WINDOW DETECTION IN COMPLEX FACADES

Michal Recky, Franz Leberl

Institute for Computer Graphics and Vision, TU Graz

ABSTRACT

The complexity of a building façade provides challenge for window detection algorithms. Especially in the gradient projection approach, the presence of gradients outside window areas significantly reduces the quality of results. We present a modified gradient projection method robust enough to process complex façades of historical buildings. In a single image scenario, this method is able to provide results even for facades under severe perspective distortion. Our algorithm is able to detect many different window types and does not require a learning step. In this paper, we also extend this method into a multi-view scenario. We examine the results of the method in multi-view and evaluate its benefits.

Index Terms— Windows Detection, Multi-view, Gradient Projection

1. INTRODUCTION

The motivation for our work is to provide a precise data for the building reconstruction and the need to interpret scenes as part of establishing an Internet-hosted Exabyte 3D World model [7]. The need to address the human scale of such a World model leads one to consider street side images, either via the use of an organized industrial sensor approach [3] or via crowd sourcing based on user-provided imagery [12]. Reconstruction of the buildings is considered a key part in this workflow [1].

In a single image scenario we consider only one image of the examined building. The method described in this paper is designed to process complex facades of historical building, containing a large variety of ornaments, arches, patterns and divisions. Algorithm is able to process façade projected from a wide angle to the façade normal, therefore under a high perspective distortion. For the purpose of testing, we created a database of various historical and modern buildings located in the urban core of the Austrian city Graz and its peripheries. The images also exhibit a variety in different lighting and weather conditions.

In our work, we consider the availability of large sets of data, currently located online in crowd-sourced internet databases as important factor for urban modelling. For many historical objects, there are usually hundreds of different views from different authors. These redundant datasets can provide additional information for improving the results of recognition algorithms. Therefore, in our work we extend the method into a multi-view scenario.

The effect of multi-view imagery on various geometric scene analyses has been established [5]. It is less well understood how the interpretation of a scene is affected by the transition from a single image to a multi-view image stack. The differences in results between both scenarios will be presented in this paper.

2. FAÇADE ANALYSIS IN A SINGLE IMAGE

The description of an algorithm for processing a single façade located in a single image is given in this section. Our work is based on horizontal/vertical gradient projection approaches, primary on a work of [9]. This is a natural approach for the façade analysis, but in the original form, it is not suitable for complex facades, where the high levels of gradient in vertical/horizontal direction can be located also outside the windows area (see Figure 1). We therefore introduce a new method to deal with this problem.

2.1. Identification of Facades

We consider the building facades to be identified in the single images by their borders. This can be achieved in several ways. In the approach of Lee, Nevatia [8] the aerial model of the scene with un-textured building frames is available and the rectified facades are obtained by projecting a digital image into this model.

In this paper, our main motivation is to process the information from the large online databases. For these types of datasets, the aerial models are generally not available and the geo-tagging is usually missing. But recent efforts towards organization of online digital information have provided us with the means of façade extraction. The ongoing work on automated block adjustments algorithms

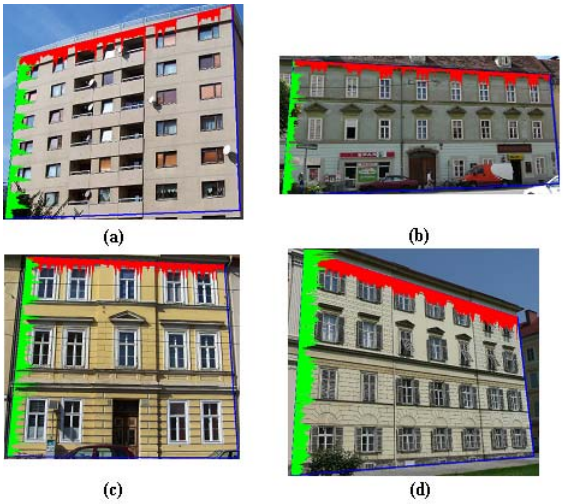

Figure 1: Four different types of facades in our database. The vertical gradient projection is marked in a green and the horizontal gradient projection is in a red color. Façade (a) is relative simple and the windows can be directly extracted from the projections. Façade (b) has additional horizontal structures, which make the horizontal separation difficult. Facades (c) and (d) have both projections highly non-regular and the extraction of windows is more complex.

(Photo-tourism, Photosynth) [12] demonstrates that the processing of large, uncalibrated image databases can be performed effectively and provide sufficient information for the extraction of sparse 3D point clouds. We also use the method of automatic context-based semantic segmentation of street-side scenes [11] to identify points belonging to the façade. In this approach we can identify the facades as the planar-like objects in the point clouds that have been classified as the facades in the segmented images (see Figure 2).

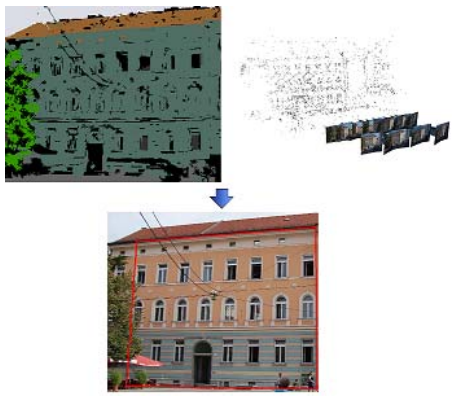

Figure 2: Top, left – semantic segmentation of the image. Façade is marked in dark green color. Top, right – a point cloud of the scene. Bottom – extracted borders of the façade (lines computed from rectification).

The semantic segmentation step allows us to automatically eliminate errors caused by occlusion (from vegetation, pedestrians, or cars), eliminate roofs and better define the borders of the façade.

After the façade is identified in the image, the computation of vanishing points (considering two-point perspective) is the next step [2]. This provides us with the projective transformation matrix and we can perform the rectification of the façade.

2.2. Extraction of Levels

The gradient projection methods are based on observation, that in the simple building facades, the strongest vertical and horizontal gradients are located at the edges of the windows. In the more complex facades (with multiple different objects other then windows), this observation is usually not valid. Strong horizontal responses can be generated at the façade rims, shop signs or aches and vertical responses at columns or stone plates. Therefore, we approach this problem in different way. The general idea is to segment the façade into rectangular areas – blocks. Subsequently, we use visual features to label each block as "façade" or "window".

We use the vertical projection to establish a horizontal division of the façade. For each local peak in the vertical projection a horizontal separator line is created. In this step, the façade is divided into a set of levels (bordered by separator lines) (see Figure 3). The horizontal projection of gradient is computed for each level separately and each level is divided into a set of blocks. The application of threshold on the horizontal projection in each level will provide the borders for the block. The areas with the overall projected gradient above the threshold and the areas below the threshold are separated into different blocks (see Figure 3).

Labelling of blocks into the "façade" and "window" category is performed next. As indicated by previous description of façade division, the block with "window" label usually does not contain entire window, but the part of it. Therefore, it is difficult to label blocks based just on the position or gradient content (glass table inside window may contain very limited gradient, but may still be inside one block as part of the window). The solution for this problem in our approach is to use the color histograms as descriptor for the façade and the window areas. The labelling is decided based on the size of the block, color and the gradient content of the block. This is done in an iterative process, where in each loop, the decision for each block is made, if it is part of façade, or not. In the pre-processing phase, the color histogram in a HSV color space for each block is computed. The blocks horizontally longer than 1/3 of the façade are automatically labelled as façade blocks.

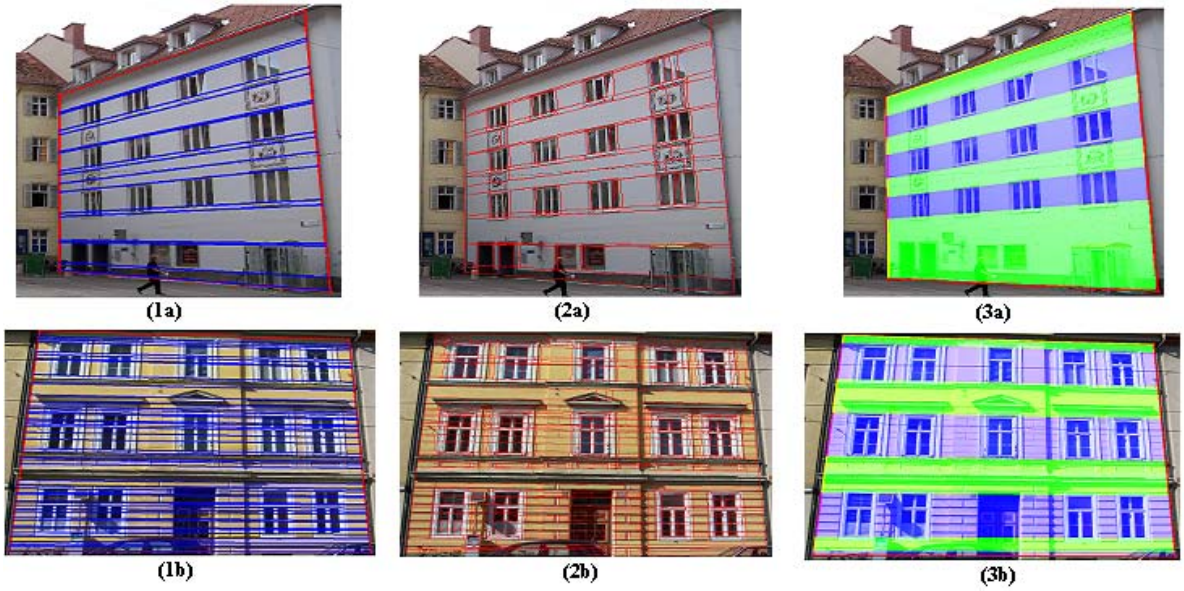

Figure 3: Analysis of a simple (a) and a complex (b) façade. In the first column, the separators between the levels are displayed. First image (1a) contains 16 levels, second image (1b) 46 levels. Second column displays the blocks located in the facades – 122 blocks for the first façade (2a), 1021 blocks for the second façade (2b). Third column shows the separation between the window levels and the façade levels.

The blocks with high gradient content are labelled window blocks, but this label can be changed in the subsequent iterative process.

In the iterative process, the overall color histograms of the façade and windows are refined in each step.Every block marked as a façade/window contributes into the façade/window color histogram. Subsequently, reclassification of each block is performed.

When all blocks in one level are labelled as the façade, entire level is excluded from the reclassification, but still contributes into the color histogram. After several iterations – usually less then ten, depending on the complexity of the façade – there are no more changes in labelling. After this step, all blocks are labelled as window or façade and the final window detection is performed, based on this information.

2.3. Window Detection

In a simple façade, the methods of horizontal/vertical projection of gradient are able to identify windows and non-windows levels directly from the projections. This is also the case in our approach. Since there is no extensive gradient outside the windows area, the divisions between the levels are located on window frames. However, the presence of different patterns on the façade of more complex historical buildings is the reason, why there are many more levels identified in the horizontal projection of these type of objects. There are usually multiple levels covering one windows row and our next step is to group these levels. The problems with grouping can be observed in the blocks bordering windows. Frames of the windows in the historical buildings are often irregular and may contain extensions into the facades, or different ornaments. Also the different types of arch windows are usually divided into several non-similar levels. Therefore, we identify the inside window and façade levels at first as levels containing blocks with strongest response to color and gradient classifiers. Subsequently we move into the in-between levels. The identification of a level as the façade/window is based on the presence of window blocks, the identification of neighbouring level and the height of the level (see Figure 2).

In the next step, we proceed with the identification of windows inside the window levels. As the levels are assumed to be located horizontally – parallel to the ground plane, the borders of the windows are vertical objects inside the window levels. As the blocks inside levels are already labelled as window/non-window, the identification of window borders is straightforward. The assumption is that the border is located in the area of intersection between the most window and non-window blocks. For

the testing purposes, the window borders are projected into the original image (see Figure 4)

3. MULTI-VIEW SCENARIO

The focus of our work is at the crowd-sourced, online open image dataset. The images are contributed by a large number of users and are taken in various lighting and weather conditions. The dataset is natively unorganized and lack additional information (camera calibration, geo-tagging,…). The main advantage is, that it exhibit a large level of redundancy, as the volume of the digital images itself will present a single object in multiple images.

For the purpose of testing the multi-view scenario, we have created the dataset that simulates the crowd-sourcing paradigm of the open databases. The images were taken mostly in the Tummelplatz – in historical centre of the city Graz, Austria and the surroundings. There are 5 main building facades, each projected into 20 – 100 images.

Figure 4**:** An example of complex facades on the left and the windows (blue) and façade (green) labels on the right. Both buildings have complex facades and are under perspective distortion.

3.1. Image Matching

In the general case of urban imaging, a block of images would be triangulated in today's typical workflows as illustrated by Photo-tourism and Photosynth [12]. We also employed this approach and created a sparse 3D point cloud. The algorithm described in [6] was used to extract this point cloud (see Figure 5).

As our goal in this paper is to match the building facades between two images pixel-by-pixel, sparse point cloud does not provide us with enough data for this. It is necessary to interpolate the positions of pixels between the points belonging to a point cloud inside the façade. We can operate with a simple assumption, that the area between two façade points is planar. In a perspective imaging, a planar object is mapped into the image plane by a projective transformation [10].

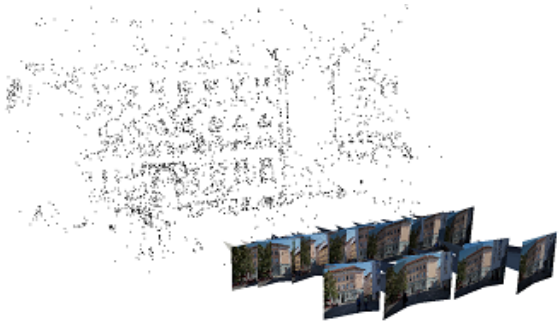

Figure 5: An example of 3D point cloud. This point cloud was created from 28 views and consists of 3498 points (thus of 125 points per image in average). Of a given façade one has 2623 points to work with.

Establishing the parameters of this transformation can provide image matching even for pixels not belonging to the point cloud.

3.2. Window Detection in Multi-view

In the presence of multiple images of the same façade we consider two approaches:

- merging the images into a single, rectified façade and performing the window detection on the merged data
- applying window detection in each image separately and merging the results in a world coordinate system

Merging a multiple images into one, rectified façade is trivial when the means of image matching in a form of point cloud are present. We simply reconstruct the façade in the world coordinates, by assigning a color into each façade pixel. This color is computed as a median from the hue, saturation and intensity from each corresponding pixel in the multiple views. The selection of median would

provide the elimination of outliers on the façade, like shadows, temporal object occlusions, or specific illumination problems. The façade analysis and window detection algorithm is applied to the rectified façade without any modification.

The application of the method on each image in the multi-view group of the façade is straightforward. After this step, we have the candidates for the window in each image located. The coordinate of the corners are projected into the world coordinates for each window candidate. For each window, the corners are computed as the average of the corners of window candidates.

4. RESULTS

In our experiments we use 5 facades, located at multiple images and their corresponding point clouds. The average probability of detecting the window is 91.4%. In subsequent experiments, we evaluate the precision of window placement (only for windows that were detected).

We compare our method on single image scenario with the typical gradient projection method, as described in the paper of Lee, Nevatia [9]. For the testing purposes, the windows were manually marked in the images. Precision of window placement (in percentage) is computed as ratio between the width/height of detected window and the closest manually labelled window width/height to façade dimension. In this experiment, we examine a relationship between the gradient content, as the measure of façade complexity and the precision of window placement. Gradient content of the façade is computed as an average of gradient value {0,...,255} for each façade pixel (windows pixels are not considered as part of the façade in this case). The results are displayed in the Figure 6.

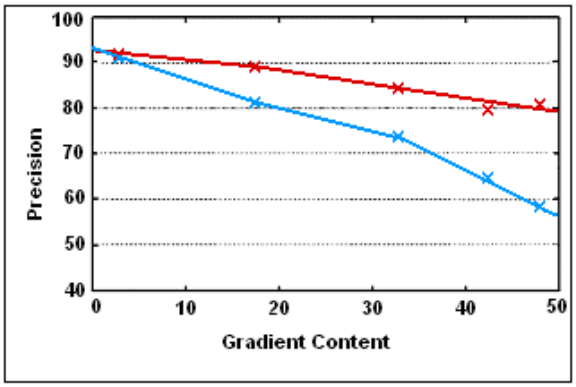

Figure 6: Relationship between the gradient content of the façade (excluding windows) and the precision of windows placement (in percentage). The blue line is displaying the relationship for the standard gradient projection method; the red line is for the method described in this paper.

From the results of this experiment we can conclude, that method described in this paper performs significantly better for the facades with high gradient content. Most historical building in our database (city core in Graz) has a gradient content between 40 and 50. In this group, the precision of window detection can improve up to 22%, using our method.

Our second experiment is focused on an implication of multi-view approach. We examine the dependency between the precision of window detection and the number of different view of the façade. Both approaches described in section 3.2 have been examined. The results can be observed in the Figure 7.

This experiment show that at the certain number of images, the precision in window detection is coming to the limit for both methods. Also, the method of first detection, then merging provides better improvements in multi-view scenario, when more images are available. This is considered to be an effect of a more robust error management for this type of approach, as the outliers are averaged and subsequently over-weighted in the merging step.

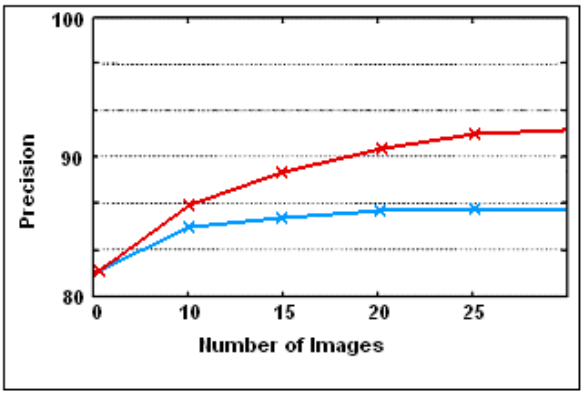

Figure 7: The relationship between the number of images in the multi-view scenario and the precision of window detection (compared in a hand-labelled, rectified façade). The blue line is for the method a) (first merging, than detection), the red line is for the method b) (first detection, then merging).

5. CONCLUSION

In this paper, we describe a gradient projection method designed for automatic processing of complex building facades. This method creates a rectangular block division of a façade and proceeds with labelling based on visual features. As such, it can be considered as a step between the gradient projection approaches [13][9] and the general segmentation methods [4].

We examined the implication of a multi-view scenario in this paper. The presence of a large volume of digital

data can be considered a typical situation for a computer vision method today. In this work, we focused on processing of data from open online databases, with crowd-sourcing paradigm.

Even this paper is focused on a window detection, we observed that many typical façade objects (arches, rims, columns, rectangular patterns,…) have specific signature in the gradient projections, and thanks to the division into multiple levels, they can be identified as well. Therefore in our future work, we will focus on the more general façade analysis.

Acknowledgments: This work has been supported by the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209

6. REFERENCES

[1] Čech J., Šára R. "Windowpane Detection based on Maximum Aposteriori Probability Labeling" *Barneva, R. P. & Brimkov, V. (ed.) Image Analysis - From Theory to Applications*, IWCIA pp. 3-11, 2008

[2] Cha J., Cofer R., Kozaitis S. "Extended hough transform for linear feature detection". *Pattern Recognition,* 39(6):1034-1043, 2006

[3] Haala N, Peter M., Kremer J., Hunter G. "Mobile LiDAR Mapping for 3D Point Cloud Collection in Urban Areas - a Performance Test". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XXXVII, ISPRS Congress, Beijing, China, 2008

[4] Han F., Song-Chun Z.. "Automatic Single View Building Reconstruction by Integrating Segmentation." *CVPRW '04.* pp 53 – 60, 2004

[5] Hartley R., Zisserman A.,. *Multiple View Geometry in Computer Vision.* 2ed,OUP, 2003

[6] Irschara A., Zach C., Bischof H. "Towards wiki-based dense city modeling." *ICCV 2007. IEEE 11th International Conference*, pages 1-8, 2007

[7] Leberl F., Gruber M. "3d-Models of the Human Habitat for the Internet." *Proceedings of Visigrapp, pp 7-15*, Lisbon, 2009

[8] Lee S. C., Jung S. K., Nevatia R. "Automatic Integration of Facade Textures into 3D Building Models with a Projective Geometry Based Line Clustering," *Computer Graphics Forum (Euro Graphics)*, 21(3):511-519, 2002

[9] Lee S. C., Nevatia R. "Extraction and integration of window in a 3d building model from ground view images." *CVPR.IEEE Computer Society*, Vol.2, pages 112-120, 2004

[10] Liebowitz D., Zisserman A. *Metric rectification for perspective images of planes*. 1998

[11] Recky M., Leberl. F. "Semantic Segmentation of Street-Side Images." *Proceedings of the Annual OAGM Workshop.* Published by the Austrian Computer Society in OCG, pp 271 – 282, 2009

[12] Snavely N., Seitz S. M., Szeliski R. "Photo tourism: Exploring photo collections in 3d." *ACM Transactions on Graphics (TOG)*. 2006

[13] Schindler K., Bauer J. "A model-based method for building reconstruction." *First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis*, Washington, DC, USA, 2003