS Comm. V, IAPRS Vol. XXXVI Part. 5 pp. 78-83.

Estrada F. J., Jepson A. D., 2005. Quantitative Evaluation of a Novel Image Segmentation Algorithm. CVPR (2) 2005: 1132-1139.

Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient Graph-Based Image Segmentation, *International journal of computer vision* 59(2) pp. 167-181.

Gorte B. 2007. Planar feature extraction in terrestrial laser scans using gradient based range image segmentation, ISPRS Workshop on Laser Scanning, pp. 173-177.

Hartley R., Zisserman A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Second Edition.

Huang, Q., Wen G., Wenjian C. 2005 Thresholding technique with adaptive window selection for uneven lighting image. *Pattern Recognition Letters* **26** pp. 801–808.

Klaus A., Sormann M., and Karner K., 2006 Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. in Proc. of ICPR 2006, pp. 15-18.

Mian, A., Bennamoun, M., Owens, R., 2006. Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes. IEEE transactions on PAMI. **28**(10), 1584-1601.

Otsu N. 1979. A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. **9**: 62-66.

Pal N. R. and Pal S. K., 1993. "A review on image segmentation techniques," *Pattern Recognition*, **26**(9): 1277-1294, 1993.

Rabanni T., 2006. Automatic reconstruction of Industrial Installations using point clouds and images. PhD thesis. NCG, publication on Geodesy 62.

Roth, V., Ommer, B., 2006. Exploiting Low-Level Image Segmentation for Object Recognition, DAGM06 pp. 11-20.

Russell, C. B., Efros A., Sivic J., Freeman T. W., Zisserman A., 2006. Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. in Proc. of CVPR 2006 **2**, 1605-1614.

Sharon E., Brandt A., Basri R., 2000. Fast Multiscale Image Segmentation. CVPR 2000: 1070-1077.

Shi J., Malik J., 2000. Normalized Cuts and Image Segmentation. IEEE transactions on PAMI. **22**(8): 888-905.

Vosselman G., S. Dijkman, 2001. 3D Building model reconstruction from point clouds and ground plans. IAPRS **34**(3/W4).