

## Automatic generation of façade textures from terrestrial thermal infrared image sequences.

by L. Hoegner and U. Stilla

Technische Universität München – Photogrammetry and Remote Sensing, Arcisstrasse 21, D-80333 München, Germany, [ludwig.hoegner@tum.de](mailto:ludwig.hoegner@tum.de), [stilla@tum.de](mailto:stilla@tum.de)

### Abstract

This paper discusses the automatic texturing of building facades from thermal infrared image sequences. A fully automatic method is presented to refine GPS based positions using a relative orientation of the image sequence and pre-knowledge of a given building model in a bundle adjustment process. The resulting refined orientation parameters are used to project the infrared images of the sequence onto the building façade. The resulting partial textures are combined to get complete façade textures in the thermal infrared domain. The quality of the presented method is evaluated using RGB textures as ground truth and stochastic parameters from the adjustment process.

### Introduction

The energy used by buildings is a significant part of the total energy consumed by humans. In European countries almost half of this energy is used for heating. Due to climate change, increasing energy costs, and energy performance directives, the energy efficiency of buildings should be improved. New technologies allow for the construction of buildings with high energy performance, likewise the state of old buildings needs to be inspected and improved. Buildings can be inspected remotely using a thermal infrared (IR) camera [1]. The thermal cameras capture the heat radiation of the building hull and record it as an intensity image where the intensity value corresponds to a measured temperature in a defined scale. Nowadays the thermal cameras can achieve high radiometric resolution (0.01K) which means that it is possible to capture even very small temperature differences. Thanks to this we can detect heat leakages in thermal images of a building hull. The knowledge about the position of heat leakages can be used for planning or quality control after a renovation.

Currently many IR inspections of buildings are based on single IR image analysis, without reference to the geometry of the captured scene. The images are often taken manually for a single building. However, we can also observe a trend to apply mobile mapping systems to carry out large area inspections for urban quarters [2] or entire cities [3]. The IR cameras can be mounted on a mobile terrestrial or flying platform. Terrestrial images taken from a vehicle can be used for documentation of frontal faces visible from the street level, while airborne images taken from an unmanned aerial vehicle (UAV) or helicopter can capture roofs. Using oblique view images inner yards can also be covered.

Combining IR images with 3D building information requires spatially referencing the data. Image data from different points in time, different platforms and multiple images (video sequences) can be fused. The assignment of the images to the 3D building model should be done automatically and can be achieved via texture mapping.

The texturing process using both terrestrial and airborne images is a similar task. The main difference consists of the geometry of acquisition. A terrestrial camera is typically close to the photographed object. Due to the fact that IR cameras have a detector matrix which is much smaller than cameras working in visible domain, to achieve an adequate geometric resolution of an observed object, a smaller viewing angle has to be chosen for normal video or photo cameras. Therefore, in terrestrial images it is not possible to record a complete building in dense urban areas with narrow streets in one image. Accordingly, textures have to be combined from multiple terrestrial images. In contrast airborne images are taken from a longer distance, so usually even large faces (façades and roofs) can be captured in one frame. Partial visibility of faces can be caused by self-occlusions or happens at the frame boundary. The geometric resolution of terrestrial images, which is connected to the distance from the camera to the object, is significantly higher than geometric resolution of airborne images. Hence, in terrestrial images many more details can be detected, including even single heat leakages caused by heating pipes, while in airborne images only larger elements can be recognized.

In recent years the process of automatic texture mapping using terrestrial or airborne images has been frequently discussed within photogrammetry and computer vision. The major problems connected to the automation of the texturing are:

- direct geo-referencing,
- self-calibration of the camera,
- model-to-image matching
- automatic visibility checking and self-occlusions,
- elimination of extrinsic occlusions,
- selection of the best image for the texture from an available set of images.

### 1.1. Geo-referencing

Geo-referencing of images consists in determination of the exterior orientation parameters of the camera. Approximated but stable position and orientation of the sensor can be directly determined using global positioning system (GPS) data. An inertial navigation system (INS) provides good short-term accuracy, but over a longer time a systematic drift occurs. Thus, the combination of GPS and INS makes it possible to avoid the INS drift and to bridge any loss of satellite signal by GPS [4]. Applying direct geo-referencing a system calibration needs to be carried out. In particular the camera position and orientation are often not identical to the position and orientation registered by integrated GPS/INS, in which case the estimation of the misalignment angles (boresight parameters) and the lever arm vector is necessary [4,5,6,7]. Furthermore, a transformation between coordinate systems is necessary, since most of building models are stored in national coordinates, while GPS/INS navigation uses a geographic coordinate system [8].

### 1.2. Camera Self-calibration

Usually the accuracy of direct geo-referencing is not high enough for precise texture mapping with high-resolution images. Furthermore, the camera interior orientation parameters are often unknown or only approximately known. These parameters can be determined a priori in a camera calibration process or in a self-calibration process together with the exterior orientation parameters. The a priori calibration is done using a known calibration pattern in an adjustment process of a series of images with known calibration marks [9]. For the self-calibration the main challenges are full automation and accuracy of the camera parameter estimation. In the majority of cases in computer and machine vision projective geometry is applied. It provides linear solutions which are quite simple and fast but have difficulty dealing with non-linear parameters such as lens distortions. In contrast, in the photogrammetric community usually perspective geometrical models are used. They are based on a system of non-linear collinearity equations which are solved by means of the bundle adjustment. In literature many publications discussing automatic self-calibration can be found [10,11].

For self-calibration metric information from the object space is needed. Usually in photogrammetry control points with known ground coordinates are selected and embedded into the bundle adjustment process. In texture mapping the coordinates of already existing 3D building models can be used as real world metric reference. The image must be matched with the projected model to deliver the necessary input for the calibration.

### 1.3. Model-to-image Matching

Various methods for model-to-image matching have been developed. Frueh et al. [12] proposes an approach based on matching line segments with model edges. In this method the edges are extracted in the image and the model is projected into the image from random camera positions. For each position a rating based on line matching is calculated and the procedure is repeated. The position with the highest rating is chosen for texture mapping. In this approach no correspondence between frames is used. For every frame many random parameter configurations have to be tested, thus the computational effort of this method is very high. Some other authors [13,14,15] assume that considered scenes consist of piece-wise planar surfaces with dominant directions (so called Manhattan or Legoland scenes) and calculate vanishing points of the vertical and horizontal lines. This, in combination with GPS data can be used for ExtOri parameter determination. These methods require a calibrated camera and many vertical and horizontal lines extracted in the image. Because of lacking vertical and horizontal lines which can be unambiguously extracted, these methods can fail in residential areas [13].

Stilla et al. [16] and Avbelj et al. [17] propose a method for matching of low resolution IR images based on intersection points of roof edges. Roof edges are selected, because in IR images these can be reliably extracted. This is related to the operation principle of the IR cameras, which record the thermal radiation of the photographed objects. This radiation depends on objects temperature and emissivity, and the emissivity depends on the material. Building façades are usually made of concrete, similar to pavement and streets and often the contrast between them is very low. In contrary, roofs are usually made of other materials, such as roof tiles or sheet metal, and can be more easily distinguish from the façades.

In close range applications such as terrestrial taken images of buildings only parts of building may be visible in one image. In these cases it is hard to find a unique solution for the model-to-image matching. A solution is the matching of a whole image sequence based in a relative orientation of the sequence and a structure-from-motion reconstruction of the camera orientations and scene. Feature points from multiple images can be used for pose estimation. Nistér [18] introduces an algorithm which uses corresponding SIFT feature points [19] in two calibrated views of a scene and calculates the essential matrix of these two images from five corresponding points to find a relative camera motion between and allows the iterative combination of hundreds of views. Further extensions of this algorithm towards the handling of possible wide-baseline image sequences have been achieved by [20,21,22]. Mayer [20] adopts Nistér's algorithm for facade extraction and texturing from multiple views of unordered sets of images. The robustness of feature descriptors to perspective changes is a key point here. Heinrichs [21] uses a set of three cameras to calculate real stereo and trifocal tensor information and included this into a structure-from-motion strategy. Pollefeys [22] presents a method for an ordered sequence of images from a moving vehicle.

#### 1.4. Visibility Analysis

For texture mapping a visibility analysis is necessary. In general, two groups of methods for visibility checking can be distinguished: (i) image based methods and (ii) polygon-based methods. In image-based methods z-buffer (depth image) or ray casting algorithms are applied. In contrast, in polygon-based (vector-based) approaches polygon clipping is used. Typically, all polygons (triangles) are projected into the image plane and. The depth-buffer method is a basic method for removing hidden surfaces adopted from computer graphics. The depth-buffer is a matrix storing the distance from the projection centre to the model surface for every pixel. This method was often proposed in some variations [12, 23]. In ray casting techniques the (depth) image is generated by tracing the rays sent from the projection centre, through the pixels of the image plane, to the intersection with a surface of a 3D scene [2,24] integrate digital images and laser scanning point clouds. They use a multi-layer 3GImage algorithm which classifies the visibility on two stages: point stage and surface stage.

Using visibility checking only self-occlusions can be detected, but for high quality texture mapping, especially when in the texture certain objects, such as windows or heat leakages need to be detected, "non-predictable" (un-modeled) occlusions have to be taken into account. For this purpose multiple images [24,25] or laser point clouds [26] can be used.

In summary, many various methods were developed and tested in existing researches on automatic texturing. Most of them were created within computer vision for good user perception. For most of them the precision of the match between the model and the image does not play a deciding role. Similarly according to the completeness of the textures, some authors suggest accepting textures with a few occluded pixels instead textures with very low resolution (taken from extremely oblique views). For the thermal inspection occluded textures means missing data, so the missing pixels should be populated even with lower resolution frames.

The presented work focusses on terrestrial scenes recorded with image sequences of a thermal infrared camera taken from a moving system. Because of the low geometric resolution of thermal cameras façade textures are composed of a big set of partial textures extracted from the images of the sequence. To fit the textures to a given building model a coregistration is done. As single images of the sequence show only small parts of the façade and facades in the 3d model consist only of the surrounding polygons of the facades a direct line matching is not suitable in this case. A relative orientation and structure-from-motion method is chosen and combined with the pre-knowledge of the building model and recorded GPS orientation parameters of the camera within a bundle adjustment process to find the best fit of the image sequence and the model.

#### 1. Methodology

In this paper, strategies for 3d reconstruction from image sequences are extended to deal with a given building model and recorded GPS camera path in a global coordinate system to match the image sequence on an existing building model. A relative orientation is calculated for every two adjacent images of the sequence from feature points. SIFT features have turned out as the most flexible and stable features for different scenarios.

In contrast to conventional IR inspection of buildings, for texturing the whole building or a building complex in dense urban areas and their narrow streets, the low resolution and the small field of view avoid a direct line matching of edges in the images and edges of the given building model. The proposed concept is based on the assumption that a 3d model of the recorded building is given containing 3d vertex coordinates and triangulated polygon surfaces with given texture coordinates. The coordinates should be given in a national coordinate system like Gauss-Krueger. GPS coordinates are often inaccurate and a direct line matching of the images and the projection of the model's edges fails because of the lack of visible façade edges in many of the images. In former works we introduced a line matching strategy for infrared image sequences [2]. 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 [20,21,22]. Instead of generating an isolated only relative oriented model, this estimation is matched with the pre-known building model and measured camera path. Several matching steps are performed: At first, a camera orientation is interpolated for every image of the sequence from the recorded GPS/INS data. Starting with the first image pair, homologous points are tracked through the images. Given the camera orientation and known camera calibration, the observed image coordinates are used to calculate 3d coordinates in a bundle adjustment. An additional constraint is introduced. Estimated 3d points of a façade should be on or close to a polygon face of the given 3d model. This constraint is added in the adjustment to minimize the distance of the extracted 3d points to the polygon surface and remove outliers. After this step, the corrected camera path and orientation are used to project the images of the sequence into the 3d model space. To be able to combine several images of a façade for the final façade texture, the resolution of the outgoing texture is defined first. Caused by the strict order of the sequence partial occlusion of facades have to be taken into account. This procedure is repeated for every visible façade in every input image of the sequence. Additionally, for every pixel of the texture, the geometric resolution of the pixel's value is stored.

In the end, for every façade a textures are generated for every image where the façade is visible in. The resulting complete texture is calculated by merging the pixel values of the partial textures. This is achieved by copying one over another in the order of the time stamps. Using an oblique view for the recorded sequence, every visible pixel of a texture has a higher geometric resolution than the texture before. To overcome remaining positioning errors, a line matching between the partial textures is included in this step.

## 2.1. Orientation of the Image Sequence

Mayer [2] has introduced an approach for wide-baseline image sequences. Given the known inner orientation of the camera, SIFT features [19] are matched via cross-correlation. RANSAC [27] is used to choose SIFT features for the estimation of the fundamental matrix  $F$  and trifocal tensor  $T$  of triplets of adjacent images of the sequence. The found inliers are used for a robust bundle adjustment [28]. To orient the whole image sequence, the triplets are linked based on homographies and already known 3d points of the already oriented sequence part. Figure 1 shows an image out of a sequence with detected homologues points and estimated point movement in the image space.



**Fig. 1.** IR image with selected SIFT features, that have correspondences in the following image. Arrows show the moving direction of the points

This method is based only on the images themselves and does not contain information of the orientation of the camera. So neither the global position with respect to a given building model is known, nor the scale of the image sequence. By introducing recorded orientation parameters from i.e. a GPS system, the images are transferred into the global coordinate system. This allows including pre-knowledge to the process. A 3d point reconstructed from homogenous pixels of a set of images that show a part of a building façade must be on or close to the model façade polygon. If it is not it is either the result of a wrong point pair or it does not belong to the façade but i.e. to a tree or car standing in front of the building.

The object coordinate estimation from more than two homologues points is an error minimization problem that is formulated here as least mean squares error in a bundle adjustment. All homologues image points  $i$  of a 3d point  $k$  are given as observations  $(x_{ik}, y_{ik})$ . The 3d coordinates of the point are set as unknowns with an initial estimation from the homologues points of the first two images. Because of the limited accuracy of the image coordinates, the two projection rays of the homologues points will not intersect in 3d space. Instead the center point of the connection of the smallest distance of the two rays is chosen. The camera calibration parameters and the camera orientation are given as observations but with higher standard deviation compared to the image coordinates of the homologues points. This is done to allow a refinement of the GPS orientation parameters and to minimize the projection errors.

The model knowledge is included as additional observations where the distance of the 3d point to the closest surface of the model is minimized. In the bundle adjustment process, the first partial derivative of the calculations of the observations from the unknown derived for every unknown is set to zero. An image point  $(x_i, y_i)$  is calculated from an object point  $(X_i, Y_i, Z_i)$  using the collinearity equations

$$\begin{aligned} x_i - x_0 &= c_x \cdot \frac{a_{11}(X_i - X_0) + a_{21}(Y_i - Y_0) + a_{31}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + dx' \\ y_i - y_0 &= c_x \cdot \frac{a_{12}(X_i - X_0) + a_{22}(Y_i - Y_0) + a_{32}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + dy' \end{aligned} \quad (1)$$

Where  $(X_0, Y_0, Z_0)$  is the camera position,  $(a_{11}$  to  $a_{33})$  is a 3x3 rotation matrix of the viewing direction of the camera, and  $(x_0, y_0, dx', dy')$  are the parameters of the geometric camera calibration.

Using the Hessian form of a plain

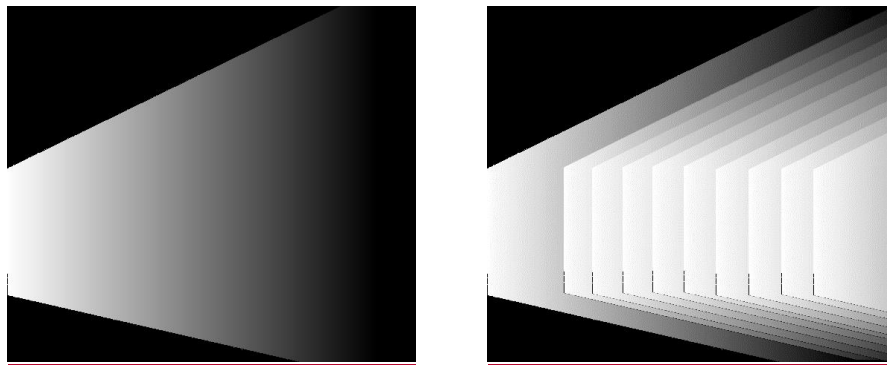
$$\vec{r} \cdot \vec{n} - d = 0 \quad (2)$$

A 3d point is on the plain if its space vector  $\mathbf{r} = (X_i, Y_i, Z_i)^T$  fulfills equation 2.

The weights for the observations are adapted in every iteration step of the bundle adjustment with respect to the remaining distance to the closest model façade as points with high distances are either wrong homologues points or 3d points that are not on the façade. As a result, the orientation parameters of the images are optimized so that the extracted 3d points fit best to the given 3d building model. The result is the best fit, not the most correct 3d positions of the cameras. But the target here is to find the best projection of the images onto the model and not to reconstruct the most correct camera positions.

## 2.2. Texture extraction

The texture extraction part is split into three steps. In the first step, based on the results of the orientation, the images are projected onto the building model. This is done using equation 1 reverse and calculating the cross point of the projection ray of every pixel of every image with the model. In a second step, based on every projected image a partial texture for every visible surface of the model is generated. Because of the perspective projection the geometric resolution of an image onto a surface of the model might be not constant. In the examples introduced in section 3, an oblique forward looking view was chosen. In this case, the geometric resolution is as seen in figure 2a. Defining a unique texture size for all partial textures of a surface, one can see that for an ordered set of input images every following partial texture only shows parts of the surface, where it has a higher geometric resolution than the already combined partial textures (Fig. 2b). This allows a very simple combination strategy. For every surface every projected image is used for a bilinear interpolation of a partial texture with the same size in pixels. Then the pixels of the partial textures that are given a value from the interpolation – which means that this part of the surface was visible in the image – are copied in recording order into the final texture. A correlation is done for every added texture to remove remaining positioning errors. In the end, every pixel of the final texture is given the value from the input image with highest geometric resolution for this pixel.



**Fig. 2a.** Schematic view of the geometric resolution of a partial texture taken from an image in oblique forward looking view. white: high resolution, grey: low resolution, black: not visible in the image **2b.** Geometric resolution of a surface textures from a set of partial textures

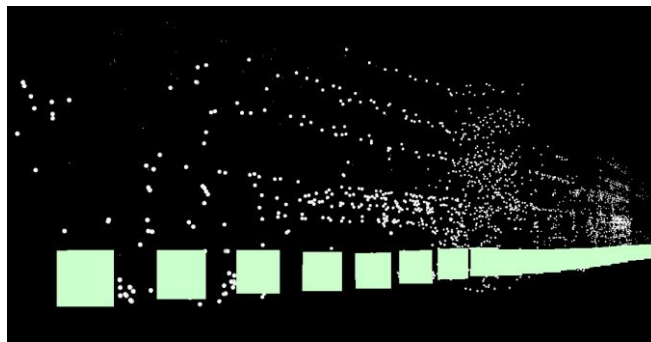
## 1. Experimental results

Our experiments have been taken using a SC3000 thermal infrared camera with a detector of 320x240 pixels and a temperature resolution of 0.02 K. Image sequences have been taken before sunrise. Figure 3 shows some image out of a sequence showing a part of a building complex. The sequence was recording with 50 frames per second. This allows a very good feature tracking (Fig. 1) due to only small movements between the images. For the 3d reconstruction every 10<sup>th</sup> image was taken to reduce the computational effort and guarantee a 3d base necessary for the initial 3d point estimation.

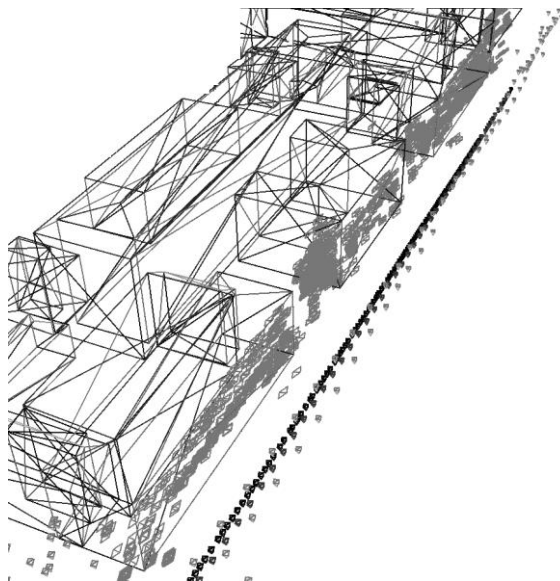
The reconstructed 3d coordinates and camera positions after the bundle adjustment can be seen in figure 4. A high number of homologues points is used to calculate 3d points. Points are grouped along the window rows where the thermal images show the highest gradients. In the center the tree is extracted. Its points are not used for the orientation improvement but are calculated in the bundle adjustment. A combination of the 3d points and the given building model is shown in figure 5. On the facades are the extracted points. The points of the tree are standing alone. In front, there are different camera paths. The bright path the right is the original uncorrected GPS path. The second darker path is the corrected GPS positions and the dense dark path is the final interpolated position for every image of the sequence.



**Fig. 3.** Four images taken out of a sequence of the SC3000 camera. Only few and incomplete parts of corners are visible.

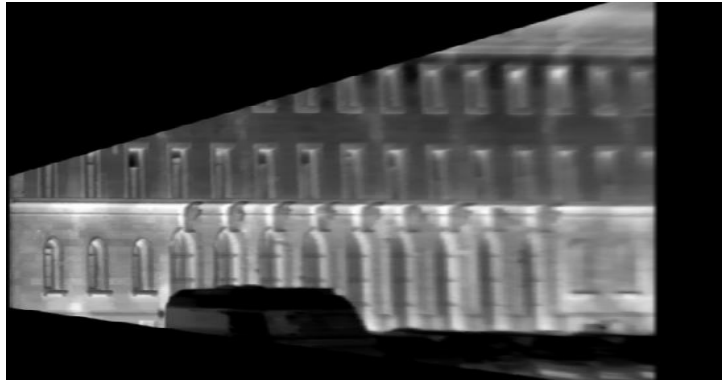


**Fig. 4.** Extracted 3d points and camera positions. Points are mainly grouped along the windows. The points of the tree (center) are also extracted.

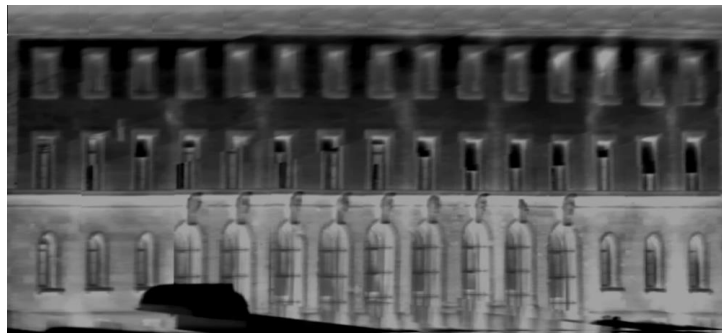


**Fig. 5.** Extracted 3d points and camera positions together with the building model

Figure 6 shows one example texture. As this texture is extracted from an oblique view, the spatial resolution is not constant. It is getting lower from the left to the right. Due to this the texture is blurred at the right. On the other hand, on the left side only the first and parts of the second floor are visible. The third floor is visible only in the low resolution right part. A correlation is done for every added texture to remove remaining positioning errors. Combining several of these partial textures leads to the texture given in figure 7. The total façade is visible in the texture but with different geometric resolutions. For the first and second floor the resolution is higher than for the third floor. The big bus standing in front of the building could only be removed to a small part because it was occluding the façade in all images.



**Fig. 6.** Partial texture interpolated from one projected image. The field of view increases with the distance to the right. The resolution decreases to the right.



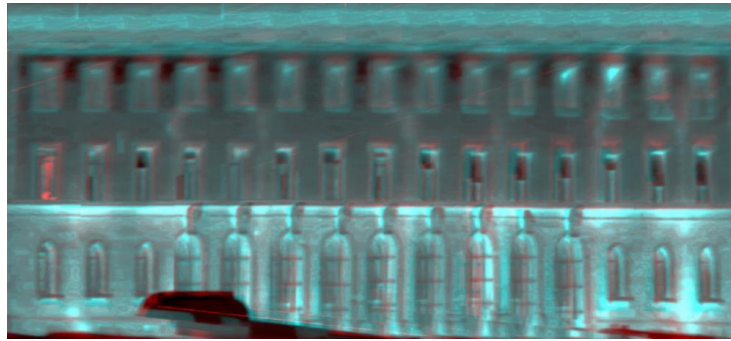
**Fig. 7.** Surface texture generated from the complete image sequence

## 2. Discussion and Outlook

Different aspects influence the quality of the extracted textures. The accuracy of the 3d reconstruction is limited compared to image sequences in the visual spectrum. One aspect is the lower image resolution that causes a higher discretization of the scene and reduces the number of details that can be matched within the sequence. On the other hand the radiometric behavior of object in the thermal infrared leads to blurred edges and a low number of intensity changes and details. Homologues points are rarer than in the visual spectrum and worse locatable. A second aspect is the geometric calibration of the cameras. Because of the two limitations already mentioned, the calibration accuracy is also limited compared to cameras in the visual spectrum.

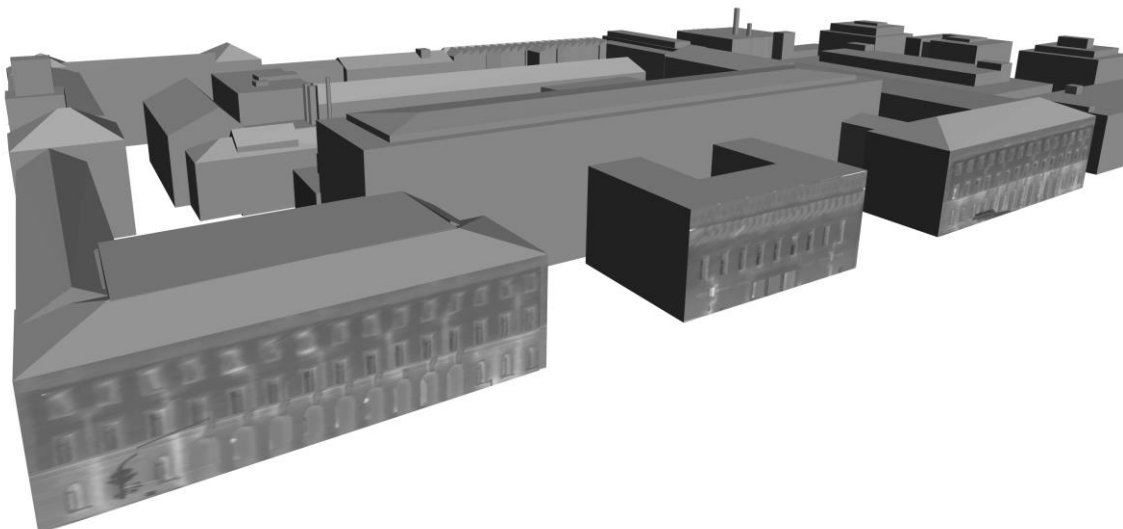
The Quality of the initial GPS positions is limited in urban scenarios caused by mirroring and occlusion of the GPS signals. The initial values have quite an influence on the final results as initial assignments of homologues points to facades influence the weightings in the iterative process. The quality of the building model itself is also a limiting factor. Building in level-of-detail 1 are only block models. In these models, the overhang at the roofs is not modelled. This causes two possible errors. If no overhang is modelled, the facades can either be set to the position of the footprint. Then, the position of the facades is correct, but the overhang is projected onto the façade. This causes misalignments of the objects on the façade like windows and heating pipes. If the building model is generated from aerial images, the roofs outline is taken as footprint and the position of the facades is incorrect in the model. In this case, the bundle adjustment tries to move the camera path which very often works but not necessarily. The number of visible elements in the thermal infrared influences the quality of the position refinement. On facades with only few objects or only repetitive patterns, the quality of the refinement is significantly reduced. If this falls together with bad initial GPS positions, the method can totally fail.

The texture extraction is quite sensitive to errors in the viewing direction estimation. The combination of relative orientation and 3d model knowledge reduces these errors significantly. Remaining errors between to partial textures. The accuracy of the texture extraction can be seen in figure 8. Two textures are generated from two different sequences. One sequence is taken in the evening and the other one in the morning. Cyan color shows a cooling during the night, red color a warm up. You can see that both textures fit quite good. Errors occur from small distortions to the upper right. This is caused by the low resolution in the images at the right. Features are the right are harder to be tracked than in the left. As the upper floor is only visible in the right parts of the images, the location of points is worse and the number of points is lower. Remaining errors in the bundle adjustment are concentrated here because the few points have lower influence on the parameter than the many points in the first and second floor. On the left side a big change in the second floor can be seen. Here, a window was opened to simulate a change on the facade.



**Fig. 8.** *Overlay of two textures from two sequences taken in the evening and in the morning. Cyan represents a cooling down during the night, red a heating up.*

Further steps in the processing of the textures are windows extraction, leakage detection and parameter estimation for facades. The given method can be used to record the building facades in whole streets to get first view on the quality of the thermal behavior of the buildings. Searching objects like windows, heating pipes, or leakages in building textures instead of images allows a 3d geo-referencing and observation of thermal relevant objects i.e. in a building information model. The evaluation of time series and changes i.e. before and after a thermal optimization, summer vs. winter, evening vs. morning are possible in an automatic way. Figure 9 shows finally the textured facades from a sequence onto a given building model. For non-visible facades, an additional record with oblique backward view is possible. Further occlusions can be removed by introducing aerial images from i.e. RPAS systems.



**Fig. 9.** *Building model with textures extracted from one sequence*



### Acknowledgment

Our work is part of the DFG (German Research Society) research project “Enrichment and Multi-purpose Visualization of Building Models with Emphasis on Thermal Infrared Data” as part of the bundle project “Interoperation of 3D Urban Geoinformation (3DUGI)”. The authors thank Dr. Clement, Dr. Schwarz and Mr. Kremer of Fraunhofer IOSB, Ettlingen, for their assistance during the recording campaign.

### REFERENCES

- [1] C.A. Balaras and A.A. Argiriou, 2002. Infrared thermography for building diagnostics”, *Energy and Buildings*, Vol. 34: 171-183
- [2] L. Hoegner, H. Kumke, L. Meng, U. Stilla, 2007. Automatic extraction of textures from infrared image sequences and database integration for 3D building models. In: PFG *Photogrammetrie Fernerkundung Geoinformation*. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung vol. 2007(6), pp. 459-468
- [3] D.L. Chandler, 2011. The big picture on energy loss, MIT news, March 16, 2011. Massachusetts Institute of Technology. Access on June 30, 2011 <http://web.mit.edu/newsoffice/2011/ir-scanning-energy-0316.html>
- [4] N. Yastikli, K. Jacobsen, 2005. Direct sensor orientation for large scale mapping – potentials, problems, solutions”. *The Photogrammetric Record* 20(111), September 2005: 274-284
- [5] H. Eugster, S. Nebiker, 2007. Geo-registration of Video Sequences Captured from Mini UAVs – Approaches and Accuracy Assessment, 5th International Symposium on Mobile Mapping Technology - Padua, Italy, 2007
- [6] U. Stilla, J. Kolecki, L.Hoegner, 2009. Texture mapping of 3D building models with oblique direct geo-referenced airborne IR image sequences. ISPRS Hannover Workshop 2009: High-resolution earth Imaging for geospatial information. 38(1-4-7/W5)
- [7] J. Kolecki, D. Iwaszczuk, U. Stilla, 2010. Calibration of an IR camera system for automatic texturing of 3D building models by direct geo-referenced images, Eurocow 2010, Spain
- [8] J. Skaloud, K. Legat, 2008. Theory and reality of direct geo-referencing in national coordinates”. *ISPRS Journal of Photogrammetry & Remote Sensing* 63 (2008: 272-282)
- [9] T. Luhmann, J. Ohm, J. Piechel, T. Roelfs, 2010. Geometric Calibration of Thermographic Cameras. In: Mills, J. P.; Barber, D. M.; Miller, P. E.; Newton, I. (Eds.), *Proceedings of the ISPRS Commission V Mid-Term Symposium 'Close Range Image Measurement Techniques'*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 5), 411-416.
- [10] T.A. Clarke, J.G. Fryer, 1998. The development of camera calibration methods and models. *The Photogrammetric Record*, Vol. 16(91), 51-66
- [11] F. Remondino, C. Fraser, 2006. Digital camera calibration methods: Considerations and comparisons”. *ISPRS Commission V, WG V/1*
- [12] C. Frueh, R. Sammon, A. Zakhor, 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)*, 2004
- [13] M. Ding, A. Zakhor, 2008. Automatic registration of aerial imagery with untextured 3D LiDAR models, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, June 2008
- [14] T.J. Cham, A. Ciptadi, W.C. Tan, M.T. Pham<sup>1</sup>, L.T. Chia, 2010. Estimating Camera Pose from a Single Urban Ground-View Omnidirectional Image and a 2D Building Outline Map”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, USA
- [15] W. Foerstner, “Optimal vanishing point detection and rotation estimation of single image from a legoland scenes”, *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, Vol 38(3B): 157-162, 2010
- [16] U. Stilla, U. Sörgel, K. Jäger, 2000. Generation of 3D-city models and their utilisation in image sequences, *International Archives of Photogrammetry and Remote Sensing*. Vol. 33, Part B2, 518-524, 2000
- [17] J. Avbelj, D. Iwaszczuk, U. Stilla, 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): 149-154
- [18] D. Nistér, 2004. An efficient solution to the five-point relative pose problem, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6), 756–777
- [19] D. Lowe, 2004. Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision*, 60(2), 91-110.
- [20] H. Mayer, 2007. 3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences. In: *Photogrammetrie – Fernerkundung – Geoinformation* 2007(3), pp. 167–176.
- [21] M. Heinrichs, O. Hellwich, V. Rodehorst, 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), 51-56

- [22] M. Pollefeys, D. Nistér, J.M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénus, R. Yang, G. Welch, H. Towles, 2008. Detailed real-time urban 3D reconstruction from video. In: *International Journal of Computer Vision (IJCV)*, 78(2-3), 143-167.
- [23] G. Karras, L. Grammatikopoulos, I. Kalisperakis, E. Petsa, 2007. Generation of Orthoimages and Perspective Views with Automatic Visibility Checking and Texture Blending. *Photogrammetric Engineering & Remote Sensing*, Vol. 73, No. 4, April 2007: 403-411
- [24] A. Abdelhafiz, W. Niemeier, 2009. Integration of Digital Images and Laser scanning Point Clouds - Texture Mapping with 3DImage. 91. DVW-Seminar Terrestrial laser scanning TLS2009, ISBN 978-3-89639-734-8, Fulda, Germany, November 19-20. *Journal of DVW* 60/2009.
- [25] J. Boehm, 2004. Multi-Image Fusion for Occlusion-Free Façade Texturing. *IAPRS&SIS*, vol. XXXV (Part B5), Istanbul, Turkey, 867-872.
- [26] S. Bénitez, E. Denis and C. Baillard, 2010. Automatic production of occlusion-free rectified façade textures using vehicle-based imagery. *PCV 2010 - Photogrammetric Computer Vision and Image Analysis. International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 38(3A): 275-280
- [27] M. Fischler, R. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM* 24: 381-395
- [28] R.L. Hartley, A. Zisserman, 2004. *Multiple View Geometry in Computer Vision*", Cambridge University Press, ISBN: 0521623049