INTERPRETING BUILDING FAÇADES FROM VERTICAL AERIAL IMAGES USING THE THIRD DIMENSION

Meixner P. and Leberl F.

Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, Graz
(meixner, leberl)@icg.tugraz.at

ISPRS Technical Commission IV & AutoCarto Symposium
ASPRS/ CaGIS 2010 Fall Specialty Conference

KEY WORDS: floor detection, window detection, 3D façade models, oblique aerial imagery, vertical aerial imagery, viewing angles, semantic interpretation, real properties

ABSTRACT:

Information is being extracted from vertical aerial photography and various data products for an efficient interpretation of terrain objects. Our focus lies on characterizing individual properties using aerial imagery as accurately as possible. We want to determine the size of buildings, their number of floors, the number and size of windows, the existence of impervious surfaces, status of vegetation, roof shapes with chimneys and sky lights etc. To achieve robust results it is very important to incorporate all data that a set of images can offer. A key aspect therefore is the inclusion of the 3rd dimension when interpreting façade images and to deal with the different imaging angles when interpreting the details of building facades from vertical aerial photography.

This paper first addresses the question which incidence angles are sufficiently large to get useful results. Secondly we show that novel oblique imagery suffers from excessive occlusions that prevent the floor and window detection to produce meaningful results. We finally explain the use of 3D point clouds to deal with complex façades with balconies, awnings and arches. Furthermore, we obtain from the 3d representation of the facades also their exact footprint. We expect to achieve an enhanced quality of the floor counts and window detection results. The topics are closely related since one first needs to understand which façade images have been taken under too small angles so that the facades are grossly distorted. Second, the plane sweep algorithm needs images with as small a distortion as possible, and with a pruned-down set of images.

1. INTRODUCTION

In previous work we have shown that the automated count of floors and windows is feasible using vertical aerial imagery (Meixner & Leberl, 2010a). The initial accuracies reach levels beyond 90% but depend strongly on the shape and structure of a façade. Current efforts are directed towards an increase of that accuracy by the use of multi-images, and by the implementation of the 3rd dimension for façade analysis. Initial results from developing 3D point clouds of façades from vertical imagery and then using these in the façade analysis do offer encouragement. For the evaluation and visualization of these results we use plane sweeping methods that will be discussed in this paper. We show how these techniques can be implemented for façade images in vertical aerial photography, and what the advantages are vis-à-vis 2 dimensional interpretations of a façade image.

Additionally an interpretation of façade images always needs to cope with limitations in the look angles under which a façade is imaged. There is a strong relationship between those look angles and the accuracy of the analysis. Below 15°, the results become poor. At 20° and beyond, the limitations of any façade analysis are more due to the occlusions by other buildings and by vegetation, or by the algorithm's ability to deal with balconies and other irregularities, than by a lack of image geometry and quality.

The recent emergence of oblique aerial photography begs the question of their merits vis-à-vis vertical photography. The topic is relevant since the Internet inspires an interest in showing urban areas and modeling them in 3D from vertical and oblique aerial photography, aerial LiDAR and from street side imagery and street side LiDAR. Vertical aerial photography continues to be the workhorse for complete maps and orthophotos, whereas many dense 3D point clouds today are being produced by LiDARs (Leberl et al., in print). With the transition to digital sensing, image overlaps can increase without adding costs. This improves the accuracy of 3D data, the automation opportunities and the extent of occlusions. One can argue that the no-cost-per-image-paradigm has changed previous value systems: LiDAR may not add the value it once had over point clouds from film imagery, and highly overlapping digital vertical images may show facades in sufficient detail to eliminate the need for oblique photography.

For the evaluation of these topics we have created a processing framework (Meixner & Leberl, 2010a). This is being applied to a test area in Graz (Austria) with about 104 buildings with 233 single façades and a total of 870 overlapping images of those 233 façades. We show that façades can be successfully analyzed from vertical aerial photography with an accuracy of the floor and window counts in the range of 90%. Since oblique photography suffers from more significant occlusions than vertical imagery does, its value to floor and window mapping is compromised. In addition one might wonder which other value exists beyond that available from vertical photography in the analysis of façades. Even a qualitative visual inspection raises doubts that oblique outperforms high quality, high resolution and high dynamic range vertical imagery.

Finally, methods are needed to deal with complex façades and spatial structures. We therefore have embarked on work to use the 3rd dimension by means of so-called Plane Sweeping to improve the floor and window counts.

2. VERTICAL VS OBLIQUE AERIAL IMAGES

We compare vertical and oblique aerial images. We concentrate on the viewing angles of the facades from two datasets and evaluate them using a floor and window detection technique previously introduced by Lee & Nevatia (2004) and adapted in our previous work to vertical aerial images (Meixner & Leberl, 2010a).

2.1 Oblique Camera Geometry

Oblique aerial imagery has been easy to come by in the form of Microsoft's BING/Maps mapping website. The images have been produced with a Maltese-Cross configuration (Petrie, 2009). For any quantitative work with oblique images we need to employ the image specifications. Since these are typically kept confidential by the data providers such as Pictometry [www.pictometry.com], we need to reconstruct them by means of a photogrammetric resection-in-space.

High quality vertical aerial photography is available for the Graz test site, identical to the material used in the Microsoft Bing Maps website. Those images have been acquired by the UltraCam-series of aerial cameras. Figure 1 presents an example in Graz (Austria) with a vertical coverage and superimposed the outlines of an oblique image. Also shown is the oblique image itself in its original geometry.

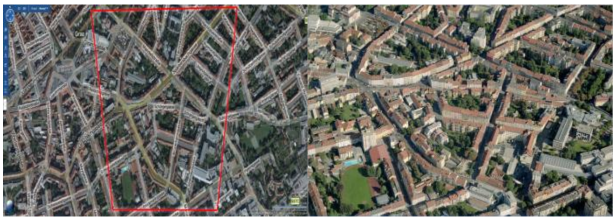

Figure 1: Detail from Microsoft Bing Maps. Left is the orthophoto and superimposed the outline of an oblique aerial image produced with the Pictometry system operated by Blom [www.blom.no]. Right is the oblique aerial image.

Well-mapped terrain with large vertical structures including a church can serve as a 3D test area to compute this resection in space. The result is summarized in Table 1.

The oblique images have 2,672 rows and 4,000 columns. These results coincide with the values calculated by Prandi (2008).

Pixel size in the image plane =	9µm
Focal length =	85.5mm
Viewing angle of the camera =	24° x 16°
Flying height above ground =	1,130 m
Distance to near range =	850 m
Near range off-nadir angle =	37 °
Far range off-nadir angle =	53 °
Horizontal GSD at near range =	14 cm
Horizontal GSD at far range =	19 cm
Distance to far range =	1530 m

Table 1: BING/Maps oblique imagery parameters reconstructed from known terrain points.

2.2 Vertical Image Geometry

In a next step we use the technical parameters of the UltraCam series. Table 2 summarizes some relevant geometric parameters of a digital aerial camera in the form of the UltraCam-X and the wide-angle UltraCam XP – WA.

	UltraCam X	**UltraCam XP - WA**
Image Rows x Columns	14, 430 x 9,420	17, 310 x 11,310
Image size in X and Y, in mm	103.9 x 67.8	
Pixel size in image plane (µm)	7.2	6
Focal length, mm	100	70
Max Look angle off-nadir (°)	27.5	36.5

Table 2: Some geometric data of two typical digital aerial cameras (from www.vexcel.com)

2.3 Pixel Sizes on Facades

For a vertically-looking camera, the pixel on a façade (FSD or Façade Sampling Distance) changes as a function of the look angle off-nadir α with:

$$FSD = GSD/ \tan(\alpha)$$

This results in values shown in Table 3 for a GSD (=Ground Sampling Distance) at 10 cm, a typical value for urban aerial photography. The façade pixels are rectangular.

Angle (deg)	0	5	10	15	20	25	30	35
Pixel vertical [cm]	∞	114	57	37	27	21	17	14

Table 3: Incidence or look angles and vertical pixel size within a façade. The horizontal pixel size is at 10 cm.

For an oblique camera, the pixel size within a vertical plane is defined by two angles. Angle β is the orientation of the optical axis off-nadir. Angle α is the angle between the optical axis and the actual imaging ray. The off-nadir angle β produces different GSD-values in two perpendicular directions. In the direction of the inclination of the optical axis we find GSDr, with r being the range direction or direction between nadir and the optical axis. In the perpendicular direction, called azimuth direction, we have GSDa:

$$GSDr = p * H * \cos(\alpha) / (f * \cos^2(\beta))$$

$$GSDa = p * H * \cos(\alpha) / (f * \cos(\beta))$$

A vertical façade is resolved as a function of where in the image it is located and this is defined by the second angle α, producing a vertical pixel dimension FSDv:

$$FSDv = p * H * \cos(\alpha) / (f / \sin(\beta))$$

Table 4 presents some vertical façade pixel sizes for the oblique camera with a look angle at 53° and compares this to pixel sizes from an UltraCam X with an angle at 22°. In this example, the vertical façade pixel sizes for the oblique camera at an angle of 53° and for an UltraCam X at an angle of 22° are almost identical. The simple conclusion is permitted that the pixel size not only is a function of the look angle, but also of the flying height and the GSD. While the look angle appears a lot less attractive from a consideration of look angles in the Ultracam, on a given façade, this does not propagate into an inferior geometric resolution.

	Degrees	Azimuth (cm)	Range (cm)	FSD (cm)
Oblique camera (near range)	37	14.7	20.9	27.7
Oblique camera (far range)	53	19.6	27.7	20.9
UC-X	22	8.1	8.1	20.0

Table 4: Size of a pixel on a façade in cm as a function of the look angle, in °. The vertical image GSD is at 10 cm.

2.4 Efficiency of Aerial Data Collection

A consideration of image pixel sizes ignores the efficiency of one versus another imaging approach and technology. Flying at a certain flying height to achieve small pixels, and producing images with a large format will be more efficient than to fly with small formats for a small swath width and at a low flying height. An UltraCam for example produces 17.5 K pixels in one single flight line. An oblique camera will have to match this number to be comparatively productive. At a frame size of 4,000 pixels, one will not easily match the productivity of a vertical mapping camera.

2.5 Experimental Results

We work with a 400m x 400m test data set in the city of Graz (Austria). The vertical images were acquired with an UltraCamX (Vexcel/Microsoft) at a GSD of 10cm and 80/60 image overlaps. The oblique images were taken from the Microsoft BING/Maps website in its “Classic” version, and have a GSD of nominally 14cm according to the near range edge (table 1).

2.6 Visual Comparison

A visual comparison of vertical versus oblique images in Figure 2 does not result in a clear advantage of one versus the other approach. At an off-nadir angle of about 45°, the oblique images have more significant occlusions, given that the vertical images show the same facades at an angle of only 27°. Regarding the radiometric range, one would give the vertical images an advantage. This visual impression from original images is overwhelmed by the differences in geometry. To eliminate that factor, we create rectified versions in the plane of the façade, as shown in Figure 3. Again, a visual comparison does not show a clear advantage of one over the other technology.

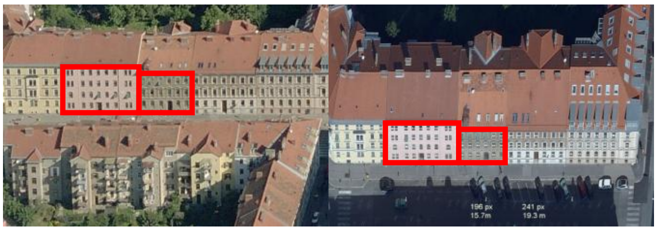

Figure 2: Oblique aerial image at 45° look angle taken from Microsoft Bing Maps (left); Vertical aerial image obtained from UltraCamX at a look angle of 27° (right).

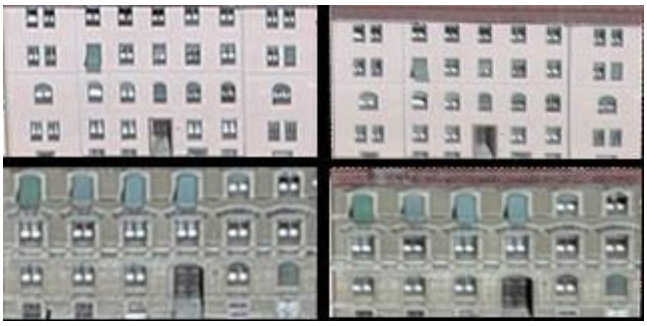

Figure 3: The marked sections of Figure 4 have been rectified. At left are two sections from the oblique data at 45°, at right from the vertical data at 27°.

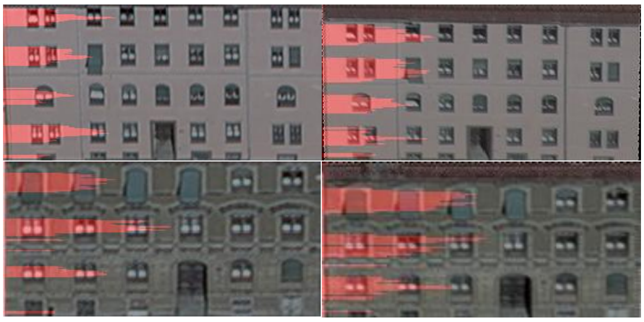

Figure 4: Two examples for floor detection using edge histograms. The oblique images are to the left.

2.7 Counting Floors

A less subjective and more quantitative comparison of oblique versus vertical is expected to result from analysing the images and extracting semantic information. A floor detection algorithm has been explained by Meixner & Leberl (2010b). Figure 4 explains that a histogram is being built from horizontal Prewitt edges and local extrema of the histogram serve to get a floor count. Applying this approach to about 870 facades in the Graz test area’s vertical images, and to a subset of 120 facades in the corresponding oblique images (from Bing/Maps) leads to Table 5. We find in this type of quantitative analysis that the result is seriously compromised by the occlusions which naturally are larger in the oblique images.

2.8 Counting Windows

A histogram-based count can also deliver the number and locations of windows. Figure 5 explains the principle of the approach. Of course one will want to apply various constraints on window size and distance between windows etc. to overcome the effects of data noise. Table 5 shows the accuracy achieved in the Graz test data set from the facades on vertical and oblique images. Again, occlusions are the main obstacle to a competitive result from oblique images.

In Table 5 the success rate of window and floor detection is calculated by dividing the number of facades where the floors and windows are correctly determined (e.g. 5-10) by the total number of facades for every angle (e.g. 7/21). As one can see the floor and window detection results for oblique images are not as good as the results using vertical aerial images. Reasons for that are the poor resolution of the oblique aerial images and occlusions from other buildings and vegetation. Concerning the floor detection occlusions are the main reason for these results. Concerning the window detection the poor resolution of the images is one of the reasons for the outcome.

Angle [deg]	< 5	5 - 10	10 - 15	15-20	20-25	> 25	Oblique
Floor Detection	0	7 / 21	79 / 103	191 / 221	255 / 279	228 / 246	90/120
Floors Percentage	0%	33%	77%	86%	91%	93%	75%
Window detection	0	6 / 21	69 / 103	174 / 221	233 / 279	212 / 246	79/120
Windows Percentage	0%	29%	67%	79%	83%	86%	66%

Table 5: Counting floors (above) and windows (below) from vertical images and results depending on look angles. Last column is from oblique images where floor counts are compromised by occlusions

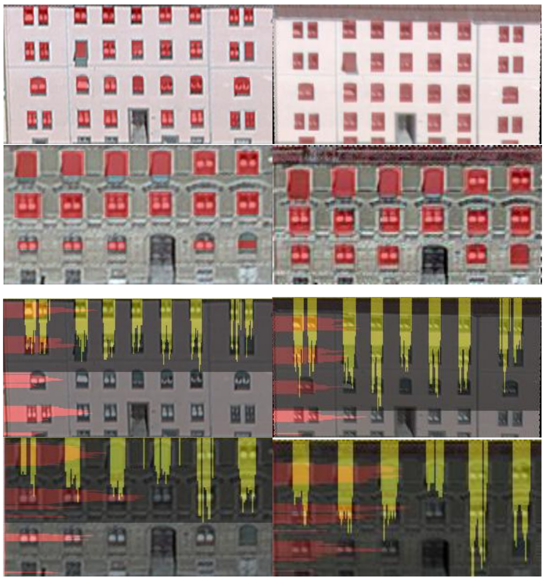

Figure 5: Window detection approach using edge histograms (upper image). Marked window locations and sizes (lower image).

2.9 Discussion

We demonstrate that the floors in façades can be counted with a 93% success rate from vertical aerial photography, and that windows can be counted with an 86% success rate. This is feasible since the images have been taken with large overlaps so as to image each façade at a sufficiently large look angle of 20° to 27°. We also show that the visual inspection of vertical versus oblique images favors the vertical data due to better radiometry at comparable pixel sizes. The major problems of oblique images are occlusions that prevent one from counting the correct number of floor and windows. The efficiency of aerial imaging may favor vertical technologies over the oblique approach. Vertical images today produce 200 Megapixels per exposure, whereas oblique cameras still operate at the 10 Megapixel level. Even if one were to consider that in a Maltese Cross arrangement one operates with 5 such cameras, this still adds up to only 50 Megapixels. A limitation of the current "normal angle" aerial cameras is the look angles one can achieve *in the direction of flight* at perhaps 17° off nadir. Solutions are either a cross flight pattern, or the use of a wide angle camera model such as the UltraCam Xp-WA with 26° in flight direction, or the use of the new single CCD-chip DMC-II, recently announced by Intergraph. Going beyond a mere "eye candy" approach for the use of oblique images, one will quickly find that novel high-redundancy vertical aerial images offer a superior source of information about urban areas, street canyons and facades. We suggest that the benefits from vertical aerial photography have been undervalued, and that conversely benefits any from oblique images have been overstated.

3. 3D RECONSTRUCTION OF FACADES

The interpretation of facades in the previous chapter was based on the assumption that a façade is planar and thus essentially has two dimensions (x and y). But there are cases where this 2-dimensional approach to detect windows and floors will fail. While problems will be caused by vegetation and by one building covering up another, our interest is in coping with balconies, bay windows, attached staircases and elevator shafts. These all do represent deviations of a façade from a planar object. Figure 6 illustrates the footprint of a façade and shows how façade details extend into the third dimension. When the emphasis is on fats throughput and a simple approach, the third dimension gets suppressed, as seen in Figure 6c.

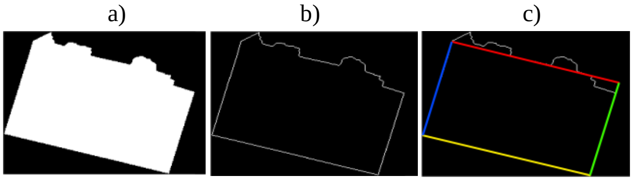

Figure 6: The classification layer "building" is based on color and texture. (a) shows the binary layer from an image classification, (b) its contour as a raster and finally in (c) the simplified geometric figure of the façade footprint.

Problems will exist if parts of the façade lie in different planes. Figure 7a is rectified image of facade with balconies and awnings. A search for "floors" and "windows" in Figure 7b fails.

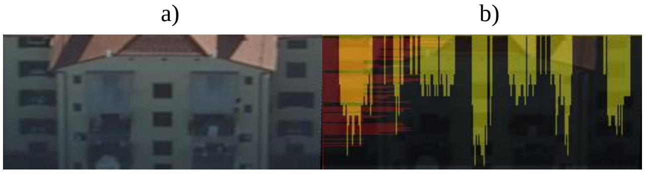

Figure 7: To the left is a rectified facade image with a depth structure, to the right a failed count of windows. The 3D structure needs to get considered.

A possible elimination of these problems could be a splitting of the façades into multiple façade fragments. However, for our experimental data set of 104 buildings with 233 facades this method would yield a quadruple number of facades, and each image would only show a small element. One will have to cope with ambiguities because of the smallness of the façade elements.

A more promising solution is the explicit consideration of the 3rd dimension. We want to use the so-called *plane sweeping* method with its advantage that one no longer needs to assume a single vertical plane per façade. One starts from an approximate location of a façade. The method produces a 3D depth map for a façade. We employ a method that has been described by Zebedin et al. (2006). It supports the definition of architectural elements that stick out from a façade.

3.1 Use of the 3rd Dimension

The result of the plane sweeping is a depth map that can be used for further analysis. The algorithm consists of three distinct steps:

- the 2D space is iteratively traversed by planes that are parallel with the main façade plane;

- an image is being warped;
- a multi-view correlation is performed

3.2 Plane Sweeping

The plane sweep approach is a well established method in computer vision for an image based reconstruction from multiple views. This is in contrast to traditional computational stereo methods. A 2D space is defined by multiple planes that lie parallel to the key-plane (see Figure 8)
A *key-plane* is the approximate façade-plane. Additional planes are set parallel to the key-plane about one pixel apart (in our test area, this is at 10 cm) in both directions from the key-plane.

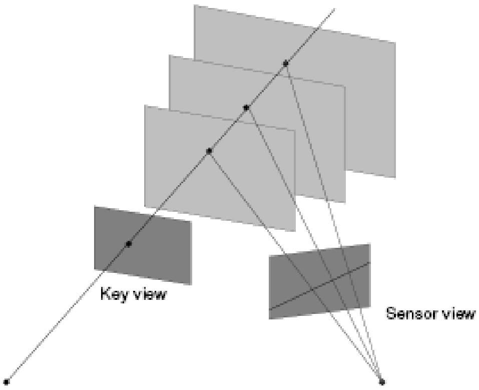

Figure 8: Plane sweeping principle. The homography between the façade's reference plane and the sensor view varies for different depths.

If the plane at a certain depth passes exactly through parts of the object's surface to be reconstructed, a match will exist between the relevant parts of the new sensor view and the key view, the match being computed as a correlation. The sensor images are warped onto the current 3D key plane using the projective transformation. This appropriate H is obtained from:

$$R = R_2 * R_1^T$$
$$t = t_2 - R_2 R_1^T t_1$$
$$H = K(R - tn^T/d)K^{-1}$$

K … intrinsic Matrix of the camera
$(R|t)$… relative pose of the sensor view

Details on epipolar geometry and the mathematics behind these equations are described in (Hartley & Zisserman, 2004).

3.3 Image Correlation

After projecting a sensor image onto the current plane hypothesis, a correlation score for the current sensor view is calculated. The final correlation score of the current plane hypothesis is achieved by integrating all overlapping sensor views. For the accumulation of the single image correlation scores a simple additive blending operation is used. The calculation of the correlation coefficient r for window-based local matching of two images X and Y is

$$r = \frac{\sum_{i \in W}(X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{\sum_{i \in W}(X_i - \tilde{X})^2 * \sum_{i \in W}(Y_i - \tilde{Y})^2}}$$

This formula is invariant under affine linear changes of luminance between images. To receive robust results for the determination of the correlation coefficient neighboring pixels are added up. This can be done for a neighborhood of 3x3 pixels. Figure 9 shows the result of the image correlation for 4 different planes.

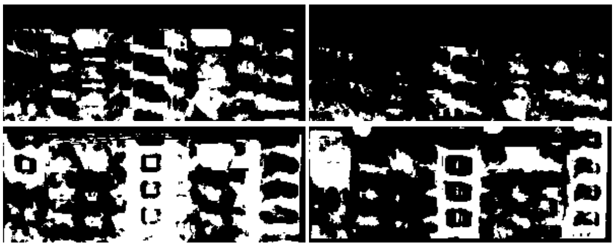

Figure 9: Correlation coefficients calculated for 4 different planes visualized as binary images (white areas have the largest correlation values)

For the determination of the best correlation coefficients we use a total generalized variation approach proposed by (Pock et al 2007). Because of poorly textured areas and repetitive structures in the façade images we receive unstable results. To avoid these problems we use the 3D structures spatial coherence and employ various surface views in a multi image approach. The therefore used variational approach focuses on stereo matching as an energy minimization problem.

The energy consists of two terms, a smoothness term to model the spatial coherence of the surface and a data term that reflects the multiple image matching qualities. Pock et al. (2007) have developed a formulation of variational stereo:

$$\min_u \left\{ \int_\Omega |\nabla u| \, dx + \int_\Omega g\big(x, u(x)\big) dx \right\}$$

Ω… image domain
g(x,u)… image matching term for each surface point

The left term is the so called total variation, ensuring that the façade surface is smooth and preserves sharp discontinuities in the solution while remaining convex. The data term to the right measures the matching quality. These results can directly be converted to a 3D depth map and visualized.

3.4 Experimental Results

In our experimental data we have produced parallel planes with an distance between each other of 1 pixel (10cm) to gain as much information about the structure of a façade as possible. That means that the depth map can have a maximum accuracy of 10cm if it would be possible to eliminate noise in the depth map. Figure 10 shows a result of this calculation, where we tried to find the major planes of a façade image.

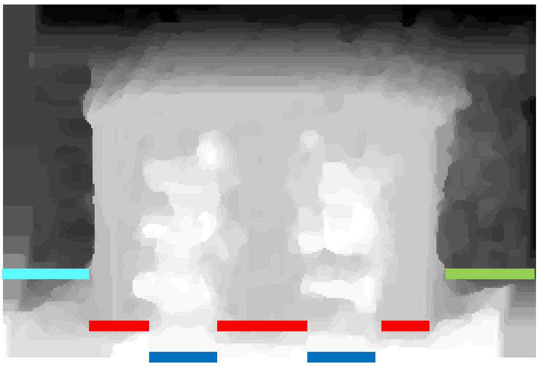

Figure 10: Simplified depth map of the façade image from figure 7a developed from 5 overlapping vertical aerial images. For an interpretation of the colored bars see figure 11.

For the determination of the major planes of one façade we were using the result shown in figure 10 and calculated column wise the most likely planes. The mean values were calculated for every column this value gets compared with the neighboring columns. Figure 11 shows the final footprint of the façade from Figure 9 and its 3 D point cloud in Figure 10 and the 4 detected major planes.

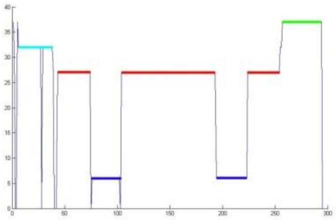

Figure 11: Footprint of the façade (different colors relate to the different determined façade planes) from Figure 5a showing the 4 major planes of the façade.

4. CONCLUSION

We have first quantitative accuracy results for the automated count of floors and windows extracted from vertical aerial photography. Current accuracies at about 90% are expected to improve with the use of multiple images and refinement of methods to better cope with shadows and partial occlusions, for example by vegetation. A second result is the advantage of vertical imagery over novel oblique photography which suffers excessive occlusions in urban street canyons.

A third result addresses the use of 3D point clouds of architecturally structured facades with balconies, arches or awnings. We show that computation of such point clouds from highly overlapping *vertical* aerial imagery is indeed feasible. One will have to cope with limitations in areas of façade homogeneity, thus in areas without any architectural structure, where matching may be ambiguous: the assignment of a pixel to a certain plane of the depth map may not be possible.

Our future work is obvious: staying with vertical aerial imagery, we will have to improve the algorithms, extend the processing framework, exploit image overlaps, and broaden of the test areas to include coastal resort environments, historical small towns, alpine terrains, urban cores with skyscrapers and industrial zones with windowless factory buildings.

5. REFERENCES

Hartley R., Zisserman A. (2004) *Multiple View Geometry in Computer Vision.* Second Edition. Cambridge University Press, March 2004, pp. 219-243.

Leberl F., Bischof H., Pock T. Irschara A., Kluckner S. (2010) *Aerial Computer Vision for a 3D Virtual Habitat.* IEEE Computer, June 2010, pp.1-8

Lee S.C., R. Nevatia (2004) Extraction and Integration of Window in a 3D Building Model from Ground View Images. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVP'04*

Meixner P. Leberl F. (2010a) *From Aerial Images to a Description of Real Properties: A Framework*. Proceedings of VISAPP International Conference on Computer Vision and Theory and Applications, Angers 2010

Meixner P. Leberl F. (2010b) *Describing Buildings by 3-dimensional Details found in Aerial Photography.* International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, in print Vienna, Austria.

Petrie G. (2009) *Systematic Oblique Aerial Photography using Multiple Digital Frame Cameras*. Photogrammetric Engineering & Remote Sensing, Vol. 75, No. 2, p. 102-107, 2009.

Thomas Pock, Michael Pock, Horst Bischof, "Algorithmic Differentiation: Application to Variational Problems in Computer Vision," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 7, pp. 1180-1193, July 2007, doi:10.1109/TPAMI.2007.1044

Prandi F. (2008) *Lidar and Pictometry Images Integrated Use for 3D Model Generation*. Proceedings of the XXI ISPRS Congress, Beijing 2008.

Weiss, Y. and Freeman, W. T. (2001). *On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs.* IEEE Transactions on Information Theory, pp. 723–735.

Zach C. (2007) *High-Performance Modeling from Multiple Views using Graphics Hardware.* Dissertation, Graz University of Technology, 2007

Zebedin L., Klaus A., Gruber B., Karner K. (2006) *Façade Reconstruction from Aerial Images by Multi-view Plane Sweeping*. Int'l Arch. Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVI (part 3), Bonn, Germany.