

Façade Segmentation in a Multi-View Scenario

Michal Recky, Andreas Wendel, Franz Leberl

Institute for Computer Graphics and Vision
Graz University of Technology
Graz, Austria
{recky, wendel, leberl}@icg.tugraz.at

Abstract— We examine a new method of façade segmentation in a multi-view scenario. A set of overlapping, thus redundant street-side images exists and each image shows multiple buildings. A semantic segmentation identifies primary areas in the image such as sky, ground, vegetation, and façade. Subsequently, repeated patterns are detected in image segments previous labeled as “façade areas” and are applied to separate specific facades from each other. Experimentation is based on an industrial street-view dataset from a moving car by well-designed, calibrated, automated cameras. High overlap images define a multi-view scenario. We achieve 97% pixel-wise segmentation effectiveness, outperforming current state-of-the-art methods.

Keywords-context; semantic segmentation; multi-view

I. INTRODUCTION

Our work aims on providing useful data for building reconstruction and the interpretation of scenes in support of the establishment of an Internet-hosted Exabyte 3D World model [7]. 3D is becoming of interest in geospatial Internet mapping platforms like Microsoft Bing Maps, or Google Maps. At the same time, the demand for improved levels of detail, the quality of visualization and the introduction of novel applications lead to the need to interpret urban scenes. Such interpreted data cannot only be stored in a compact form and thereby reduce the amount of data needed to be transferred through internet connections, but also represent the opportunity for users to build applications that would not be possible if the Internet maps were merely images.

Aerial image processing is typically used to extract 3D frame models of buildings [18]. An increasing “appetite for detail” at the human scale of a world model has caused the emergence of street-level photography and other sensor data for an immersive experience of the Internet user.

Recently, automated 3D building modeling and façade interpretation has become a research topic [9][11][17]. Previous work of Recky and Leberl [11] addressed façade analysis based on gradient projection. Elements of a façade such as windows, entrances, arches or columns were identified. Façade textures can be replaced by parameters such as shape grammars [9].

However, methods of 3D building modeling require that single, geometrically rectified building façades are known to serve as an input. In general, street-side images present complex urban scenes and a building landscape with many

façades. In this paper, we introduce a robust approach to solve this problem and thereby are increasing the likelihood of success of 3D modeling from street-side images. In the most basic case, we consider a single street-side image as an input. First, we label principal areas into classes, such as sky, vegetation, ground, buildings. In this process we make use of the natural geometry of urban scenes and employ the context between classes. Subsequently, we proceed with the segmentation of building fronts into separate façades based on the detection of repetitive patterns built by windows and architectural styles. This segmentation subsequently can be used to interpret the details in each façade.

In a typical case, the single image of an urban scene is augmented by many additional overlapping images. We find that street-level images represent large amounts of redundant visual data. Therefore, we extend our method into a multi-view scenario, where a redundant dataset with overlapping street-level images exists. The effect of multi-view imagery on various geometric scenes has been discussed previously, for example in [3]. However, it is less well understood how the semantic interpretation of a scene is affected by the transition from a single image to a multi-view image stack. Our work addresses a test dataset taken from a moving car designed for the visual mapping of urban environments. We selected a challenging subset of images taken from a frontal-sideway camera. In this subset, most facades are under perspective distortion, testing the limits of detecting repeated patterns and of segmentation. The dataset is complemented with laser scanner data to help in matching overlapping images within pixel accuracy.

II. RELATED WORK

Our approach is based on two methods. First we apply the semantic segmentation method of [10] to extract contextual information from the image. We use this information as prior knowledge for façade segmentation [16]. We thus combine results from a previously described façade segmentation with a context-based image segmentation and achieve an improvement of the segmentation towards 97% effectiveness.

A. Semantic Segmentation based on Context

In this section, we describe an approach for semantic segmentation of a street-side image. This method labels every pixel in the image into classes – sky, cloud, roof,

vegetation, building, ground, shadow and undecided. The semantic segmentation we use is described in [10]. It is based on the idea of Hoiem [5], but works with segments instead of superpixels and uses the spatial and contextual information to improve the results.

Labeling segments in lieu of superpixels is introduced to better consider spatial context. The context in the image has to be naturally examined for the real objects (i.e., the context between the roof and the building façade, see Fig. 1). Superpixels often do not represent real objects in the scene; therefore the spatial context at the level of superpixels is not a valid concept. In our work, we try to locate the real objects in the scene as segments and subsequently, we examine the context between them.

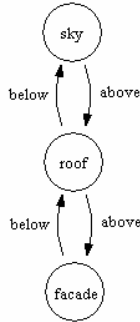


Figure 1. An example of geometric context between object classes.

The first step is image segmentation. Small patches get identified in a watershed segmentation. We iteratively merge patches into larger segments. The merging is based on color and texture differences between neighboring patches. Ground-truth data of urban scenes is used to train a merging algorithm. This allows to achieve the required segmentation results, i.e., the merged segments approximate real objects. After this step, a graph is constructed for the image. Each segment is considered a node in the graph and the edges are placed between neighboring regions (see Fig. 2).

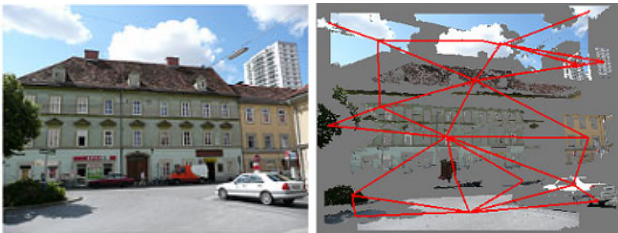


Figure 2. Graph-structure placed over the segmented image. Segments are represented as graph nodes. An edge is placed between segments that are mutual neighbors [10].

In each node we have a set of descriptors for visual information of the corresponding segment. The edges of the graph define the contextual relationship for subsequent examinations.

To examine the context between segments, we employ Discriminative Random Fields (DRF) [6]. Random fields have been used for context representation previously [5][8]. The most commonly used models in vision are Markov

Random Fields (MRF). Given the observed data $\mathbf{y} = \{y_i\}_{i \in S}$ from the image, and corresponding labels $\mathbf{x} = \{x_i\}_{i \in S}$, where S is the set of sites, the posterior distribution over labels for MRFs can be written as

$$P(\mathbf{x} | \mathbf{y}) = \frac{1}{Z_m} \exp \left(\sum_{i \in S} \log p(y_i | x_i) + \sum_{i \in S} \sum_{j \in N_i} \beta_m x_i x_j \right), \quad (1)$$

where Z_m is the normalizing constant, β_m is the interaction parameter of the MRF and N_i is the set of neighbors of site i .

The pairwise term $\beta_m x_i x_j$ in MRFs can be seen as a smoothing factor, given the input data \mathbf{y} . This makes the MRF applicable mainly for simpler forms of context. In our work we use the DRF to overcome this limitation. DRFs are a discriminative model designed to directly model distribution over labels, using discriminative classifiers in both unary and pairwise terms. The principal probability formula of DRFs can be described as

$$P(\mathbf{x} | \mathbf{y}) = \frac{1}{Z} \exp \left(\sum_{i \in S} A_i(x_i, \mathbf{y}) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, \mathbf{y}) \right), \quad (2)$$

where Z is the normalizing constant, $-A_i$ is the unary and $-I_{ij}$ the pairwise potential. The pairwise term

$$I_{ij}(x_i, x_j, \mathbf{y}) = \beta (K x_i x_j + (1 - K) (2 \sigma(x_i x_j \mathbf{v}^T \boldsymbol{\mu}_{ij}(\mathbf{y})) - 1)), \quad (3)$$

where $0 \leq K \leq 1$ and \mathbf{v} are the model parameters, can be seen as an extension of Markov's pairwise term (when $K = 1$ in DRF, pairwise terms are identical). However, this formulation allows us to apply more complex context as a feature vector $\boldsymbol{\mu}_{ij}(\mathbf{y})$. To represent context, we examine spatial relations between segments as listed in Table I.

TABLE I. SPATIAL RELATIONS ARE BASED ON THE RELATIONS BETWEEN BOUNDING BOXES AND CENTERS OF ANY TWO REGIONS [10].

Site i	Relation	Site j
Bounding box	Inside Enveloping Partially above/below Fully above/below	Bounding box
Region center	Inside Above Below	Bounding box
Region center	Above Below	Region center

Another advantage of the DRF is that we can use a discriminative classifier as a unary term $A_i(x_i, \mathbf{y})$ [6]. In our case, we can use visual features (histogram of HSV values, histogram of gradients, region covariance) of the segments to train classifiers and apply them directly in the DRF framework.

Recky and Leberl have shown in [10] that this approach can improve classification mainly for those areas that have strong spatial relations, for example the pair of façade - roof. For the classification of building façades the achieved success rate was 94%, thus having this portion of all façade pixels actually identified as belonging to a façade. Examples of semantic segmentation can be observed in Fig. 3.



Figure 3. Original images (left) and semantic segmentation (right) [10]. Each class is represented by a color: building (dark green), sky (blue), roof (brown), vegetation (green), ground (grey), clouds (white). Black areas are considered unidentified.

B. Detection of Repetitive Patterns

We build on the method of Wendel et al. [16] for finding repetitive patterns within a single image. First, the Harris corners [2] are detected as interest points. Subsequently, color intensity profiles are extracted on lines between each of them. The resulting complete graph structure has interest points in nodes and each edge corresponds to a color profile. We limit the complexity of the graph by connecting only the 30 closest neighbors to each node.

The color profiles are constructed as 60-dimensional normalized descriptor: for every color channel of RGB, we compute a 20-dimensional descriptor sampled by linear interpolation along the line. In this approach, the scale invariance is achieved by setting the number of coefficients for interpolation as a constant. For an example of intensity profiles, see Fig. 4.

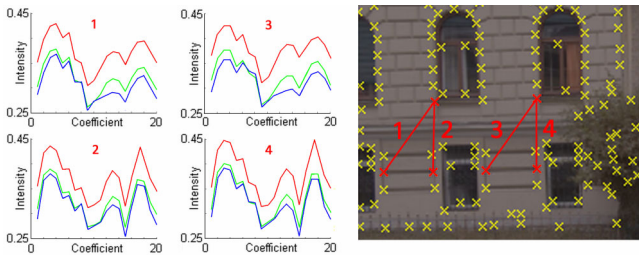


Figure 4. We describe the image content between Harris corners by extracting intensity profiles with 20 values for every RGB color channel [16].

For matching of the descriptors, we use a kd-tree method [1]. The tolerated threshold was set to 5% deviation off the descriptor prototype for finding repetitive patterns. Matches with more than 10 descriptors involved are ignored, as they showed to be insufficiently discriminative. To achieve the required robustness, an additional voting step has been included in the process. All matching profiles vote for their corresponding interest points. The vote is counted only if the descriptor has not contributed to a correspondence so far. This method is described in [14][15] and removes the bias in a voting matrix.

We finally locate the repetitive patterns in the voting matrix by thresholding the votes a correspondence received. Interest point correspondences are established if at least 3 of 30 intensity profiles were matched. An example for arbitrarily shaped repetitive image areas obtained by this voting process can be observed in Fig. 5(a). We eliminate outliers by the assumption that repetitive patterns are not likely to occur across an entire image, and are not likely to be very small. Therefore, we restrict the horizontal and vertical distances of accepted matches. The result of this approach can be observed in Fig. 5(b) as color-coded lines.

C. Façade Segmentation

Several methods for processing a single façade have been introduced recently. However, façade separation or segmentation itself has not received much attention. Mueller et al. [9] introduced an algorithm which is able to summarize redundant parts of a façade and thus subdivide images into floors and tiles. A major limitation is the dependency on single façade images, and automatic processing fails for scenarios with blurry texture, low contrast, chaotic ground floors, and occlusions caused by vegetation. Other works on façade separation [4][17] are based on the evaluation of directional gradients, which did not prove to be robust for our dataset because it only works for highly regular façades.

The previous section summarized a method for detecting areas of repetitive patterns. Due to the natural setting of objects in street-side images, we assume that the repetitive patterns are located along the horizontal direction and separations between façades occur in a vertical direction. Subsequently, we project lines between matched interest points into the horizontal axis, constructing the histogram of match counts (see Fig. 5(c)). We compute the repetition likelihood as a percentage of all matches in a given interval of the histogram.

The next step is the detection of separation “areas”, as an extended interval between repetitive areas (marking the positions where one façade ends and another starts). Computing the separation areas from minima on the repetition likelihood is not sufficient, as the global minimum does not account for images with more than two connected façades and local minima can be detected in common false positive cases, such as between rows of windows. If all parts of the façade contribute equally, we would get a uniform repetition likelihood. Setting this value as a threshold, areas

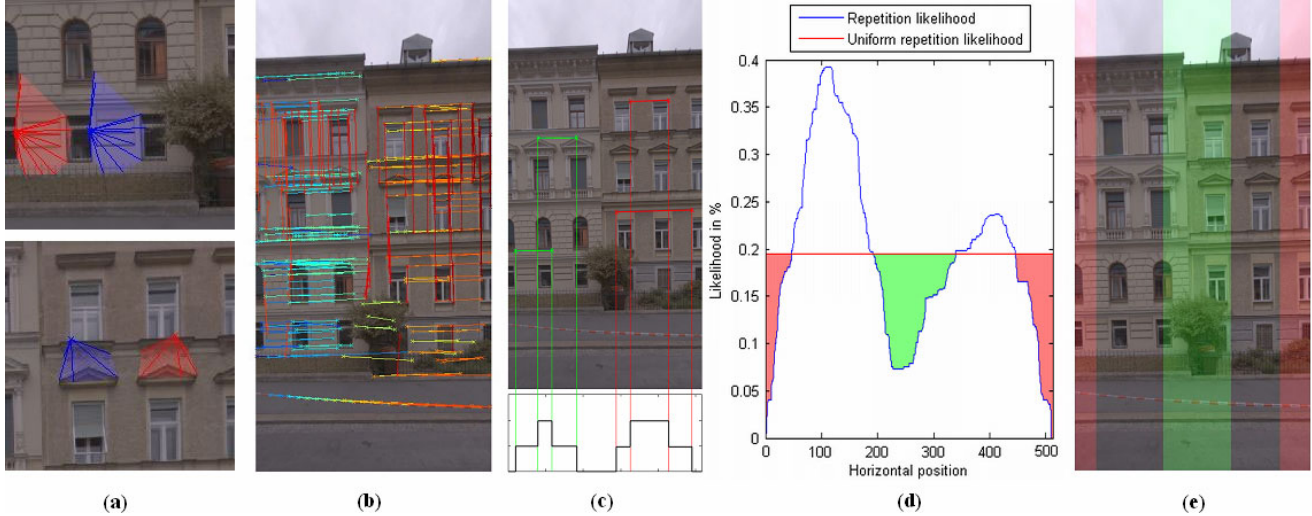


Figure 5. From streetside data to separation (best viewed in color) [16]: (a) Matching of arbitrary areas (b) Detected repetitive patterns (color-coded lines) (c) Projection results in a match count along the horizontal axis (d) Thresholding the repetition likelihood with the uniform repetition likelihood (e) Resulting repetitive areas, separation areas (green), and unknown areas (red).

with low likelihood are defined as separation areas and areas with higher likelihood as repetitive areas (see Fig. 5(d)).

To cope with narrow fields of view, where the location of repetitive areas is not detected, we define the areas on borders of an image as “unknown”. These areas start at the image boundary and end at the first repetitive area. An example of repetitive, separation and unknown areas can be seen in Fig. 5(e).

III. FAÇADE IDENTIFICATION IN A SINGLE IMAGE

In this section, we enhance the previously presented method of façade separation using prior knowledge. Given the semantic segmentation described in Section II(A), we consider all areas in the image labelled as building or unidentified to potentially be part of repetitive areas, defining a separate façade. Subsequently, we apply the repetitive pattern detection only to those pixels. The resulting set of Harris corners gives us a better basis for interest point matching (see Fig. 6).

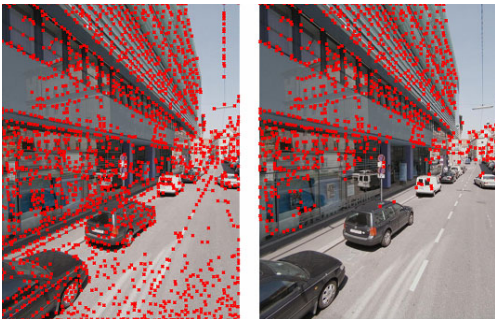


Figure 6. Set of Harris corners without prior knowledge (left) and with prior knowledge from semantic segmentation (right).

Each repetitive pattern area gives us a unique façade. To detect the entire façade area, we use the results of the segmentation described in Section II. As described, the

segmentation was designed to detect natural objects in the scene such as a building façade. Therefore, façades are usually represented as one or a small number of large segments, labelled into a façade class. Subsequently, we compute the ratio of pixels of segments which belong into a repetitive pattern area to the number of pixels outside of that area. If the ratio is larger than one, segments are labelled as separate façade (see Fig. 7).

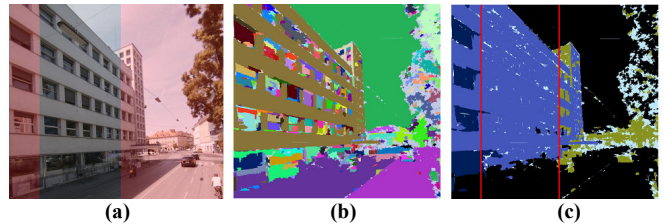


Figure 7. Façade identification in a single image: (a) Original image with repetitive area (b) Segmentation of the image (notice the façade segmented as one, brown segment) (c) Identified façade (original repetitive area is marked by red lines).

It is a matter of definition if objects like windows, doors or shop signs are considered part of the façade or separate objects. In the semantic segmentation, these objects are segmented separately and as there is no class for them, they are usually labeled as unidentified. However, in our ground truth, only the border of a façade is labeled, so these objects are included in the definition of a façade. To cope with this problem, we consider all unidentified segments as façade segments and we include them in the evaluation by repetitive areas. This solution sometimes gives unwanted results, as other objects, like cars, or pedestrians originally labeled “unidentified” are often included into façades. This problem will be solved by extending the set of classes in semantic segmentation, which is the subject of ongoing work.

IV. MULTI-VIEW SCENARIO

Automatic processing of real world digital image data continues to present challenges for researchers and many computer vision tasks are considered unsolved. “Hard problems” may become more tractable if one generalizes the input data to consist not of a single image, but of a stack of multiple images. We can denote this as “redundant” or “multi-view” input data. Therefore in our image databases, we usually want to employ multiple images of any given scene. Each of these images, when processed individually, will provide us with a specific interpretation. In this section, we examine how multiple interpretations from multiple images differ and complement one another to improve the overall result.

In our database (denoted as an “Industrial System dataset”), images are taken by a calibrated multi-camera apparatus mounted on a car (see Fig. 8).



Figure 8. An example of a camera system mounted on a car. It is designed to cover a wide viewing range.

This setup creates overlapping images with a rigorous and calibrated geometry from a single image-taking position, and delivers for each object point multiple images from that single sensor position. By moving the car and repeating the image collection, the level of redundancy is increased. Carrying along a scanning laser arrangement with the imaging sensors provides one with additional range information and the means to directly match the overlapping images without a need for an image-based point cloud. Fig. 9 is an example of a data set that consists of 250 images from each camera on the car platform. In our work, we use the input of only one, frontal-sideways tilted camera. The data base supports investigations into the issues of the types of redundancies, namely multiple images, all taken with parallel optical axes from different camera positions [12].

In the dataset, we have identified 9 separate building façades, each located in approximately 20-50 images. 7 of the façades are under severe perspective distortion, which present the worst case scenario for a repeated pattern algorithm (as the repetition is less evident in the perspective) and provides a challenge for segmentation.

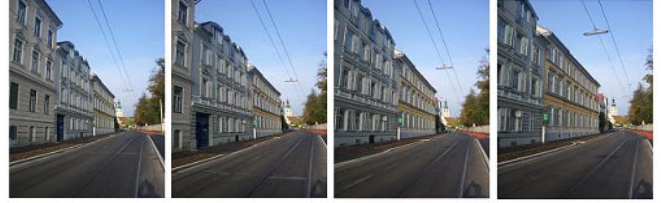


Figure 9. Example for images from the Industrial System database with the parallel optical axes providing a high level of redundancy from sequential exposures in a moving vehicle. In this example the optical axes are pointing halfway forward.

A. Image Matching

Matching the images of the stack is based on LiDAR laser scanner data, provided as a supplement to the Industrial System dataset. In each position of a car, a set of LiDAR points has been measured. Given the geo-position of the camera system, a set of global 3D points of the scene is available from the LiDAR in the same coordinate system. The projection of these 3D points into individual images gives us an instrument for superimposing the separate images. In this case, the set of LiDAR points can be considered an equivalent to an image-based sparse 3D point cloud.

As the sparse distribution of points does not provide us with enough data to match pixel-by-pixel, we proceed with matching segments. The precision of matching depends on the number of LiDAR points located in the segment. In the worst case scenario, there are no LiDAR points that can be projected into a segment. In this case, we use the assumption of planarity of the façade. As most of the façades are nearly planar objects (or are mostly composed of planar regions), we consider the interpolation between points to provide a valid approximation with sufficient accuracy. In the case of no LiDAR points projected into a segment, we find the two closest points and create a new point by interpolation. Considering the planarity, the corresponding interpolation between 3D coordinates of such points will provide the means of matching. This solution is necessary only for small segments, as major ones have usually a large number of LiDAR points projected into them.

B. Labeling of Segments in a Multi-View Scenario

Given the set of LiDAR points projected into a segment, we have the means of matching a segment to the points in other images. When the façade identification has been performed in these images, as described in Section III, we can gather data from multiple images about the labeling of segments as part of separate façades. Subsequently, we can make the decision for the segment based on the following criteria:

- If the segment was classified as a *building* in the semantic segmentation and was labeled as a part of separate façade *in at least one image*, we consider the segment to be part of that façade.

- If the segment was classified as *unidentified* in the semantic segmentation and was labeled as part of a separate façade in the majority of images, we consider it to be a part of that façade. Segment classification can then be projected from corresponding images.

This distinction between façade class and unidentified class removes most of the problems with the labeling of unidentified objects like cars or pedestrians as façades (described in Section III), but still keeps windows and doors as parts of a façade.

V. RESULTS

Our evaluation procedure is similar to that of Wendel et al. [16] to ensure compatibility. We use precision and recall [13] to evaluate our algorithm and combine these by a harmonic mean to obtain the measure of effectiveness, also called F_1 -measure. We obtain ground truth by manually labeling individual façades. We estimate the point matching quality by clustering the matches and assigning them to the ground truth, resulting in a set of inliers and outliers for every segment. We only use the match precision (PR_{match}), as it is not possible to estimate the amount of false negatives required to compute the recall. Facade separation quality (Separation F_1) is estimated by checking if the detected repetitive area lies within the ground truth segment. More or fewer splits lower the effectiveness except when they occur in an unknown area. The facade segmentation quality (Segmentation F_1) is estimated using a pixel-wise comparison between the automatically obtained segment and the ground-truth segment. Our approach depends on the parameters of the Harris corner detector: $\sigma_D = 0.7$, $\sigma = 3.0$. All other parameters are defined with respect to the image scale.

In our first experiment, we compare the separation precision using the semantic segmentation as a prior knowledge to the original algorithm proposed by Wendel et al. [16] (see Table II).

TABLE II. SEPARATION RESULT WITHOUT (ORIGINAL) AND WITH THE PRIOR KNOWLEDGE FROM SEMANTIC SEGMENTATION (MASK).

	Original	Mask
Matching precision (%)	43,7	45,5
Separation precision (%)	83,6	96,1
Separation recall (%)	75,3	85,3
Separation F_1 (%)	79,2	90,4

In the next experiment, we compare the results of façade segmentation. We tested three different approaches, namely the segmentation method of Wendel et al. as proposed in [16] without prior knowledge, Wendel’s segmentation with the prior knowledge, and finally the approach described in this paper – repetitive areas acquired with prior knowledge applied on context based segmentation (see Table III).

In these results, we can observe a significant improvement with the application of prior knowledge (more than 15%). The segmentation with prior knowledge (Single) has performed less effectively in segmentation precision, but outperformed the original approach in recall, resulting in better overall performance (by 3%).

TABLE III. SEGMENTATION RESULTS FOR THREE DIFFERENT APPROACHES. WENDEL’S SEGMENTATION WITHOUT PRIOR KNOWLEDGE (ORIGINAL), WITH PRIOR KNOWLEDGE (MASK) AND THE SEGMENTATION DESCRIBED IN THIS PAPER (SINGLE).

	Original	Mask	Single
Segmentation precision (%)	53,2	91,3	90,2
Segmentation recall (%)	79,7	69,5	74,6
Segmentation F_1 (%)	63,8	78,9	81,7

Our final experiment is focused on the difference between the single-view and multi-view scenarios. In both approaches, we use the prior knowledge from semantic segmentation to extract repetitive areas. The segmentation method for both scenarios is the one described in this paper. In this test, we also examine the effect of segmentation post-processing, using morphological operators. As there were some gaps within the segmentation due to patches not completely merged into the segments (see Section II(A)), or mislabeled areas of the façade, we applied the morphological closing and opening to close the gaps and avoid false negatives (see Table IV).

TABLE IV. SEGMENTATION RESULT IN THE SINGLE-VIEW AND THE MULTI-VIEW SCENARIO. IN BOTH SCENARIOS, WE TESTED THE EFFECT OF MORPHOLOGICAL OPERATORS WITH KERNEL DIAMETER OF 3 (SMALL) AND KERNEL DIAMETER OF 30 (LARGE).

	Single (small)	Single (large)	Multi (small)	Multi (large)
Segmentation precision (%)	92,5	90,2	96,9	97,1
Segmentation recall (%)	68,2	74,6	92,3	96,2
Segmentation F_1 (%)	78,5	81,7	94,6	96,6

Results show that given the laser scanner 3D point cloud, or another robust image matching method, the transition from single-view to multi-view can improve the output by approximately 15%. Also, with the segmentation described in this paper, we can achieve better results with larger kernels of morphological operators.

An example for the separation and segmentation of façades is presented in Fig. 10 for a single-view scenario and in Fig. 11 for a multi-view scenario (both with large kernels). The visual comparison shows that the multi-view scenario does not only provide more complete façade segments in terms of coverage, it also enables our algorithm to detect more separate façades in a single image. For better visualization, videos are provided online¹.

¹ <http://www.icg.tugraz.at/Members/wendel/>



Figure 10. Separation and segmentation results for different facades in a *single-view scenario* (large kernel setting). While separate facades are found in general, the algorithm in the single-view scenario does not always provide complete façade segments and is not able to detect as many separate facades as in the multi-view scenario.



Separation and segmentation results for different facades in a *multi-view scenario* (large kernel setting). Even for facades under perspective distortion, the multi-view scenario provides robust results.

VI. CONCLUSION

In this paper, we present a method for identifying separate façades from a single image and in multi-view scenarios. This work is a bridge between street-side digital data and procedural modeling of façades [9][11]. In a first step, we use a semantic segmentation as prior knowledge for façade separation, further improving the state-of-the-art. Subsequently, we combine the façade separation with segmentation results to improve the identification of separate façades.

Our second contribution in this paper is the application of the method in a multi-view scenario. A majority of computer vision methods are constructed and tested for single images, yet the presence of multi-views is often the current reality. Therefore, we investigate the transition from single-views into multi-views and the challenges that come with it. We propose a matching method based on LiDAR data and the method for façade labeling in multi-views.

For experimentation we rely on a challenging dataset of 250 images containing a variety of objects from historical buildings to modern architecture. Our results show significant improvements when the prior knowledge from semantic segmentation is applied and when the single-views are replaced by multi-views.

ACKNOWLEDGMENT

This work has been supported by the Austrian Research Promotion Agency (FFG) project FIT-IT CityFit (815971/14472-GLE/ROD) and by the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209.

REFERENCES

- [1] J.H. Friedman, J.L. Bentley, R.A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software* (3) 209-226, 1977
- [2] C. Harris, M. Stephens. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*. Volume 15. 1988
- [3] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*. 2ed, Cambridge University Press, ISBN: 052-154-0518, 2004
- [4] J. Hernandez, B. Marcotegui. Morphological segmentation of building façade images. In *Proceedings of ICIP*, 2009
- [5] D. Hoeim, A. A. Efros, M. Herbert, Geometric context from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 654–661, 2005
- [6] S. Kumar, M. Herbert, Discriminative random fields. *International Journal of Computer Vision*, 68(2), pp. 179–201, 2006

- [7] F. Leberl, M. Gruber. 3d-Models of the Human Habitat for the Internet. In Proceedings of Visigrapp, Lisbon, pp.7-15, 2009
- [8] Li, Z. Stan. Markov Random Field Modeling in Image Analysis. ISBN: 978-1-84800-278-4, 2009
- [9] P. Mueller, G. Zeng, P. Wonka, L.V. Gool. Image-based procedural modeling of facades. ACM Transactions on Graphics 26(3), 2007
- [10] M. Recky, F. Leberl, Semantic Segmentation of Street-Side Images. In Proceedings of the Annual OAGM Workshop. Austrian Computer Society in OCG, pp. 271–282, 2009
- [11] M. Recky, F. Leberl. Windows Detection Using K-means in CIE-Lab Color Space. In Proceedings of the International Conference of Pattern Recognition. Istanbul, Turkey, pp. 356-360, 2010
- [12] M. Recky, F. Leberl. Multi-View Image Interpretation for Street-Side Scenes. CGIM, Innsbruck, Austria, pp 48-54, 2010
- [13] C.J.V. Rijsbergen. Information retrieval. Butterworth-Heinemann Newton, USA, 1979
- [14] D. Tell, S. Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In Proceedings of ECCV, pp. 814-828, 2000
- [15] D. Tell, S. Carlsson. Combining appearance and topology for wide baseline matching. In Proceedings of ECCV, pp. 68-81, 2002
- [16] A. Wendel, M. Donoser, H. Bischof. Unsupervised facade segmentation using repetitive patterns. In Proceedings of the 32nd Annual Symposium of the German Association for Pattern Recognition (DAGM'10), LNCS 6376, pp. 51-60, Springer, 2010
- [17] J. Xiao, T. Fang, P. Zhao, M. Lhuillier, L. Quan. Image-based street-side city modeling. ACM Transactions on Graphics 28(5), 2009
- [18] L. Zebedin, A. Klaus, B. Gruber-Geymayer, K. Karner. Towards 3d map generation from digital aerial images. Journal of Photogrammetry and Remote Sensing 60(6), pp. 413-427, 2006