International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B3, 2012
XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia

QUALITY ASSESSMENT OF MAPPING BUILDING TEXTURES FROM INFRARED IMAGE SEQUENCES

L. Hoegner, D. Iwaszczuk, U. Stilla[a]

Technische Universitaet Muenchen (TUM), Germany

Commission III Working Group 5

KEY WORDS: thermal imagery, relative orientation, image sequences, texture extraction

ABSTRACT:

Generation and texturing of building models is a fast developing field of research. Several techniques have been developed to extract building geometry and textures from multiple images and image sequences. In this paper, these techniques are discussed and extended to automatically add new textures from infrared (IR) image sequences to existing building models. In contrast to existing work, geometry and textures are not generated together from the same dataset but the textures are extracted from the image sequence and matched to an existing geo-referenced 3D building model. The texture generation is divided in two main parts. The first part deals with the estimation and refinement of the exterior camera orientation. Feature points are extracted in the images and used as tie points in the sequence. A recorded exterior orientation of the camera s added to these homologous points and a bundle adjustment is performed starting on image pairs and combining the hole sequence. A given 3d model of the observed building is additionally added to introduce further constraint as ground control points in the bundle adjustment. The second part includes the extraction of textures from the images and the combination of textures from different images of the sequence. Using the reconstructed exterior camera orientation for every image of the sequence, the visible facades are projected into the image and texture is extracted. These textures normally contain only parts of the facade. The partial textures extracted from all images are combined to one facade texture. This texture is stored with a 3D reference to the corresponding facade. This allows searching for features in textures and localising those features in 3D space. It will be shown, that the proposed strategy allows texture extraction and mapping even for big building complexes with restricted viewing possibilities and for images with low optical resolution.

1. INTRODUCTION

The analysis and refinement of buildings in urban areas has become an imported research field in the last few years (Pu & Vosselman, 2009; Mayer & Reznik, 2006). Buildings in city models normally consist of simple façade structures and optical textures. To refine and automate the extraction of geometry and textures in urban scenes different solutions have been proposed during the last years (Mayer, 2007; Heinrichs et al., 2008; Lo & Quattrochi, 2003; Pollefeys et al., 2008). Those strategies are extracted geometry and textures together to form a new building or façade and are not merged with an existing geo-referenced city model.

Much urban information cannot be extracted from normal optical images but from other optical domains like infrared. Ground cameras are recording the irradiation of building façades (Klingert, 2006), for the specification of its thermal behaviour for thermal building passports. IR data of buildings are collected from several photos and analysed directly in the acquired images without any geometric or 3d processing.

The focus of this paper is the integration of pre-known building models in the texture extraction process from image sequences to improve the textures' quality. There are two reasons to add infrared textures to existing building models instead of generating a building model and extract the textures from the same infrared image source. The geometric resolution of infrared images is quite low compared to images in the visible domain and geometric features like edges show different appearances which lead to problems in correct extraction of facade planes. For many buildings, accurate building models from databases like CityGML or a building information model (BIM) already exist and the given model has to be improved. Infrared textures are an additional data source for the extraction of windows (Iwaszczuk et al, 2011; Sirmacek et al., 2011).

Thermal cameras record electromagnetic radiation in the invisible infrared (IR) spectra. Thus, surface characteristics of object can be detected, that stay invisible in normal visible spectra. For recognition of objects with little difference in temperature and for identification of small details from distance, thermal cameras must be able to resolve temperatures with an accuracy of 0,01 Kelvin. High-quality infrared cameras are able to record image sequences with standard video frame rate (25 fps) or even higher. Because of special cooling technique, camera optics and the low production numbers, the expenses of infrared cameras are very high compared to normal video cameras. Today, thermal image data is used for many different applications. Typically, IR data of buildings are collected from photos and analyzed directly in the acquired images. Bigger building parts are acquired by combining several images. The results of the analysis are stored in the 2d photos without any geometric or 3d processing. This can be a problem, when images from different cameras or views are combined and stored for further processing.

In contrast to conventional IR inspection of buildings, in this paper an automated strategy is used for texturing an entire building model. Narrow streets and the low resolution and small field of view of IR cameras are serious problems. Only small parts of the building façade are visible in one image. Direct line matching of the images and the projection of the model's edges is used especially for aerial images (Frueh et al., 2004; Avbelj et al., 2010) fails because of the lack of visible façade edges in many of the images. In this case we have to deal with the fact, that structures found in IR images do not always have correspondences in the building model and vice versa. This

problem is even getting worse with a moving camera and inaccurate orientation parameters caused by the GPS system.

Different strategies for matching of given 3d models and images are well known in computer vision. Single image processing is working with 3 or more correspondences between image and model. An overview over 3-point algorithms is given in Haralick et al. (1994). Triggs (1999) introduces a generalisation of the 6-point Direct Linear Transformation (DLT) for camera pose estimation and calibration from a single image with 4 or 5 known 3D points. There are also iterative methods proposed in Haralick et al. (1989). Longuet-Higgins (1981) introduce the 8-point algorithm to projectively reconstruct a relatively oriented scene from two different views without previously known 3D coordinates of the points. The eight point correspondences are used to calculate the Fundamental matrix (Hartley and Zisserman 2003), which describes the relative orientation of the two views. Two images showing the same planar surface are related by a homography. This homography is used to find the relative orientation of the two images via their corresponding plane in the scene (Hartley and Zisserman 2003). When using image sequences, multiple images can be used for pose estimation.

Another approach to relatively orient two images is introduced by Nistér (2004). The algorithm uses corresponding SIFT feature points (Lowe, 2004) in two calibrated views of a scene and calculates the essential matrix of this two images from five corresponding points to find a relative camera motion between. Improvement of robustness is samling sets of five points within a random sample consensus scheme (RANSAC) (Fischler and Bolles 1981). A hypothesis test deals with mismatched points (Torr and Murray, 1997; Zhang, 1998) and allows the combination of hundreds of views using trifocal tensors (Nistér 2000). Further extensions of this algorithm towards the handling of possible wide-baseline image sequences taken with digital still-images and video cameras have been achieved by Mayer (2007), Pollefeys et al., (2008) and Heinrichs et al. (2008). Mayer (2007) adopts Nistér's algorithm for facade extraction and texturing from multiple views. In this paper, these strategies are extended to deal with a given building model and camera path in a global coordinate system to match the image sequence on an existing building model.

2. METHODOLOGY

The matching process between the images and the building model is done in two steps. The usage of continuous image sequences taken from a moving car allows performing a relative orientation of the images of a sequence to extract estimated façade planes and a relative camera path (Mayer, 2007; Heinrichs et al., 2008; Lo & Quattrochi, 2003; Pollefeys et al., 2008). In this paper the observed exterior camera orientation is added as additional observation in the bundle adjustment. This allows recovering the scaling factor of the image sequence which is unknown in only relative orientation. The transfer of the image sequence to the global coordinate system allows the matching of edges detected in the images and the given building model. As mentioned, a matching of image edges and model lines only is not successful in most of the images due to the small part of building models visible in one image. But possible corresponding parts of edges in the image and lines of the model can be introduced to the bundle adjustment. They are of special interest for the correct borders of facade planes.

The second step uses the images and their global coordinates from the first step to perform a coregistration of images and building in 2d and 3d. In the 2d image space a matching of extracted image edges and projected lines of the 3d model can be performed (Avbelj et al., 2010). In the 3d space 3d points are generated from the homologous points of the image sequence orientation process. These points are grouped in planes and matched with facade planes of the 3d model. Both 2d and 3d matching are combined in a bundle adjustment.

The quality of the bundle adjustment of the orientation of the image sequence as well as the matching of the estimated 3d point cloud and the building facades depends on the number and accuracy of homologous point features in the images of the sequence. Two representatives of different point feature classes have been used in this paper. Gradient based features like Foerstner (Foerstner & Guelch, 1987) points are compared to blob detectors like SIFT features (Lowe, 2004). The advantage of gradient detectors is their stability and accuracy for small changes of the camera orientation and scene which is the case in adjacent images of the image sequence. The advantage of blob detectors is their tolerance to changing viewing directions and scales of features which is the case if we try to find two images of the sequence with a big stereo base and many homologous points. Both bundle adjustments, the image sequence orientation and the matching of the sequence and the building model are performed with both feature classes to compare their quality.

The quality of the matching of the image sequence and the building model is done by analyzing the extracted textures. Textures from different sequences at different times and with different orientation parameters are compared through correlation. The textures are extracted by projecting the voxels of a predefined grid from a surface into the image space and interpolate the intensity values. This is done for every image where a surface is partially visible. The final texture has to be combined from these partial textures by choosing the best texture for every pixel. In general, different aspects have to be taken into account for this and especially if no textures covers the hole surface, the best quality solution is quite difficult. If we concentrate on image sequences with a constant oblique viewing direction, this problem can be simplified. Every following image has a higher resolution of all visible parts of all surfaces than all images before. This means, that we can overwrite partial texture 1 with all parts of partial texture 2, where partial texture two was visible, if image 2 is a follower of image 1 and the camera is forward looking.

3. EXPERIMENTS

Current IR cameras cannot reach the optical resolution of video cameras or even digital cameras. The camera used for the acquisition of the test sequences offers an optical resolution of 320x240 pixels with a field of view (FOV) of only 20°. The FLIR SC3000 camera is recording in the thermal infrared (8 - 12 µm). On the top of a van, the camera was mounted on a platform which can be rotated and shifted. Like in the visible spectrum, the sun affects infrared records. In long-wave infrared the sun's influence appears only indirect, as the sun is not sending in the long wave spectrum, but of course is affecting the surface temperature of the building.

Caused by the small field of view and the low optical resolution it was necessary to record the scene in oblique view to be able to record the complete facades of the building from the floor to the roof and an acceptable texture resolution. The image sequences were recorded with a frequency of 50 frames per second. Small changes between two images reduce the number of mismatches on regular, repetitive structures like windows in facades but reduce the accuracy in 3d coordinate estimation. To guarantee anyhow a good 3d reconstruction, the features are tracked through the hole sequence and images are taken for the

3d point generation which show a sufficient number of homologue points and a big stereo base within the sequence. As only small parts of a building facade are visible within one image, a direct matching of edges extracted from the images and a given building model cannot be performed directly. An example of a recorded sequence is shown in figure 1. The position of the camera was recorded with GPS and, for quality measurements from tachymeter measurements from ground control points.

Figure 1. Example images from one sequence along a building.

3.1 Feature Tracking

In a first observation, the number of features per image and the number of images in a sequence a feature can be tracked is investigated. Figure 2 shows an image of a sequence with SIFT features detected in two images and the movement of the features from the first image to the second.

Figure 2. IR image with selected SIFT features, that have correspondences in the following image. Arrows show the moving direction of the points and numer of pixels they move.

For small distances between the images, Foestner points and SIFT features show almost the same number of homologous points in two images. With a bigger distance of the images, the number of homologous points from Foerstner points decreases faster than the number of homologous points from SIFT features. For sequences of several images, this decrease is much smaller but also shows a better performance for SIFT features. Table 1 shows the decrease of the mean number of homologous points with the distance of images in the sequences. For comparison reasons the number of features for the first image was set to 100 for both feature detectors. For selected sequences, the features have been tracked manually to see the decreasing number of features due to features running out of the image.

Distance in frames / seconds	1 / 0.02	10 / 0.2	50 / 1.0	200 / 4.0
Manually tracked	100	91.6	83.2	51.7
Foerstner point single pair	99.5	89.7	47.5	24.8
SIFT feature single pair	99.4	89.8	61.4	38.5
Foerstner points with 10 frames distance step	99.5	89.7	55.8	31.3
SIFT features with 10 frames distance step	99.4	89.8	64.5	42.4

Table 1. Decrease of the mean number of homologous points with the distance of the images for Foerstner points and SIFT features in the sequences.

3.2 Orientation of Image Sequences

The homologous points are used for the calucaltion of trifocal tensors as introduced by Mayer (2007). The bundle adjustment is additionally given the obervations of the position of the camera of every image. The resulting oriented image sequence is used to derive 3d coordinates of the homologous points. The generated 3d point cloud of the SIFT features and the estimated camera position for every image of the sequence can be seen in figure 3. The structure of the facades is already visible. Most of the points are located in the edges of the windows and grouped in lines. The variance analysis of the bundle adjustment of the relative orientation of the image sequences shows smaller errors for SIFT features compared to Foerstner points due to the weak edges in IR images and mismatches in window regions for Foerstner points.

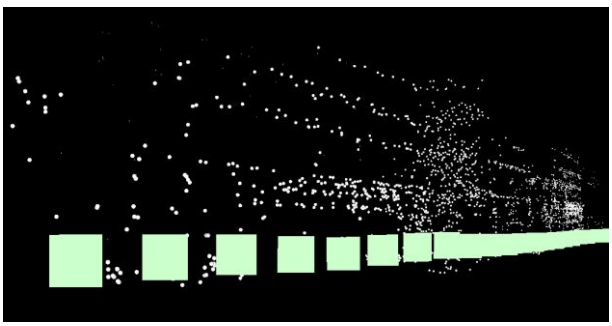

Figure 3. Point cloud of one image sequence along a group of facades. The squares are representing the estimated camera positions.

3.3 Matching with the building model

The 3d point cloud generated in the image sequence orientation step is now matched with the given building model (Fig. 4). A grouping of the points is done before the matching to remove non façade and wrong points i.e. of trees. The local neighborhood of every point is analyzed to derive an estimated plane the point is on and its normal. Every point is assigned to the surface with the smallest distance and similar normal direction. Points with a distance or normal direction that differs beyond a threshold for all facades are rejected. The remaining points are now used for the least squares matching with the facades of the building model.

Additionally, a line matching (Frueh et al, 2004) in the image space is done to refine the exterior orientation of every image (Fig. 5).

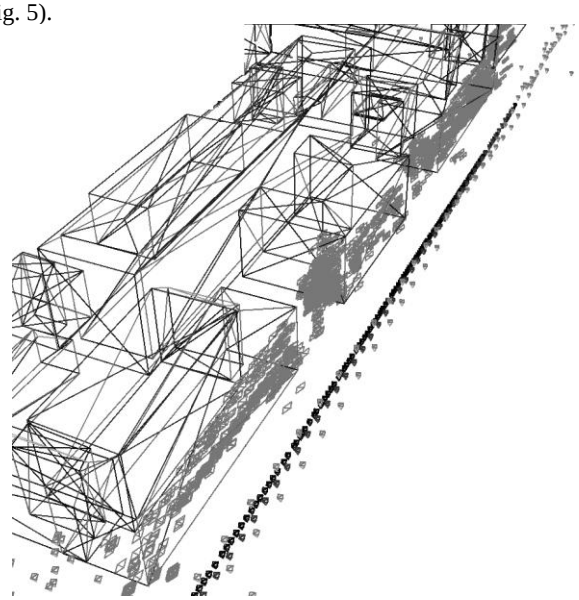

Figure 4. Grid model of the building with point cloud (light grey) and camera path: light grey: GPS path, dark grey transformed estimated camera positions

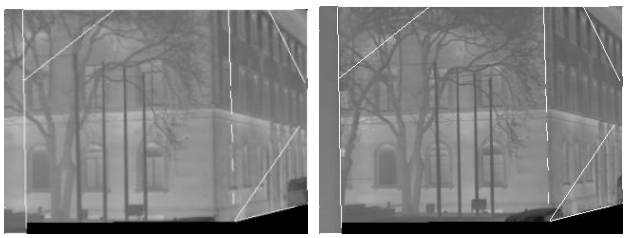

Figure 5. Image overlayed with the grid model of the building. Left: before line matching, Right: after line matching

3.4 Comparison of extracted Textures

For every surface of the model, partial textures are generated one from every image where the surface is visible. These textures normally do not cover the hole surface (Fig. 6). Due to the recording configuration, the geometric resolution decreases to the roof and the right and shows only a small part of the facade on the left.

Figure 6. Partial texture of one façade extracted from one IR image

In a first step, the partial textures of one surface from one sequence are combined (Fig. 7). The resulting combined textures show a good fitting in the middle of the images but disturbances at the roof and especially on the ground. The roof disturbances seem to be caused by the viewing angle and the low resolution in all images, the errors near the ground are caused by occlusion.

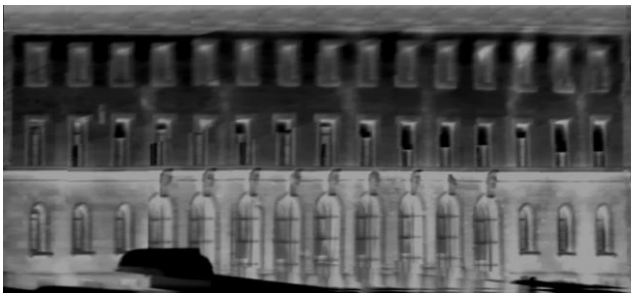

Figure 7. Surface texture generated from the complete image sequence

If a texture is associated with a building surface, we are able to compare different textures from different recording times and conditions. A straight forward way is to overlay these textures (Fig. 8). In this example two infrared textures, one from a sequence in the evening and one from a sequence in the next morning, are combined. The blue color indicated a cooling effect over night. One can see, that the position of the windows in the first and second floor are fitting very well, whereas as the third floor seems to be blurred. One can also see small diagonal lines. These lines are the result of the combination of partial textures of one sequence. This combination are not exact the same for to textures from different sequences and thus can show small differences in the intensity.

Figure 8. Temperature differences of two textures from different image sequences.

4. DISCUSSION AND CONCLUSION

The matching of the point cloud and the building model shows different behaviour depending on the building geometry. In many cases, occlusions reduce the number of visible facades in the image sequence and thus no 3d points for the matching exists. This sometime leads to remaining shifts of the point clouds in the facades. These shifts can be reduced by checking the edges of the building model against the images using projected voxels of the facade borders. It can be seen, that this method allows to handle geometric details on the facade which have not been modelled in the 3d building model itself. The geometry of the recording leads to an quite unbalanced distribution of feature points compared to the model surfaces.

The upper floor is visible only in small parts of the images, only in few images and with a very low resolution with causes blurring and thus a very low amount of feature points. A necessary improvement would be a prediction of the movement of the features. In almost every image there are enough features found from the image before to predict a movement of lost feature. This predicted features could be search in following images. This could improve the density of the point cloud and the accuracy in the bundle adjustment.
Further improvements can be achieved by combining the forward looking image sequence with a backward looking sequence to reduce occlusions and to add more images showing a specific voxel in the 3d model space.
At the moment we are working on a pssobility to integrate both bundle adjustments, the orientation step and the matching step, as the given building model with its lines should be usable as ground control points in the orientation step directly.

References

Avbelj, J., Iwaszczuk, D., Stilla, U., 2010. Matching of 3D wire-frame building models with image features from infrared video sequences taken by helicopters. PCV 2010 - Photogrammetric Computer Vision and Image Analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 38(3B): pp. 149-154

Fischler, M.A., Bolles, R.C. 1981. Random Sample Consensus: A Paradigm for Model Fit-ting with Applications to Image Analysis and Automated Cartography. In: Communications of the ACM, 24(6), pp. 381-395.

Foerstner, W. and Guelch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In: Proceedings of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data, pp. 281-305.

Frueh, C., Sammon, R., Zakhor, A., 2004. Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery, *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission* (3DPVT'04)

Haralick, R.M., Joo, H., Lee, C.N., Zhuang, X., Vaidya, V.G. and Kim, M.B. 1989. Pose estimation from correspondence point data. In: SMC 19(6), pp. 1426-1446

Haralick, R.M., Lee, C.N., Ottenberg, K. and Nolle M. 1994. Review and analysis of solutions of the 3-point perspective pose estimation problem, In: International Journal of Computer Vision 13(3), pp. 331-356.

Hartley, R.L. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision, In: Cambridge University Press, ISBN 0-521-54051-8.

Heinrichs, M., Hellwich, O., Rodehorst, V. 2008. Robust Spatio-Temporal Feature Tracking. In: Chen Jun, Jiang Jie and Wolfgang Förstner: Proc. of the XXI Congress of the Int. Society for Photogrammetry and Remote Sensing, Beijing, China, *International Archives of Photogrammetry and Remote Sensing* 37(B3a), pp. 51-56

Iwaszczuk D., Hoegner L., Stilla U., 2011. Detection of windows in IR building textures using masked correlation. In: Stilla U, Rottensteiner F, Mayer H, Jutzi B, Butenuth M (Eds.) Photogrammetric Image Analysis, ISPRS Conference - Proceedings. Lecture Notes in Computer Science, Vol. 6952, Springer: 133-146

Klingert, M.,2006. The usage of image processing methods for interpretation of thermography data ,17th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering, Weimar, Germany, 12-14 July 2006.

Lo, C.P., Quattrochi, D.A.,2003. Land-Use and Land-Cover Change, Urban Heat Island Phenomenon, and Health Implications: A Remote Sensing Approach, Photogrammetric Engineering & Remote Sensing, vol. 69(9), pp. 1053–1063

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstruction a scene from two projections, In: Nature 239, pp. 133-135.

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. In: International Journal of Computer Vision, 60(2), pp. 91-110.

Mayer, H., 2007. 3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences. In: *Photogrammetrie – Fernerkundung – Geoinformation* 2007(3), pp. 167–176.

Mayer, H. and Reznik, S., 2006. MCMC Linked with Implicit Shape Models and Plane Sweeping for 3D Building Facade Interpretation in Image Sequences. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. (36) 3, pp. 130–135.

Nistér, D., 2000. Reconstsruction From Uncalibrated Sequences with a Hierarchy of Trifocal Tensors. In: *Proc. European Conference on Computer Vision* 2000(1), pp. 649-663

Nistér, D., 2004. An efficient solution to the five-point relative pose problem, In: *IEEE Transactions on Pattern Analysis and Machine Intelligenc*e 26(6), pp. 756–777

Pollefeys, M., Nistér, D., Frahm, J. M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S. J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G. and Towles, H., 2008. Detailed real-time urban 3D reconstruction from video. In: International Journal of Computer Vision (IJCV), 78(2-3), 143-167.

Pu, W. and Vosselman, G., 2009. Refining building facade models with images. *ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation* (CMRT‘09)

Sirmacek, B., Hoegner, L., and Stilla, U., 2011. Detection of windows and doors from thermal images by grouping geometrical features, *Joint Urban Remote Sensing Event (JURSE'11)*, Muenchen, Germany.

Torr, P. and Murray, D., 1997. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. In: International Journal of Computer Vision 24(3), pp. 271-300.

Triggs, B., 1999. Camera pose and calibration from 4 or 5 known 3D points. In: Proc. International Conference on Computer Vision (ICCV'99) 7, pp. 278-284

Zhang, Z., 1998. Determining the Epipolar Geometry and its Uncertainty: a Review. In: International Journal of Computer Vision 27(2), pp. 161-195.