Liu and Seipel *Visualization in Engineering* (2015) 3:16
DOI 10.1186/s40327-015-0028-0

Visualization in Engineering
a SpringerOpen Journal

RESEARCH Open Access

Infrared-visible image registration for augmented reality-based thermographic building diagnostics

Fei Liu[1,2*] and Stefan Seipel[1,2]

Abstract

Background: In virtue of their capability to measure temperature, thermal infrared cameras have been widely used in building diagnostics for detecting heat loss, air leakage, water damage etc. However, the lack of visual details in thermal infrared images makes the complement of visible images a necessity. Therefore, it is often useful to register images of these two modalities for further inspection of architectures. Augmented reality (AR) technology, which supplements the real world with virtual objects, offers an ideal tool for presenting the combined results of thermal infrared and visible images. This paper addresses the problem of registering thermal infrared and visible façade images, which is essential towards developing an AR-based building diagnostics application.

Methods: A novel quadrilateral feature is devised for this task, which models the shapes of commonly present façade elements, such as windows. The features result from grouping edge line segments with the help of image perspective information, namely, vanishing points. Our method adopts a forward selection algorithm to determine feature correspondences needed for estimating the transformation model. During the formation of the feature correspondence set, the correctness of selected feature correspondences at each step is verified by the quality of the resulting registration, which is based on the ratio of areas between the transformed features and the reference features.

Results and conclusions: Quantitative evaluation of our method shows that registration errors are lower than errors reported in similar studies and registration performance is usable for most tasks in thermographic inspection of building façades.

Keywords: Multimodality image registration, Augmented reality, Thermal infrared imaging, Façade

Introduction

Due to diverse mechanisms of different kinds of imaging sensors, multimodal images are capable of characterizing various distinct properties of a scene. Through image registration, these properties are integrated to provide us with a thorough understanding of the scene represented. Such a benefit has made multimodality image registration a vital process in fields like remote sensing (Dawn et al. 2010; Le Moigne et al. 2011) and medicine (Oliveira and Tavares 2014). Visible images captured by standard digital cameras have high spatial resolution and are rich in visual details. However, their quality is greatly degraded when there is no sufficient lighting. On the other hand, thermal cameras work with the heat radiation emitted by objects so they do not require visible illumination to function; instead daylight is often avoided when thermal imaging is used for diagnostic purposes. Among the drawbacks inherent to thermal images are low spatial resolution, poor quality and lack of visual detail (Choi et al. 2011; Morris et al. 2007; Prakash 2000). The respective pros and cons of thermal and visible images make them complement each other very well so both modalities are widely employed in building diagnostics (Balaras and Argiriou 2002; Gade and Moeslund 2014; Kylili et al. 2014) and surveillance systems for human detection and tracking (Gade and Moeslund 2014; Kong et al. 2007; Kumar et al. 2014).

Augmented reality (AR) technology enables users to view the real world enhanced with computer-generated

*Correspondence: feiliu@hig.se
[1] Department of Industrial Development, IT and Land Management, University of Gävle, 80176 Gävle, Sweden
[2] Centre for Image Analysis, Uppsala University, 75105 Uppsala, Sweden

information. In recent years the applications of AR in the architecture, engineering, construction and facility management (AEC/FM) industry have been constantly explored (Behzadan et al. 2015; Chi et al. 2013; Wang et al. 2013). When adopted to thermographic building diagnostics, AR can provide a valuable and intuitive tool for integrating images from a thermal infrared (TIR) survey into the real architectural context. Such an integration frees inspectors from mentally comparing information presented in more than one medium and thus facilitates the purpose of identifying insulation deficiencies in buildings or localizing failures of structural and functional building components that are not directly visible to naked eyes. The key challenge to be overcome by such an AR system, as illustrated in Fig. 1, is to register images of building façades taken at different times with two very different modalities. This paper focuses on a novel method for registering TIR and visible images of façades in order to tackle the challenge.

Related works

The goal of image registration is to estimate a transformation model (a mapping function) which can align sensed images with a reference image. Registering images of different modalities, in this case, TIR and visible, is challenging because both modalities represent identical objects in a scene disparately in terms of, e.g., intensity levels, gradients and textures. Additionally, as mentioned before, TIR images have specific shortcomings in comparison with visible images, namely, low spatial resolution, poor quality (noise) and scarce visual details. There are two major categories of approaches towards image registration in general, area-based and feature-based (Zitova and Flusser 2003). The area-based approach computes pixel-wise similarity measure within a portion of the images or the entirety of them to determine the registration quality. Concerning multimodality image registration, mutual information (MI) is the most prominent representative of this category and has been applied successfully in medical image registration (Pluim et al. 2003). However, because TIR images have low spatial resolution and contain far fewer visual details compared with their visible counterparts, images of these modalities are less statistically dependent and MI-based methods are thus likely to fail in such a registration scenario (Keller and Averbuch 2006). Additionally, a façade typically exhibits a repetitive layout. Such a nature can cause a misalignment modeled by, e.g., a horizontal or vertical translation (namely, shifting the sensed image in one direction) to be considered as a correct registration since the overlapping parts of both images are still quite similar thus obtaining a high MI value. The feature-based approach, on the other hand, relies on detecting salient, distinctive objects (termed features) in images and then matching them to estimate the transformation model. A plethora of literatures regarding multimodality image registration in this category exist in the field of remote sensing, where images of the same scene taken by various sensors need to be registered for further analysis. We found that features based upon Scale Invariant Feature Transform (SIFT) (Lowe 2004) were the most studied ones in those literatures (Kupfer et al. 2013) because of SIFT's invariance to rotation, scaling and partial invariance to viewing point difference. However, as (Kupfer et al. 2013) points out, all these SIFT-based methods attempt to improve the feature matching quality of the standard SIFT. The standard SIFT does not perform well in multimodal remote sensing images because the gray-level values of the same object in different modalities can vary significantly or even be reversed in some cases (Li et al. 2009; Yi et al. 2008). The change in the layout of gray-level values is likely to alter pixel gradients, giving rise to disparate SIFT feature descriptors. Figure 2 clearly demonstrates these limitations for images used in our study. Although these proposed SIFT-based variants do improve the registration results in the field of remote sensing, their applications in multimodality image registration involving the TIR band are rarely reported.

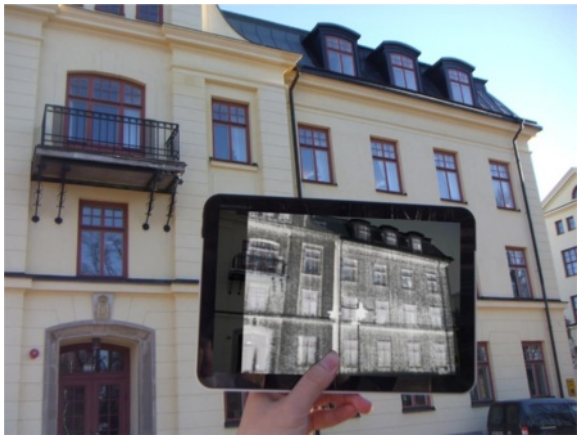


Fig. 1 Conceptual illustration of the AR system for infrared-thermographic building diagnostics

Owing to dissimilar gray-level values of the same scene object and greatly reduced texture details in the TIR band, most existing TIR/visible image registration methods take advantage of features derived from substantial discontinuities in gray-level values, for instance, edge line segments and corners. Dana and Anandan (1993) presented an edge map-based coarse-to-fine approach to infrared/visible image registration. Instead of measuring the similarity of the two edge maps, edge line segments are grouped to form triangle features for estimating the transformation model in Coiras et al. (2000). Without exploiting the knowledge of underlying geometry of image

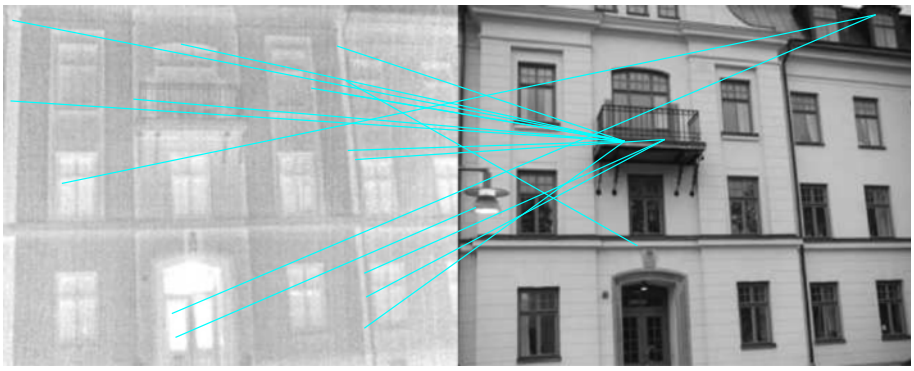


Fig. 2 Corresponding SIFT features between a TIR image and its visible counterpart

contents, all possible pairs of triangles are simply evaluated through a quality function based on the angles and the center-to-center distance between the transformed edge line segments and the ones in the reference image. Han et al. (2013) proposed hybrid visual features for the registration. Their method consists of dominant edge lines for global transformation estimation and Harris corner points (Harris and Stephens 1988) with maximal response for subsequent local transformation refinement. In Hrkać et al. (2007), the authors studied a similar problem as presented herein. They adopted Harris corner as the registration feature and assumed a similarity transformation model. Possible feature matches (hypotheses) are established by searching corners within a circular area in the reference image. The k^{th} smallest partial Hausdorff distance (Huttenlocher et al. 1993) between the transformed corners and the reference corners is used to select the best hypothesis.

Façade quadrilateral features

Following the observation that windows are typically ubiquitous elements on a façade (Friedman and Stamos 2013; Mesolongitis and Stamos 2012), we propose quadrilateral features which are derived from rectangular façade elements and a hypothesis-evaluation framework for estimating the transformation model. Our method draws on image edge line segments and perspective information to generate the quadrilateral features, as edges are likely to be preserved across these two different modalities (Coiras et al. 2000; Dana and Anandan 1993). In order to establish feature correspondence between two images, we assume the façade under scrutiny is the major content in the view and at the same time images of both modalities are taken from roughly similar viewing positions. This assumption is reasonable because the purpose for the registration is thermal inspection of buildings and naturally the focus should be a façade or a portion of it. Although visible and IR cameras tend to have different fields of view, through gathering test image data, we found it feasible in practice to capture TIR/visible images both emphasizing the same portion of a façade. Besides, standing in front of a façade, we often do not have many very different viewing positions to choose from due to road obstructions and occlusions from trees. Under the aforementioned assumption, the features can be hypothetically matched based on their spatial distances and geometric properties (e.g., aspect ratio). The estimation of transformation models is governed by a simple forward selection algorithm: at each of the total n steps, where n is the number of feature pairs we desire, a feature pair is added to the current set of selected corresponding features such that this set produces the best registration quality. The quality measure is computed based on the overlapping areas between the transformed features and the features in the reference image. Figure 3 shows an overview of the registration method.

The difference between our method and the previous works is that we take one step further by grouping low-level edge line segments into geometric entities that model the shape of major façade elements. Through utilizing these entities as higher-level features for registration, more knowledge is gained to help with the feature matching and the design of transformation quality measure, which engenders a promising solution to the challenging cross-modality registration problem.

Detection of quadrilateral features on façade

Since rectangular structures on a façade, e.g., windows and doors, comprise mostly horizontal and vertical line segments, we started with detecting these two groups of line segments on an image. Under a perspective distortion, a pencil of parallel lines converges on a vanishing point. We followed the method described in Liu and Seipel (2014) to compute two major vanishing points (representing horizontal and vertical directions in the real world) and categorize edge line segments according to these two orientations. Figure 4 shows these two groups on both TIR and visible images respectively.

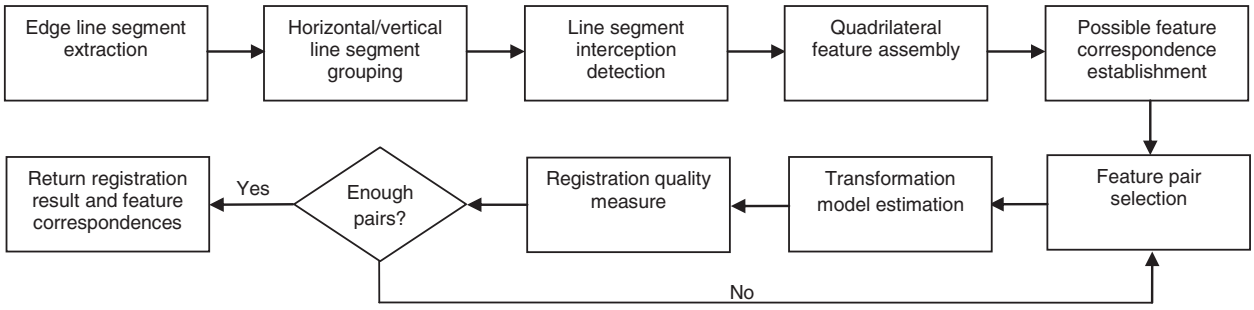


Fig. 3 Overview of the proposed registration method

Line segment interceptions

A quadrilateral can be considered as a composition of four interceptions, each of which is formed by a pair of horizontal and vertical line segments. Therefore, with line segments of both orientations detected, the next step is to find such intercepting pairs among all previously identified line segments. To eliminate impossible combinations while also considering the influence of noise in the line segment detection process, we define that a horizontal line segment l_h intercepts with a vertical line segment l_v if one of l_h's end points is within p pixels from one of the end points of l_v, where p is set to 5 for this work. We then enumerate all the horizontal line segments and searched for vertical line segments whose end points satisfy the distance criterion. The result is a group of preliminary interceptions each of which is represented by two line segments of different orientations and their involving end points. Figure 5a illustrates a possible configuration of the preliminary interceptions. In the figure, we have interceptions formed by line 2 and line 4, line 2 and line 5, line 2 and line 1 as well as line 3 and line 4. For the convenience of discussion, these interceptions are denoted sequentially I_{24}, I_{25}, I_{21} and I_{34}.

Before we can generate quadrilaterals from these interceptions, we need to address two problems existing in the preliminary set. First, since we use a distance threshold

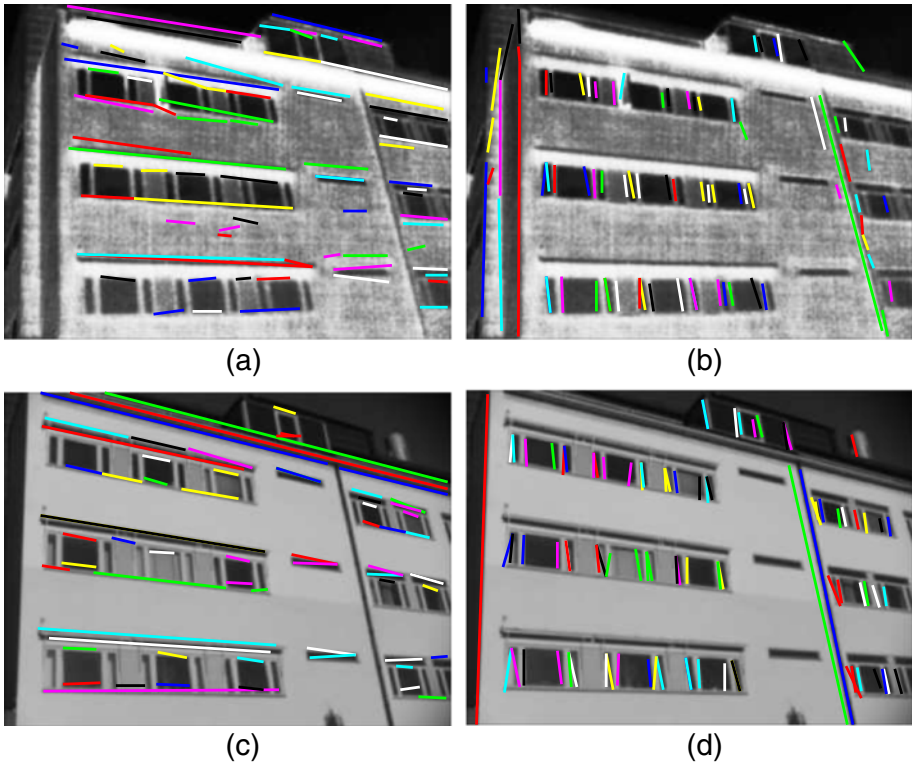


Fig. 4 Grouped edge line segments (colors are used here to distinguish individual line segments): **a** and **b** TIR horizontal and vertical, **c** and **d** Visible horizontal and vertical

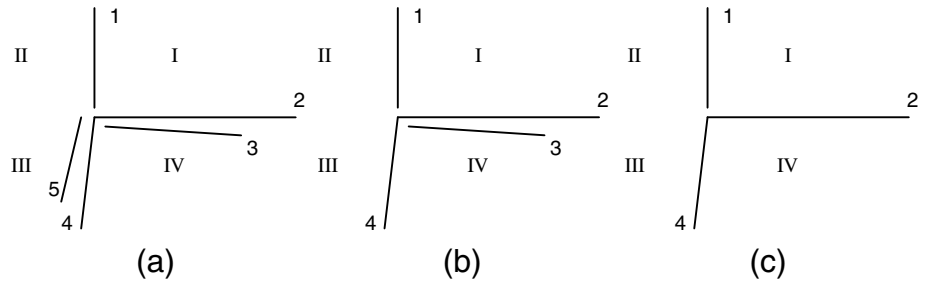


Fig. 5 Interception refinement process and the numbering of quadrants (in Roman numerals): **a** Preliminary, **b** and **c** Refinement based on horizontal and vertical line segments respectively

between two end points, a pair of horizontal and vertical lines might not intercept at an exact point, e.g., I_{21} in Fig. 5a. We call this kind of interceptions two-point interceptions. The solution to this problem is using the centroid of the two involving end points to represent the interception point. Second, a horizontal line can be paired with multiple vertical lines and vice versa. In this case, however, not all of these one-to-many pairs should be used further because, e.g., some of the lines can result from noise. Therefore, when this situation arises, we need a strategy to retain the most reliable pairs and discard the rest. In general, we incline to the preliminary interceptions whose horizontal and vertical line already intercept at an exact point, namely, one-point interceptions (e.g., I_{24} in Fig. 5a). More specifically, we first assign a quadrant number to each interception. For instance, I_{21} in Fig. 5a is an interception of quadrant I while I_{24} is an interception of quadrant IV and so on. Then we examine each horizontal line segment shared by more than one interception. For each quadrant, we keep a one-point interception if there is any and delete the rest. If a quadrant in question only has two-point interceptions, we will keep all of them and replace the two points using their centroid as discussed before. Figure 5b shows the result after the aforementioned refinement. Note that I_{25} is eliminated because it has the same quadrant as I_{24}, which is a one-point interception. Since such one-to-many interception case can happen to vertical line segments as well, we repeat the refinement process once from the perspective of vertical line segments, too. The refined result of the exemplary configuration is depicted in Fig. 5c. This time I_{34} is left out for the similar reason. The detected interceptions are displayed in Fig. 6.

Deriving quadrilaterals

With a list of interceptions formed by horizontal and vertical line segments, an intuition is that a quadrilateral is formed by a group of four connected interceptions which represent the four vertices of the quadrilateral (the first pattern in Fig. 7). However, a close look at Fig. 6 reveals that many potential quadrilaterals may be derived from groups of connected interceptions that resemble the remaining patterns or their variants of different orientations in Fig. 7. In order to recover as many quadrilaterals as possible, we propose a method which walks along a group of connected interceptions and derives a quadrilateral that can best characterize the underlying rectangular façade element this group originates from. Given an interception, there are already three known points (the interception point and two end points opposite to

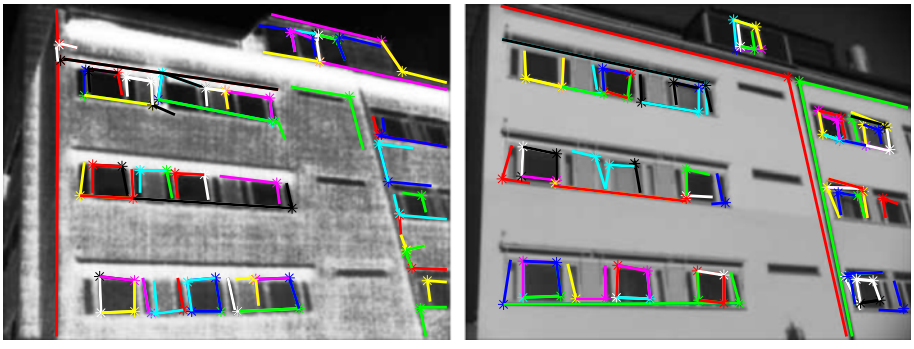


Fig. 6 Interceptions formed by horizontal and vertical line segments: **a** TIR and **b** Visible. (Colors are used here to distinguish individual interceptions)

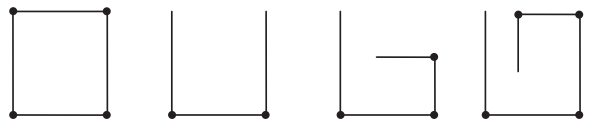


Fig. 7 Example cases of connected interceptions

it). Since we assume these interceptions are from rectangular façade elements, the fourth point should be easily computed leveraging properties of a parallelogram. However, we cannot simply treat the two line segments of an interception as adjacent edges of a parallelogram due to the presence of perspective distortion. Consequently, a distortion removal process is needed prior to deriving quadrilaterals. Since we have already computed two major vanishing points for the detection of horizontal and vertical line segment at the beginning of this section, we can derive the line at infinity $l_\infty = (l_1, l_2, l_3)^T$ from the two vanishing points using

$$l_\infty = vp_h \times vp_v \tag{1}$$

where vp_h and vp_v are the vanishing points of horizontal and vertical orientations respectively and the symbol "$\times$" represents vector cross product. Subsequently, an affine rectification procedure (described in Hartley and Zisserman (2003)) can be carried out by transforming the original image with the following homography

$$H = H_A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \tag{2}$$

where H_A is any affine transformation and in this case it usually contains scaling information. After removing the perspective distortion, we can start with an interception in the group and use the three known points it provides to calculate the fourth point of the quadrilateral as mentioned above. The process continues with the neighboring interception until all the interceptions within this connected group are visited. The derived quadrilateral with the largest area (in number of pixels) is the final result we are looking for and interceptions of this group are then removed from the list before we proceed to the next group. This idea is demonstrated in Fig. 8 using the fourth pattern in Fig. 7 as example. The first row shows which intersection and the three points are selected at each step (in red) while the second row shows the derived quadrilaterals. In this example, the process begins with the bottom right interception and chooses to walk along its vertical line segment first. Once all the interceptions in that direction have been visited, the process goes back to the starting interception and tries to walk in the direction of its horizontal line segment. In this manner, our method ensures that all interceptions in the group are visited once no matter where it starts. Eventually, the quadrilateral

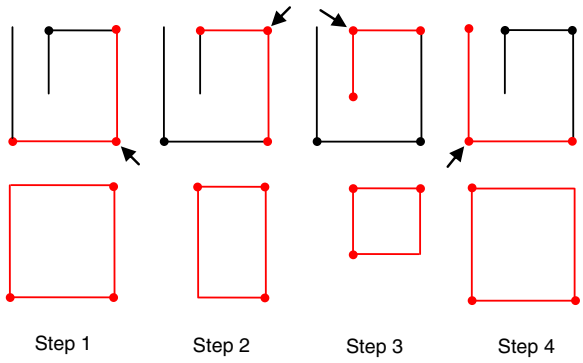


Fig. 8 Illustration of quadrilateral construction from a connected interception group. In each step the arrow points at the interception in question

from either Step 1 or Step 4 is chosen to represent the group.

After all interceptions have been processed, we have a list of quadrilaterals represented by their respective four vertices. Quadrilaterals whose areas (measured in number of pixels) are smaller than a fraction of the median area of all the quadrilaterals are discarded because these small quadrilaterals likely result from noise rather than actual façade elements. Through experiments, we found that setting the fraction factor to 0.4 gave the best results. In addition, it can also occur that more than one quadrilaterals are derived from a single rectangular façade element. These quadrilaterals are similar in shape and close to each other spatially. To reduce redundancy, we merge such quadrilaterals. Finally, we transform this list of quadrilaterals back to their positions in the original image space (an inverse process of the aforementioned affine rectification) for the purpose of identifying possible registration control point (CP) correspondences, which will be described in the next section. Figure 9 illustrates the final quadrilateral features.

Identification of possible feature correspondences

In the previous section, we detected quadrilateral features based on the interceptions of horizontal and vertical edge line segments. Although the quadrilaterals are represented by their respective vertices, we instead chose edge centers as CPs for further registration processes. The motivation for such a choice is that positions of edge centers are less influenced in situations where a few vertices deviate from the true corresponding corners of underlying façade elements (refer to Fig. 10). Therefore, this choice can improve the accuracy of CP positions in those situations. With CP defined, the next question is how to match them in order to estimate the transformation model. Determining corresponding CPs for multimodality

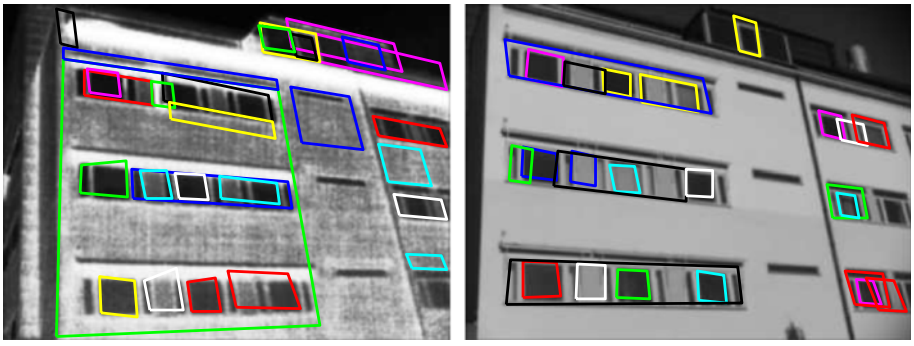


Fig. 9 Quadrilateral features reconstructed from interceptions of horizontal and vertical line segments: **a** TIR and **b** Visible. (Colors are used here to distinguish individual features)

registration is a challenge. Due to the different imaging mechanisms, image content-based descriptors are not viable options. However, the assumption we discussed in the introduction, which is that the façade should be the major content in the view and images of both modalities are taken from similar viewing positions, indicates if a CP from the TIR image has a corresponding CP in the visible image, the image coordinates of both CPs should be similar (assuming both images have the same size). In other words, given a TIR CP, we can establish a set of visible CPs as matching candidates by searching within a circular region in the visible image. The circular region is defined by the coordinate of the TIR CP as its center and with a certain radius r. As shown in Fig. 11, the coordinate of a TIR CP p is used to define a circular search region with radius r. So far, our searching strategy of corresponding features has resembled the one presented in Hrkać et al. (2007). However, since we are using higher-level quadrilateral features, their geometric properties can be exploited to further reduce the number of candidates:

1. The edge which a candidate CP represents and the edge which a TIR CP represents should have the same edge number (please refer to Fig. 11 for the numbering of quadrilateral vertices and edges).
2. The aspect ratios of involved quadrilaterals need to be similar.

Like the derivation of quadrilateral features, their aspect ratios were computed in affine space as well (after the affine rectification procedure introduced in the previous section). Doing so improves the reliability of aspect ratio comparison because without the influence of perspective distortion, rectangular façade elements of the same size (e.g., windows of the same design) will yield parallelograms with similar aspect ratios.

According to the two searching criteria, the aforementioned process gives each TIR CP a set of potential visible CP correspondences. In the example of Fig. 11, the possible candidate set for TIR CP p contains p_1 and p_2 from the visible image. The cardinality of the set can be 0 or any positive integer. The reason for one-to-many matches can be ascribed to the repetitive nature of façade layout as well as feature detection errors. For instance, a quadrilateral representing a window can be matched to quadrilaterals from its neighboring windows, which usually share

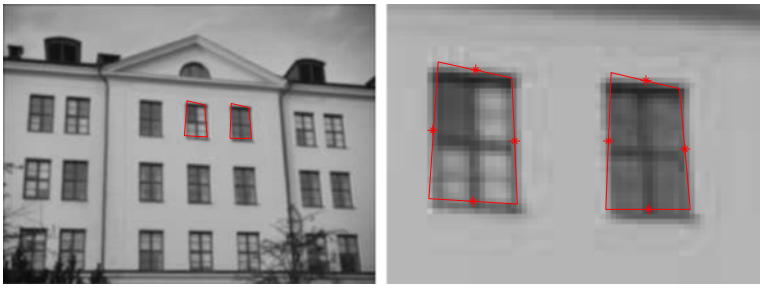


Fig. 10 a Two quadrilaterals which can be benefited from edge centers as CPs and **b** Close-up look with edge centers plotted

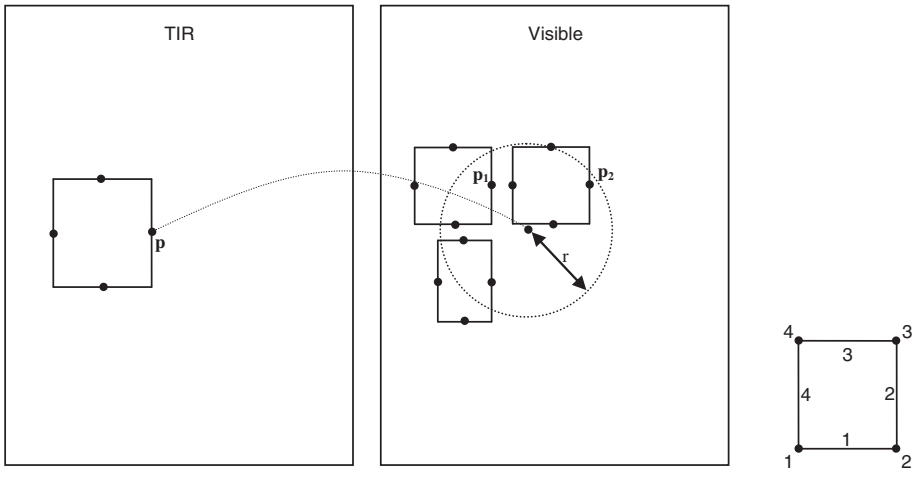


Fig. 11 The searching process for possible corresponding CPs

the same design or it can be matched to a quadrilateral derived from a pipe-ledge interception, which is accidentally within the vicinity and has a similar aspect ratio. Once the potential candidate set has been established for each TIR CP, we convert these CP-level matching results back to quadrilateral-level ones via the following voting scheme: given 4 CPs of a TIR quadrilateral, all their candidates are merged into one set, denoted C. For each visible CP in C, its quadrilateral receives one vote. In the end, all visible quadrilaterals with at least 3 votes are designated as matching candidates of the TIR quadrilateral in question. Similar to the CP-level matching results, some TIR quadrilaterals can also be matched to more than one visible quadrilateral. In the next section, we will discuss a strategy for resolving false correspondences and arriving at a correct transformation model. After the conversion process, all TIR quadrilaterals without any visible counterparts matched are removed, excluding them from subsequent registration processes.

Transformation model estimation

Images of the same façade photographed with cameras modeling central projection are related by a projective transformation, which is represented by a homogeneous non-singular 3×3 matrix H. In order to estimate the elements of H, at least 4 pairs of corresponding image points are required with no 3 points collinear. However, in practice, positions of corresponding points in their respective images do not accord exactly. Estimations from the minimal amount of point pairs are often unreliable. The common solution is to include far more point correspondences than the minimum requirement to build an over-determined equation system and solve it as an optimization problem (Hartley and Zisserman 2003). In the last section, a list of possible feature correspondences were established and in this section we will present how we select the set of correct feature pairs in order to find the best estimation of H.

Suppose we will need m ($m > 1$) quadrilateral feature pairs for the estimation. A straightforward approach towards feature correspondence selection is to enumerate all the possible combinations of m pairs from the total n pairs of correspondences, namely, $\binom{n}{m}$ iterations. The number n is typically around 30 to 40 in our test image pairs, which means as m increases, it quickly becomes impractical to evaluate all the combinations. In view of this, we chose to adopt the forward selection algorithm for this task. The algorithm starts with an empty result set and repeats m times. At each iteration, it adds only one feature pair to the result set such that the current selection produces the best registration result. Since true correspondences result in good registration and vice versa, the verification of feature correspondences can be carried out through evaluating their registration results. Hence, each iteration blends both selection and verification. Within an iteration, we first obtain a projective transformation from the current set of feature pairs. Afterwards, the remaining TIR quadrilaterals (i.e., excluding the m ones used for estimating the model) are transformed into the visible image coordinate system. If the selected corresponding pairs were in fact true, hence the genuine estimated transformation model, a transformed TIR quadrilateral $\hat{Q}$ would overlap with a quadrilateral Q in the visible image to a large extent whereas a selection of erroneous feature correspondences would lead to less or no overlapping. Ideally, a perfect alignment of a pair of quadrilaterals means a unity ratio between $A_{\hat{Q}Q}$ and A_Q, where $A_{\hat{Q}Q}$ is the area of the overlapping portion of $\hat{Q}$ and Q and A_Q is the area of Q. The ratio decreases as the overlapping diminishes. This insight leads us to a quality measure based on the

aforementioned ratio. More specifically, for each visible quadrilateral j which a transformed TIR quadrilateral i overlaps with (blank boxes 2, 3, 4 in Fig. 12), the quality score S_{ij} of the transformed TIR quadrilateral i with regard to the visible quadrilateral j is computed according to

$$S_{ij} = w_{ij}(A_{ij}/A_j) \tag{3}$$

where w_{ij} is a weighing term to penalize the score and A_{ij} is the area of the overlapping region between quadrilateral i and j while A_j is the area of quadrilateral j. The weighing factor is defined as

$$w_{ij} = min(A_i, A_j)/max(A_i, A_j) \tag{4}$$

where A_i is the area of quadrilateral i. All areas are measured in pixels. We designate the maximum of S_{ij} as the quality score for the transformed TIR quadrilateral i. The final score for the current estimated transformation model can then be expressed as

$$S = \sum_i max(S_{ij}) \tag{5}$$

Method evaluation and results

To evaluate our method, we acquired images of façades from buildings on the university campus. This test location comprises mixed architectures. Therefore, both old and modern façade elements and structures are represented in the test image database. All TIR images were taken using an AGEMA Thermovision 570 infrared condition monitoring system. We shot them during nights to reduce the influence of heat from the sun. For several façades, images from different views were taken where this was possible (viewing positions can be constrained by tree occlusion or ground obstructions). For each acquired TIR image, its visible counterpart was taken from the approximately same viewing position during the day using a FUJIFILM FINEPIX REAL 3D W3 digital camera. In total, we gathered 41 TIR/visible pairs for the experiment. The size of TIR and visible images are 320 × 240 and 2592 × 1944 respectively.

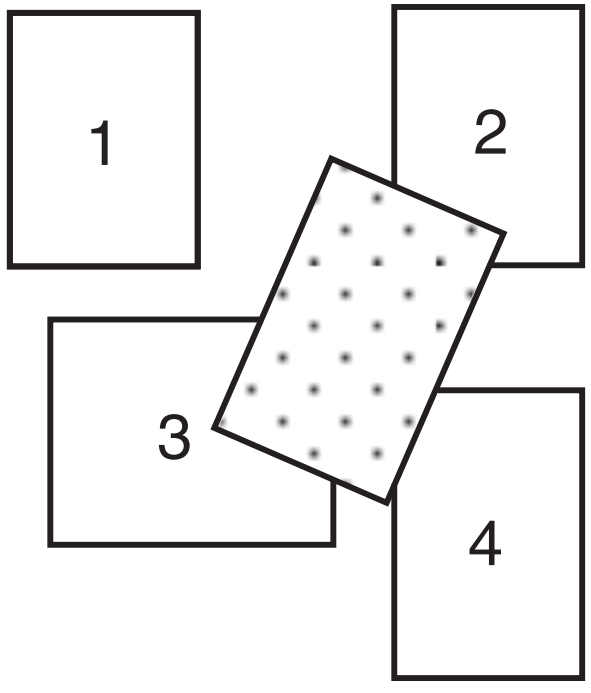


Fig. 12 A transformed TIR quadrilateral (dot-patterned) overlaps with quadrilaterals in the visible image (numbered)

During the test, we treated visible images as references and attempted to register TIR images with them. Prior to feature detection, histogram equalization was applied to TIR images to enhance the contrast while visible images, which have much larger size, were scaled down to match the size of their TIR counterparts. Resizing the visible images ensures the subsequent feature matching based on image space distance is meaningful. We set the feature searching radius r to 50 pixels, the similarity threshold for the aspect ratio to 0.5 and the number of feature correspondences m for estimating a transformation model to four. Functions for model estimation and image transformation are provided by the image processing toolbox in MATLAB. Among the 41 pairs of testing images, 33 pairs succeeded in finishing the registration process and Fig. 13 displays some examples of the results visually. The remaining eight image pairs did not exhibit a sufficient number of corresponding features so the proposed method could not register them.

To quantify the registration error, we manually selected ten pairs of corresponding salient points in each of the 33 successfully registered image pairs to establish a ground-truth for computation of the point-wise registration errors in terms of the L^2-norm. These points are typically corners of windows and we tried to have them spread out as much as possible across the image. Figure 14a reports the mean registration errors for each of the 33 image pairs in ascending order of errors. The overall registration error for all images is on average 3.23 pixels and the standard deviation of the error is 1.89 pixels. We also did a registration of the ground-truth points based on themselves as corresponding CPs and through that we determined registration errors inherent to the manual selection procedure. These results are shown in Fig. 14b in ascending order of mean error. The average registration error with manual ground-truth CPs was 1.08 pixels with a standard deviation of 0.65 pixels.

While Fig. 14 illustrates levels and variation of registration errors between images in the entire successful set, it does not reveal typical variations of errors within images. Visual inspection of the test results suggests that registration errors seem to be larger in the periphery as compared

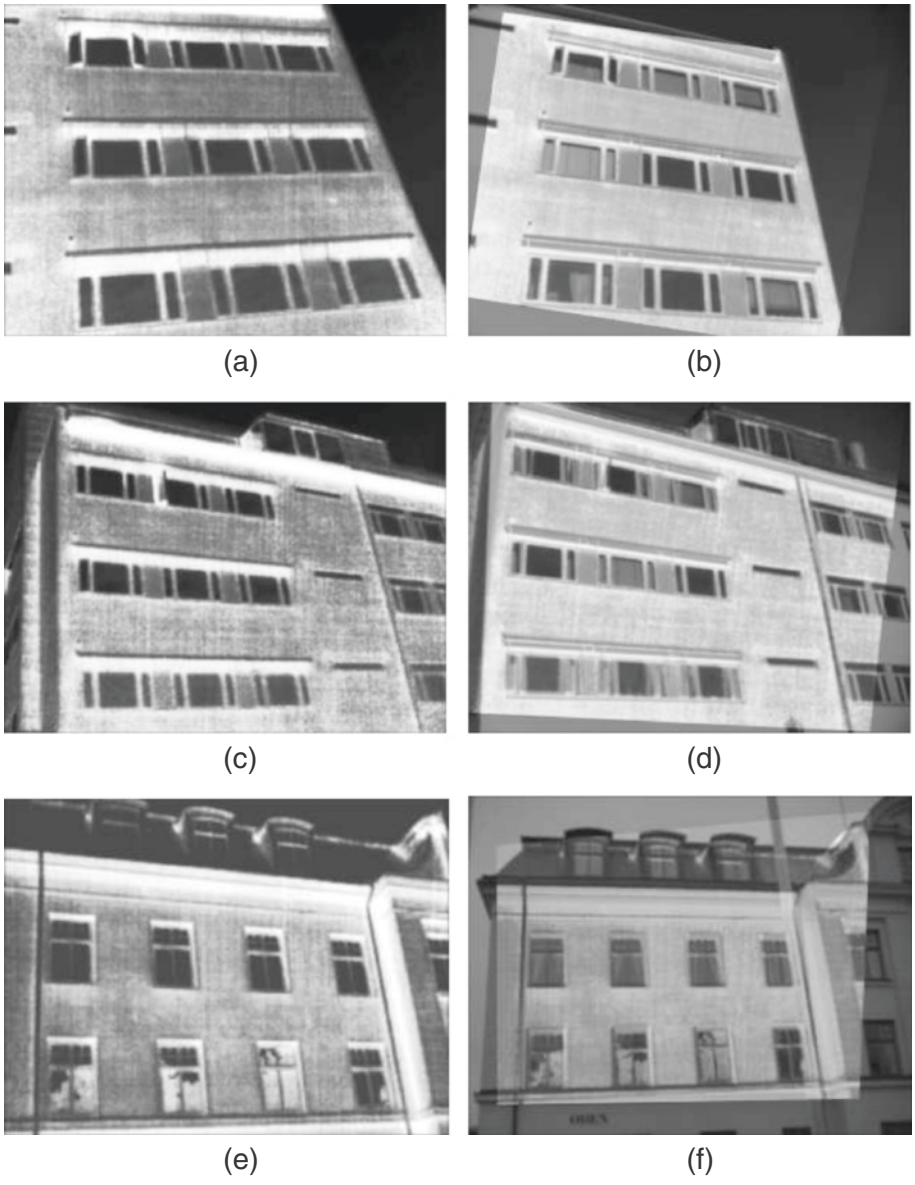
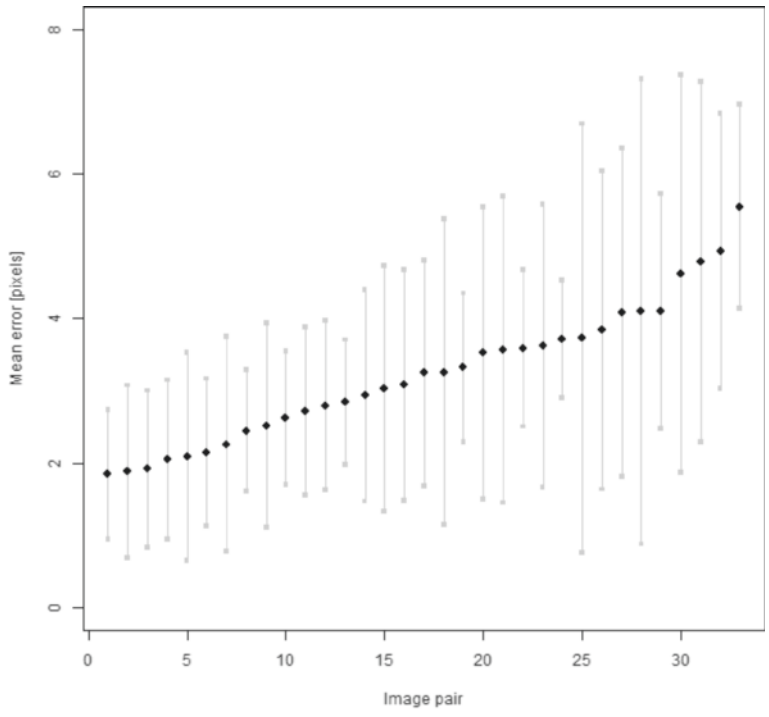


Fig. 13 Registration results. *Left*: original TIR images. *Right*: fused images through alpha blending

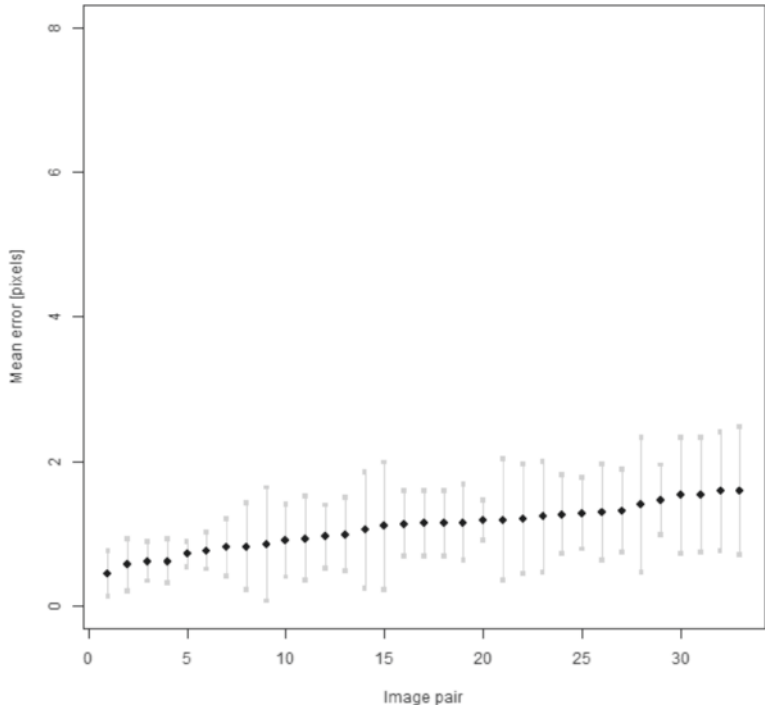
to central parts of the image. We therefore grouped registration errors obtained from all 330 ground-truth points by their distance to the image center using three equally spaced distance intervals where the maximum distance is

$$R_{max} = \frac{\sqrt{w^2 + h^2}}{2} \tag{6}$$

where w is the width and h is the height of the image. Table 1 gives an account on the areas in the image covered by these distance intervals, the number of ground-truth points contained therein and the median registration errors for those points obtained with the proposed registration method. It shows that the registration error is less than three pixels within a radial region from the center ($I_1 + I_2$) which covers 73% of the image area. The boxplot in Fig. 15 summarizes graphically what was assumed from visual inspection of registration results, namely, the increase of registration errors for regions towards the boundaries of the images. Since errors in the three groups of ground-truth points showed are not normally distributed, *median* rather than *mean* values are listed in Table 1. Also, the increase of errors seen in Fig. 15 is a significant effect, as revealed by a Wilcoxon rank-sum test



(a)



(b)

Fig. 14 Average registration errors for ground-truth points of all successfully registered image pairs (ascending order of error magnitude). Gray bars represent twice of standard deviations: **a** Registration using edge centers of auto-selected quadrilaterals as CPs and **b** Registration using ground-truth points as CPs

Table 1 Distance intervals and image coverage

Interval	Range	Coverage	GT Points	Median
I_1	$d \leq \frac{R_{max}}{3}$	18.2 %	104	2.57
I_2	$\frac{R_{max}}{3} < d \leq \frac{2R_{max}}{3}$	54.5 %	197	2.99
I_3	$d > \frac{2R_{max}}{3}$	27.3 %	29	3.76
$I_1 + I_2 + I_3$	$d < R_{max}$	100 %	330	2.94

between I_1 and I_2 ($W = 8551.5, p = 0.01845$), as well as between I_2 and I_3 ($W = 2191.5, p = 0.04324$).

Discussion and conclusion

Robustness and limitations of our method

The challenge of registering TIR and visible images lies in significant content difference and the generally poor quality of TIR images, which can impede our algorithm in deriving identical quadrilateral features from the same rectangular façade element imaged in the two modalities. For example, the four selected features for registration as well as their respective CPs of an image pair with relatively high mean registration error are depicted in Fig. 16. Here, CPs with the same color in different images are matched to estimate the transformation model. Quite obviously, corresponding CPs are representing different façade positions, especially those in red circles. In spite of such difficulty, visual inspection of the registration results in our test (a few examples are shown in Fig. 13) reveals generally very good registration quality characterized by little, if any, notable ghosting effect. This is also expressed by an overall mean registration error of 3.23 pixels from the ground truth points, which is less than the error of 4.28 pixels reported in the other related work (Han et al. 2013) for a comparable yet rather limited set of images of an urban scene. Concerning the visible modality, external factors such as different weather conditions and times of a day when images are taken should not confine our method. This is because the proposed quadrilateral features originate from edge line segments, which are computed from local contrast and therefore are not so sensitive to illumination levels. On the other hand, the application of image perspective information makes the feature detection robust against changes of image capturing positions, especially different viewing angles. Nevertheless, as we mentioned earlier in this paper, when standing in front of a façade, we often do not have many very different viewing positions to choose from due to road obstructions and occlusions from trees.

Interesting insight is gained from the error analysis of the manual ground-truth points. As described in the *Method evaluation and results* section, instead of using CPs from quadrilateral features as identified by our method, transformation models were also estimated from the 10 pairs of ground-truth points for each image pair and then these point pairs were used again to determine registration errors. The average registration errors

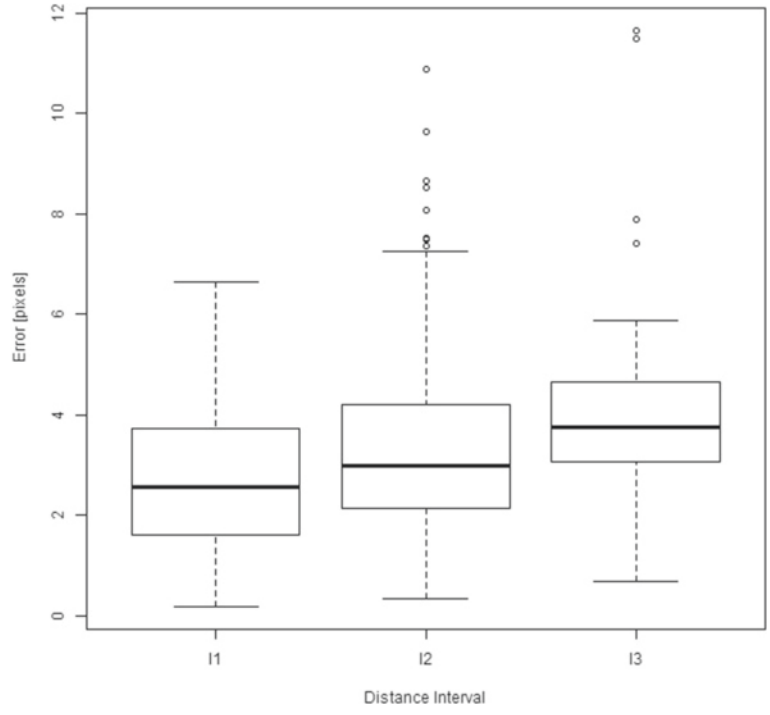


Fig. 15 Boxplots of registrations errors with respect to distance intervals from image center

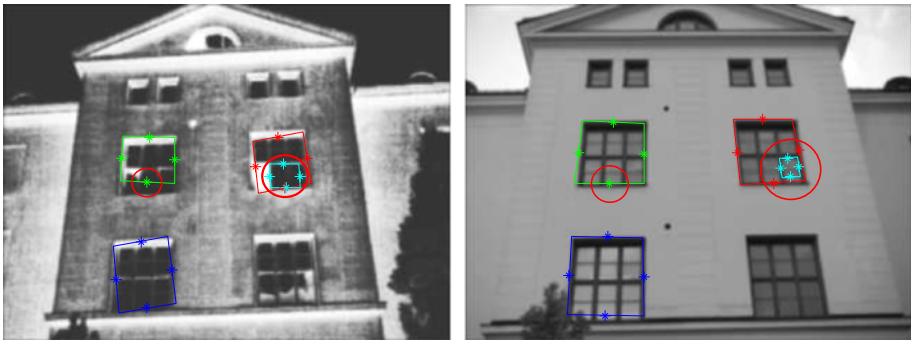


Fig. 16 Incongruence of corresponding CPs (in red circles) from features selected by our method leads to a relatively high mean registration error: **a** TIR and **b** Visible

are plotted in Fig. 14b with the overall mean error 1.08 pixels. This amount of error in the manually designated ground-truth points should be considered when interpreting the registration errors of our algorithm presented in Fig. 14a because they are related. On the other hand, it also illustrates that a human observer performing manual registration is not perfect and the errors from both approaches only differ by as little as two pixels on average in our test.

Since our method relies on abundant rectangular elements on a façade (typically windows) to derive features for registration, its application can be limited by the intrinsic differences between TIR and visible images as well as the visual design of a façade. To demonstrate the limitation, we take two examples from the eight image pairs that failed to be registered. In Fig. 17a, b quadrilateral features are derived from individual windows but most windows generate two smaller features in the TIR image

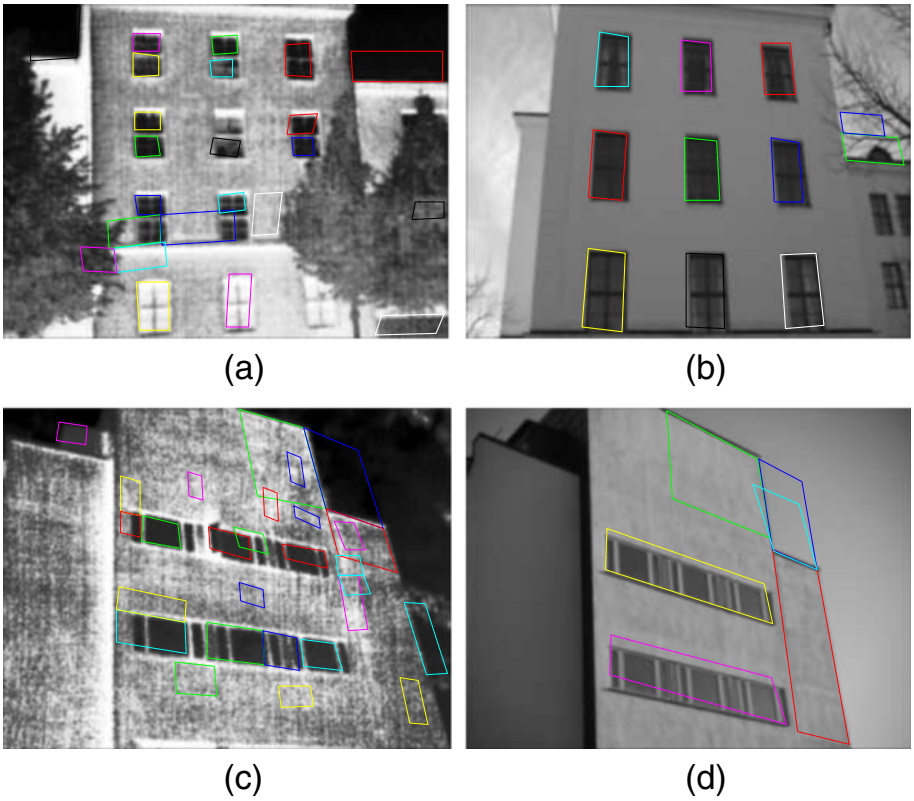


Fig. 17 Two examples where there are not enough corresponding features between TIR and visible images: **a** and **b** Case 1, **c** and **d** Case 2. (Colors are used here to distinguish individual features))

due to the highly visible horizontal bars in the center of windows. In Fig. 17c, d the design of the façade causes windows of the same floor to exhibit similar gray level values in the visible image so they are detected as a single large quadrilateral feature whereas the TIR modality represents windows as dark rectangular regions and the high contrast between windows and their peripheral structures results in detection of features characterizing those windows. Apparently, neither of the above cases can lead to enough correct feature correspondences much less a successful registration. In general, if a façade features many rectangular windows and/or other rectangular elements which can be delineated from the background wall, we expect our method to perform well.

Given that the objective of this work is to develop and evaluate a method that enables successful registration of TIR images with visible images for an AR-based system for building diagnostics, developing the final system is not on the current research agenda and therefore an algorithm performance analysis in terms of run-times is not part of the studies presented here. On a short note, however, we can state that the entire image processing and registration pipeline according to Fig. 3 requires a few seconds to 20 seconds for image pairs from our dataset. This amount of time spent registering images is insubstantial in comparison to the whole process of diagnosing a building considering that, for example, it might take an inspector several minutes to walk from one façade of the building to another.

AR system and future works

Considering the proposed method will be applied to build a final AR system, it is also relevant to interpret registration performance with respect to errors for real use cases. As our distance-based error statistics show, average error is about three pixels or less in the central parts ($I_1 + I_2$) of an image which cover almost three quarters of the image. This is also the area of interest where users will naturally focus on when pointing the portable device (camera) towards the buildings under investigation. TIR imaging devices of the kind we used in the study typically have a fairly limited field of view. For the Amega Thermovision system that has a horizontal field of view of 24 degrees and an image resolution of 320 pixels horizontally, a registration error of three pixels corresponds to 0.23 degrees visual angle. Depending on viewing distance, pixel misalignment on the real façade corresponds to 3.93 cm at near range inspection (10 m) up to 11.78 cm at long range inspection (30 m). Meanwhile, users of the system will inevitably bring about performance and perception errors. Hence, future experiments should also be conducted to find out what these errors are as well as characterizing them.

Additionally, given that the current version of the method has been implemented in MATLAB without any attempt so far to optimize run-time performance, there is obviously good potential to achieve interactive processing in the final system. Possible measures include, e.g., utilizing GPU or multi-kernel processing and adopting other state-of-the-art computer vision libraries such as OpenCV to realize certain computation-intensive steps in the pipeline, especially vanishing point detection for horizontal and vertical line segments.

Conclusion

In this paper, we have presented quadrilateral features derived from visible and TIR façade images and an image registration pipeline revolves around such features. These features are abundant on façdes thanks to the existence of plentiful rectangular objects, e.g., windows. Because the quadrilateral features stem from horizontal and vertical edge line segments, their strength lies in the robustness against changes of illumination and image capturing position. The higher-level nature of them comparing to discrete edge line segments also makes possible the proposed feature matching process and transformation quality measure. Both qualitative and quantitative results obtained from the evaluations demonstrated our method succeeded in a majority of image pairs in the test dataset with a satisfactory mean registration error. For identification and localization of structures hidden inside façades, which are parts of the intended application of thermographic building analysis, this amount of error seems tolerable considering the size of those structures, in particular for larger artifacts of interest such as insulation defects, failure of heating pipes or ventilation system components. Therefore, we believe obtained registration results are certainly sufficient and the presentation of them through an AR interface on-site is more insightful than traditional non-colocated presentation of mere thermographic images.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Both authors worked closely to design and evolve the registration method. FL implemented the method, gathered image data and carried out tests on the data while SS supervised the whole process and provided advice for improvement. FL drafted the paper and then SS joined for the revision, in particular the refinement of "Method evaluation and results" and "Discussion and conclusion" sections. Both authors read and approved the final manuscript.

Acknowledgements

This research was supported by funding from Faculty of Engineering and Sustainable Development at University of Gävle.

Received: 30 June 2015 Accepted: 22 October 2015
Published online: 09 November 2015

References

Balaras, C, & Argiriou, A (2002). Infrared thermography for building diagnostics. *Energy and buildings*, *34*(2), 171–183.

Behzadan, AH, Dong, S, Kamat, VR (2015). Augmented reality visualization: A review of civil infrastructure system applications. *Advanced Engineering Informatics*, *29*(2), 252–267.

Chi, H-L, Kang, S-C, Wang, X (2013). Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in Construction*, *33*, 116–122.

Choi, K, Kim, C, Kang, M-H, Ra, JB (2011). Resolution improvement of infrared images using visible image information. *Signal Processing Letters, IEEE*, *18*(10), 611–614.

Coiras, E, Santamarı, J, Miravet, C (2000). Segment-based registration technique for visual-infrared images. *Optical Engineering*, *39*(1), 282–289.

Dana, KJ, & Anandan, P (1993). Registration of visible and infrared images, In *Optical Engineering and Photonics in Aerospace Sensing* (pp. 2–13): International Society for Optics and Photonics.

Dawn, S, Saxena, V, Sharma, B (2010). Remote sensing image registration techniques: A survey. In A Elmoataz, O Lezoray, F Nouboud, D Mammass, J Meunier (Eds.), *Image and Signal Processing. Lecture Notes in Computer Science,* vol. 6134 (pp. 103–112): Springer.

Friedman, S, & Stamos, I (2013). Online detection of repeated structures in point clouds of urban scenes for compression and registration. *International journal of computer vision*, *102*(1–3), 112–128.

Gade, R, & Moeslund, TB (2014). Thermal cameras and applications: a survey. *Machine vision and applications*, *25*(1), 245–262.

Han, J, Pauwels, EJ, De Zeeuw, P (2013). Visible and infrared image registration in man-made environments employing hybrid visual features. *Pattern Recognition Letters*, *34*(1), 42–51.

Harris, C, & Stephens, M (1988). A combined corner and edge detector, In *Alvey Vision Conference*, 15 (p. 50). Manchester, UK.

Hartley, R, & Zisserman, A (2003). *Multiple View Geometry in Computer Vision,* 2nd edn: Cambridge University Press.

Hrkać, T, Kalafatić, Z, Krapac, J (2007). Infrared-visual image registration based on corners and hausdorff distance, In *Image Analysis* (pp. 383–392): Springer.

Huttenlocher, DP, Klanderman, GA, Rucklidge, WJ (1993). Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *15*(9), 850–863.

Kong, SG, Heo, J, Boughorbel, F, Zheng, Y, Abidi, BR, Koschan, A, Yi, M, Abidi, MA (2007). Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition. *International Journal of Computer Vision*, *71*(2), 215–233.

Kumar, S, Marks, TK, Jones, M (2014). Improving person tracking using an inexpensive thermal infrared sensor, In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference On* (pp. 217–224): IEEE.

Keller, Y, & Averbuch, A (2006). Multisensor image registration via implicit similarity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(5), 794–801.

Kupfer, B, Netanyahu, NS, Shimshoni, I (2013). A sift-based mode-seeking procedure for efficient, accurate registration of remotely sensed images, In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International* (pp. 4142–4145).

Kylili, A, Fokaides, PA, Christou, P, Kalogirou, SA (2014). Infrared thermography (irt) applications for building diagnostics: A review. *Applied Energy*, *134*(0), 531–549.

Le Moigne, J, Netanyahu, NS, Eastman, RD (2011). *Image Registration for Remote Sensing*: Cambridge University Press.

Li, Q, Wang, G, Liu, J, Chen, S (2009). Robust scale-invariant feature matching for remote sensing image registration. *Geoscience and Remote Sensing Letters, IEEE*, *6*(2), 287–291.

Liu, F, & Seipel, S (2014). Detection of façade regions in street view images from split-and-merge of perspective patches. *Journal of Image and Graphics*, *2*(1), 8–14.

Lowe, DG (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91–110.

Mesolongitis, A, & Stamos, I (2012). Detection of windows in point clouds of urban scenes, In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference On* (pp. 17–24): IEEE.

Morris, NJ, Avidan, S, Matusik, W, Pfister, H (2007). Statistics of infrared images, In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference On* (pp. 1–7): IEEE.

Oliveira, FP, & Tavares, JMR (2014). Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, *17*(2), 73–93.

Pluim, JP, Maintz, JA, Viergever, MA (2003). Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, *22*(8), 986–1004.

Prakash, A (2000). Thermal remote sensing: concepts, issues and applications. *International Archives of Photogrammetry and Remote Sensing*, *33*(B1; PART 1), 239–243.

Wang, X, Kim, MJ, Love, PE, Kang, S-C (2013). Augmented reality in built environment: Classification and implications for future research. *Automation in Construction*, *32*, 1–13.

Yi, Z, Zhiguo, C, Yang, X (2008). Multi-spectral remote image registration based on sift. *Electronics Letters*, *44*(2), 107–108.

Zitova, B, & Flusser, J (2003). Image registration methods: a survey. *Image and vision computing*, *21*(11), 977–1000.