Advanced Engineering Informatics 27 (2013) 120–130

Contents lists available at SciVerse ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

A Self-adaptive ASIFT-SH method

Peter Podbreznik [a,*], Božidar Potočnik [b]

[a] University of Maribor, Faculty of Civil Engineering, Smetanova 17, 2000 Maribor, Slovenia
[b] University of Maribor, Faculty of Electrical Engineering and Computer Science, Smetanova 17, 2000 Maribor, Slovenia

ARTICLE INFO

Article history:
Received 8 March 2012
Received in revised form 27 September 2012
Accepted 5 November 2012
Available online 11 December 2012

Keywords:
Corresponding point
Arbitrary-selected point
Widely-separated views
3D reconstruction
Segmentation

ABSTRACT

When monitoring events on a building site using a system of multiple cameras, it is necessary to establish correspondence between the acquired imaging material. The basic problem when attempting this task is the establishment of any correspondence between points located on uniform areas of the images (e.g. regions with uniform colour or texture). The basic version of our ASIFT-SH method can mainly solve such a problem. This method consists of four steps: (i) determining the initial corresponding points within the images of both views by using the ASIFT method, (ii) grouping of initial corresponding points from the first step into subsets, based on segmented regions, (iii) calculation of local homographies for a particular subset of corresponding points, and (iv) determining any correspondence between arbitrary points from a particular camera's viewpoint, by using a suitable local homography. The critical step of this method concerns segmentation. Therefore, we have introduced into our algorithm a step for adaptive adjustment, the segmented regions being remodelled so that they better meet the required coplanarity criterion. This introduced step is based on a 3D reconstruction of the initial corresponding points and a search for the minimal number of planes within the 3D space to which these points belong. Those points that belong to a particular plane, represent a newly-created subset of the initial corresponding points. The results point out that the introduction of this adaptive step into ASIFT-SH significantly improves the accuracy of corresponding points' calculation. The mean error is 1.63 times lower and the standard deviation is 2.56 times lower than by the basic version of the ASIFT-SH method.

1. Introduction

Many IT-supported methods have been introduced for monitoring construction projects [1–4], but so far none has been able to tackle this problem satisfactorily and reliably. An automated vision-based tracking framework, which can provide 3D spatial coordinates of entities across time has been introduced by Brilakis et al. [5], where images are acquired from two or more static cameras placed on building sites. A recent systematic review of 2D vision tracking methods was presented for finding the most appropriate for 3D vision tracking construction resources [6]. A more detailed review of building site monitoring systems can be found in [7]. The concept of an automated construction activity monitoring system was introduced by Podbreznik and Rebolj [8] and fully formulated in [9,10]. A subsystem, called 4D-ACT [10], is aimed at discovering any differences between the planned and actually-built elements from building site images. Major problems, such as the overlapping of building elements and limited camera fields-of-view have been solved by using widely-separated multiple cameras when acquiring imaging material. A system with widely-separated multiple cameras and a 4D model [11,7] as an learning set, requires established correspondence between any widely-separated views of the same observed building's object, for an arbitrary selected point.

In literature [12–18], several methods exist for establishing initial corresponding points between images from widely-separated views. The turning point in this researched field regarding the establishment of correspondence between camera views was the scale-invariant feature transform (SIFT) proposed by [14]. This method is invariant to image scaling and rotation; and is partially invariant to illumination and viewpoint changes. The affine-SIFT (ASIFT) method [19,20] is a very important extension of the SIFT method. The ASIFT method is a fully affine invariant. This algorithm detects corresponding points within two widely-separated views much more reliably than other related methods [7]. In our previous work, this method was upgraded by the so-called ASIFT-SH method [19]. A new feature detector and descriptor was proposed by Hauagge and Snavely [21]. This new method is based on detecting and representing the local symmetries that are fundamental characteristics of many urban images. In many cases regarding urban environment, the proposed features are more invariant to large viewpoint changes than a SIFT feature.

* Corresponding author. Tel.: +386 2 229 43 18.
E-mail address: peter.podbreznik@um.si (P. Podbreznik).

1474-0346/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aei.2012.11.001

Over recent years, a very wide-range of computer vision applications have been developed that deal with building-site monitoring. They mostly require correspondence estimation. In [22], a fast and simple application was proposed for location recognition by using imaging materials. The matching of crowded images is based on SIFT features. Images were taken from large Internet photo collections, provided by different authors. This matching algorithm can also be used to find any images of specific persons in the scene [23]. The matching procedure is based on summing the number of positively classified pixels inside and those pixels immediately surrounding each aligned part. 3D reconstruction from large unstructured collections of images, downloaded from the Internet, is mainly carried out by using incremental algorithms that solve progressively larger bundle adjustment problems. In [24], the authors presented an alternative approach based on hybrid discrete–continuous optimisation for determining an absolute 3D camera position, where the SIFT method is used for correspondence estimation. The same matching method was applied during similar research, where a 3D reconstruction system was updated with novel parallel-distributed matching and reconstruction modules [25]. An extension of 3D reconstruction from a large collection of urban images' are results in 4D cites [26], which is actually time-varying 3D model. The reconstruction process requires a 3D pose of the camera, internal parameters, and the date of photographing, where image-matching is also applied using the SIFT method.

Establishing correspondence between building site images and 4D model is, of course, an essential step for the successful recognition of a particular building's element and for identifying actual situations on a building-site (e.g., for detecting any discrepancies with respect to as-planned). Classical approaches for establishing correspondence are unsuitable, because the vertexes from a 4D model have a certain geometrical meaning and are, in general, not invariant to image rotation, translation, and scale. Fig. 1 depicts an example of properly established correspondence between a 4D model and images from both cameras' views.

This paper deals with upgrading the basic version of our ASIFT-SH method. This method [7] is one of the few efficient methods for establishing correspondence between arbitrarily-selected points within two widely-separated views. Its main strength is that it is capable of finding a correspondence, even for pixels, within an area with certain uniform properties. This is especially important for building elements (e.g., walls) having similar colours or textures. The above-reviewed methods developed earlier than the ASIFT-SH method, fail in such regions.

The ASIFT-SH method during its segmentation step presupposes that the segmented regions meet a coplanarity criterion (i.e., all points from the segmented region can be 3D reconstructed to the same plane). This is an unattainable requirement for the majority of segmentation methods. Therefore in this article, we have introduced a step for the adaptive adjustment of obtained segmentation results into the existing ASIFT-SH method. This step is based on a 3D reconstruction of the initial corresponding points and a search for the minimal number of planes within the 3D space, to which these point belong. Our proposed method can also established correspondence in more general situations that on building sites. However, the observed objects should have planar surfaces.

This paper is organised as follows. The ASIFT-SH method is briefly reviewed in Section 2. The newly-introduced refinement of the ASIFT-SH method is explained in detail in Section 3, followed by a presentation of the results in Section 4. The results are discussed in Section 5, and this paper concludes with some future work directions, as presented in Section 6.

2. Review of the basic ASIFT-SH method

The ASIFT-SH method from our previous work was developed to establish correspondence between arbitrarily-selected points (e.g., points satisfying certain geometrical criteria with respect to a 3D model viewed over time) within two widely-separated views [7]. This four-step method is a mixture of the ASIFT method, segmentation, and local homography. The method is briefly reviewed in the following subsections.

2.1. Establishing an initial corresponding point

Candidates for initial corresponding points are established from the images of both the widely-separated camera views, by using the ASIFT method [19]. An established rough correspondence is employed to reduce number of the outliers. Some of outliers are eliminated by applying an epipolar constraint.

2.2. Grouping the coplanar initial corresponding points (segmentation)

The initial corresponding points are then grouped into regions by using image segmentation. It is assumed that each of the obtained segmented regions contains only coplanar image points. The condition of coplanarity is fulfilled if the entire segmented region is part of a flat surface. A subset of the corresponding point pairs $\mathcal{S}_j$ is created for any segmented region $\mathcal{R}_j$. The subset $\mathcal{S}_j$, defined as $\mathcal{S}_j = \{(\boldsymbol{p}_i, \boldsymbol{p}_i'); \forall i \boldsymbol{p}_i \in \mathcal{R}_j$, consists of the initial corresponding points $\boldsymbol{p}_i$, which belong to the region $\mathcal{R}_j$ on the first camera's image, whilst $\boldsymbol{p}_i'$ denotes its corresponding point from the second camera's image.

2.3. Calculating local homographies

The relationship between the arbitrarily-selected and its corresponding point is established through the equation:

$$\boldsymbol{p}_j' = \mathcal{H}_j^{-1} \boldsymbol{p}_j, \tag{1}$$

where $\boldsymbol{p}_j$ is the arbitrarily-selected point from the first image $\mathcal{I}$, $\boldsymbol{p}_j'$ denotes its corresponding point from the second image $\mathcal{I}'$, and the matrix $\mathcal{H}$ defines the homography [27]. In the ASIFT-SH method, the DLT algorithm is used for determining the local homographies (matrices $\mathcal{H}_j$) for all the sets $\mathcal{S}_j$ [7].

Fig. 1. Matching between 4D/3D models and an image of the first camera (dotted line); correspondence between the points from the first and the second images (solid line).

2.4. Determining correspondence for an arbitrary point

By using Eq. (1), the homography is applied for directly determining the corresponding point on the second image for a selected-point on the first image. Let us consider an arbitrarily-selected point $\boldsymbol{p}$ from the first image that belongs to a segmented region $\mathcal{R}_j$. Region $\mathcal{R}_j$ contains a set of initially corresponding points' pairs, $\mathcal{S}_j$. The local homography (defined by matrix $\mathcal{H}_j$) is calculated for region $\mathcal{R}_j$ based on the subset $\mathcal{S}_j$. A corresponding point $\boldsymbol{p}'$ on the second image for the arbitrarily selected point $\boldsymbol{p}$ is then uniquely determined by using Eq. (1) and the local homography $\mathcal{H}_j$.

3. Self-adaptive ASIFT-SH method

The most sensitive, and from an accuracy point of view, the most critical step of the ASIFT-SH method is segmentation. The segmentation procedure presupposes that the obtained segmented regions meet a coplanarity criterion (i.e., all points from the segmented region can be 3D reconstructed to the same plane). In general, this cannot be ensured.

The coplanarity condition might be fulfilled to some extent if a suitable segmentation method (e.g. [28]) is applied using a correct parameter set but, in general, this condition cannot be assured in all cases. The extension to the ASIFT-SH method aims at reshaping or remodelling those segmented regions that are not coplanar according to our definition.

The extension of the ASIFT-SH method, i.e. the Self-adaptive ASIFT-SH method, consists of two phases. During the first phase, the non-coplanar segmented regions $\mathcal{R}$ are identified, whilst during the second phase each of the identified regions is reshaped into two or more regions that better suit the coplanarity criterion. Both steps are described, in detail, in this sequel.

3.1. Identification of non-coplanar segmented regions

The identification of non-coplanar segmented regions is based on the following procedure. Each segmented region $\mathcal{R}_j$ holds a subset, $\mathcal{S}_j$ of the initial corresponding points. In special cases, the particular subsets could also be empty. The initial corresponding points from the subset $\mathcal{S}_j$ are 3D reconstructed. These reconstructed points are then used to calculate the best-fitting plane for the points from the segmented region $\mathcal{R}_j$. A plane for the segmented region $\mathcal{R}_j$ is established by using the singular value decomposition method (SVD), as follows:

1. The 3D reconstructed initial corresponding points from subset $\mathcal{S}_j$ are composed into a matrix $\mathcal{P}$ as:

$$\mathcal{P}_j = \begin{pmatrix} P^j_{1_x} & P^j_{1_y} & P^j_{1_z} & 1 \\ P^j_{2_x} & P^j_{2_y} & P^j_{2_z} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ P^j_{n_x} & P^j_{n_y} & P^j_{n_z} & 1 \end{pmatrix}, \tag{2}$$

   where $P^j_{1_x}, P^j_{1_y}$, and $P^j_{1_z}$ denote the x, y, and z coordinates of the first reconstructed point from the subset $\mathcal{S}_j$, whilst n stands for the number of all the reconstructed initial corresponding points in subset $\mathcal{S}_j$.
2. The matrix $\mathcal{P}$ is decomposed by using the SVD method as:

$$[\mathcal{U}, \mathcal{D}, \mathcal{V}] = \mathrm{SVD}(\mathcal{P}), \tag{3}$$

   where $\mathcal{U}$ and $\mathcal{V}$ denote the orthogonal matrices and $\mathcal{D}$ stands for a diagonal matrix. The eigenvector from the matrix $\mathcal{V}$, which belongs to the smallest value from matrix $\mathcal{D}$, presents the equation's coefficients (in implicit form) for plane Π. The coefficients of plane Π are determined in such a way that the mean distance is minimal for all the reconstructed points from subset $\mathcal{S}_j$ to this plane. The eigen values are sorted in descending order in matrix $\mathcal{D}$, therefore, the last column of matrix $\mathcal{V}$ presents the solution, i.e. the coefficients of the equation in implicit form for the reconstructed plane Π.

It is known that the SVD method can be affected by the outliers (i.e., far off non-coplanar points). If several outliers are expected then the RANSAC-SVD approach is suggested instead of the SVD method. Identification of the non-coplanar segmented region is based on measuring the Euclidean distance between the 3D reconstructed corresponding points from subset $\mathcal{S}_j$ and their associated reconstructed plane Π_j, i.e. the calculated plane for segmented-region $\mathcal{R}_j$. If the mean distance $\bar{\mathrm{d}}$ and the standard deviation σ of these distances are greater than the predefined threshold, the following conclusions can be drawn:

1. If the mean distance $\bar{\mathrm{d}}$ is much smaller than the standard deviation σ, then a small group of points exists, which are non-coplanar with respect to other points from this subset inside subset $\mathcal{S}_j$; or
2. If the mean distance $\bar{\mathrm{d}}$ and standard deviation σ have the same magnitude, then a larger group of points exists, which are non-coplanar with respect to other points from this subset inside subset $\mathcal{S}_j$.

3.2. Reshaping a non-coplanar segmented region

If the segmented region $\mathcal{R}_j$ does not meet the condition of coplanarity, then this region has to be remodelled or split. At this point of our algorithm, an additional restriction is set: each new region obtained after reshaping region $\mathcal{R}_j$ has to contain at least 4 initial corresponding points. This restriction is set to enable the calculation of local homography $\mathcal{H}_j$ (see Sub Section 2.3). A reshaping of the non-coplanar segmented region is based on the RANSAC method [29]. The RANSAC method is an iterative method that enables the grouping of entities with selected properties. In our case, the distance of the ith reconstructed point, as denoted by $\boldsymbol{P}^j_i$, from the reconstructed plane Π_j, is considered as a selected property. The procedure for the reshaping of the non-coplanar segmented region $\mathcal{R}_j$ (and consequently the subset $\mathcal{S}_j$) by using RANSAC method, consists of the following steps:

1. A set of samples is created from all the reconstructed corresponding points from subset $\mathcal{S}_j$, i.e. $\{\mathcal{Q}^j_i, i = 1, \ldots, m\}$, where m is the number of samples. Each sample $\mathcal{Q}^j_i$ contains p randomly-selected reconstructed corresponding points from subset $\mathcal{S}_j$. The reconstructed plane Π^j_i is determined by using Eqs. (2) and (3) for the points from sample $\mathcal{Q}^j_i$. This plane belongs to sample $\mathcal{Q}^j_i$.
2. For each plane Π^j_i, determined by using the corresponding points from sample $\mathcal{Q}^j_i$, a degree of fitting is calculated for all the reconstructed corresponding points from subset $\mathcal{S}_j$. Two criteria are observed:

(a) the mean distance $\bar{\mathrm{d}}^j_i$ of all the reconstructed corresponding points from subset $\mathcal{S}_j$ to the plane Π^j_i, and
(b) the number of reconstructed corresponding points from subset $\mathcal{S}_j$, the distances from within the reconstructed plane Π^j_i are smaller than threshold r. The number of these points is denoted by n^j_i.

   A coefficient q^j_i is formed from both calculated values for the particular sample $\mathcal{Q}^j_i$ as:

$$q^j_i = n^j_i / \bar{\mathrm{d}}^j_i. \tag{4}$$

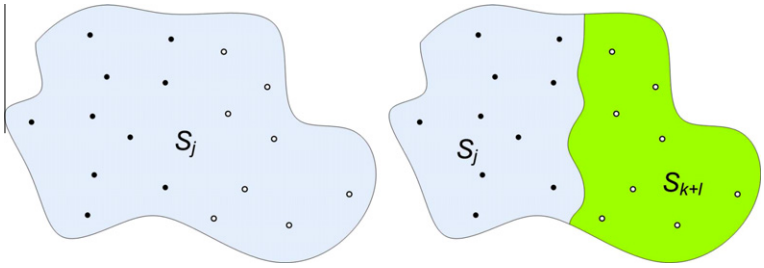

Fig. 2. Schematic presentation of the non-coplanar region division: situation before division (left) and situation after division of the segmented region $\mathcal{R}_j$ (right). Filled circles represent corresponding points from the reduced subset $\mathcal{S}_j$, whilst the empty circles represent the corresponding points from the newly-created subset $\mathcal{S}_{k+l}$.

In this way, the defined coefficient q_i^j equally considers the above-defined criteria (a) and (b).

3. Afterwards, the largest coefficient q_{max}^j is sought, i.e. $q_{max}^j = \max_i q_i^j$, where $i = 1, \ldots, m$. The reconstructed plane Π_{max}^j for sample $\mathcal{Q}_{max}^j$ associated with the largest coefficient q_{max}^j is a solution in the current iteration. The plane Π_{max}^j has the best degree of fitting for all 3D reconstructed corresponding points from subset $\mathcal{S}_j$ of segmented region $\mathcal{R}_j$. In other words, the plane Π_{max}^j is a plane to which the majority of reconstructed points from subset $\mathcal{S}_j$ fit the best. Those reconstructed corresponding points with a distance from the plane Π_{max}^j greater than the pre-defined threshold r, are moved into the new subset $\mathcal{S}_{k+l}, l = 1, \ldots, z$, where z is the number of all identified non-coplanar segmented regions (k is the number for all the initial segmented regions).
4. The non-coplanar region/subset $\mathcal{S}_j$ is, after step 3, split into a reshaped (reduced) subset $\mathcal{S}_j$ and into the newly-created subset $\mathcal{S}_{k+l}$. This division of the initial corresponding points from the original subset $\mathcal{S}_j$ into two subsets, also affects the segmented region $\mathcal{R}_j$. The segmented region $\mathcal{R}_j$ is divided by being reshaped into two parts, as well. The reshaping of segmented region $\mathcal{R}_j$ is carried out by using a region-growing segmentation method, as proposed by Potočnik and Zazula [30], where two initial kernels (i.e., homogenous regions) are used. Each kernel consists of the initial corresponding points from subsets $\mathcal{S}_j$ and $\mathcal{S}_{k+l}$, respectively. Both kernels start the growing process at the same time and at the same speed. The merging of (both) regions is thus prevented. The schematic results from this procedure are depicted in Fig. 2.

After the first iteration of the above-described procedure, the z-segmented regions are reshaped and the same number of new regions created. The reshaped segmented regions successfully fulfil the condition of coplanarity (see subset $\mathcal{S}_j$ in Fig. 2 on the right), but for the newly-created regions (e.g. see subset $\mathcal{S}_{k+l}$ in Fig. 2) this condition still needs to be checked. The next iteration validates the coplanarity of these newly created regions. The steps described above are iteratively repeated, until all the regions successfully fulfil the condition of coplanarity. It should be stressed that regions with less than 4 corresponding points cannot be non-coplanar.

3.3. Parameters estimation

The Self-adaptive ASIFT-SH method is a mixture of several methods, such as: the ASIFT method for the initial corresponding points' estimation, the DTL method for the homography calculation, graph-based method for segmentation, and the RANSAC method. The segmentation method and RANSAC have a strong impact on the final results. A procedure for estimating the parameters of the segmentation method has already been exactly described by Podbreznik and Potočnik [7]. The following parameters' values were determined: $\gamma = 1$, $K = 1000$, and $min = 2000$.

The second method having a major impact on the final results, is RANSAC. This method requires the following parameters: (i) the number of samples, m, (ii) the number of points within a particular sample p, and (iii) the threshold r. The threshold r controls the allowed distance of the reconstructed point $\boldsymbol{P}_i$ from the associated reconstructed plane Π. The number of samples m, is in correlation with the number of outliers from the subset $\mathcal{S}_j$ (i.e., points with a distance from the reconstructed plane Π_j greater than the threshold r). The number of points within the particular sample $\mathcal{Q}$, mostly depends on the considered property. The number of samples m, is calculated, as proposed by [31]:

$$m = \frac{\log(1-\Upsilon)}{\log[1-(1-\varepsilon)^p]}, \tag{5}$$

where Υ stands for the degree of probability that the points from the winning sample $\mathcal{Q}_{max}^j$ will provide the best determination of those parameters that describe the selected property. Variable Υ denotes the ratio of those reconstructed corresponding points that have distances from the reconstructed plane Π greater than the pre-defined threshold r ($\varepsilon \in [0,1]$), whilst the number of points in the samples is denoted by the variable p.

For an example, let us assume that the reconstructed corresponding-points from a region $\mathcal{R}$ lie on two different planes. Statistically, the value of variable ε is then *0.5*. The winning sample has to have a high-degree of probability that the obtained solution will ensure the best fittings for all point from the subset $\mathcal{S}_j$. In this case, the variable Υ must be very close to 1 (e.g. $\Upsilon = 0.99$). Let the number of points in a particular sample be fixed at 3. For these assumptions, the required number of samples m is calculated by using Eq. (5) as 35 ($m = 35$).

The parameters for the RANSAC method were obtained experimentally in our research. Several tests of the Self-adaptive ASIFT-SH method were carried out by varying the parameters for the RANSAC method. The best results were obtained when the number of samples was fixed at 300 (i.e. $m = 300$), the number of points within a particular sample was 6 (i.e. $p = 6$), and the degree of probability Υ was 0.99. The determination of variable ε is based on the assumption that more than 50% of the reconstructed corresponding-points have smaller distances from the appropriate plane, as is defined by threshold r. Thus, the variable ε was estimated to be $\varepsilon = 0.5$.

4. Results

The results from our experiments are presented in this section. Firstly, the experimental environment is described, which was used for the testing of our proposed method, then the obtained qualitative and quantitative results are presented.

4.1. Experimental environment

The proposed Self-adaptive ASIFT-SH method designed for the correspondence-estimation of arbitrarily-selected points from two widely-separated views, was tested on real-world images, i.e. building site images. Fifty-five testing image pairs were considered in our experiments, whereas 10 image pairs are depicted in Fig. 3. For the term "testing image pair" we also use in the sequel the expression "measurement". The sizes of images were 2670×2003 pixels for the landscape-oriented images and 2003×2670 pixels for the portrait-oriented images. The image resolution varies with the distance of the camera from an object. The location of the calculated corresponding point in the second image for an arbitrarily-selected point from the first image was observed during these experiments. For evaluation purposes, the so-called "testing" corresponding point pairs $(\boldsymbol{p}_i, \boldsymbol{p}_i')$ were determined in the following two ways:

1. The corresponding points $(\boldsymbol{p}_i, \boldsymbol{p}_i')$ were determined manually for the first 15 tested image pairs, as follows:
   (a) 100 points were manually selected within the first image. These points were also visible in the second image;
   (b) for each selected point from the first image, an observer manually determined its corresponding point in the second image;
   (c) three different observers carried out step 2;

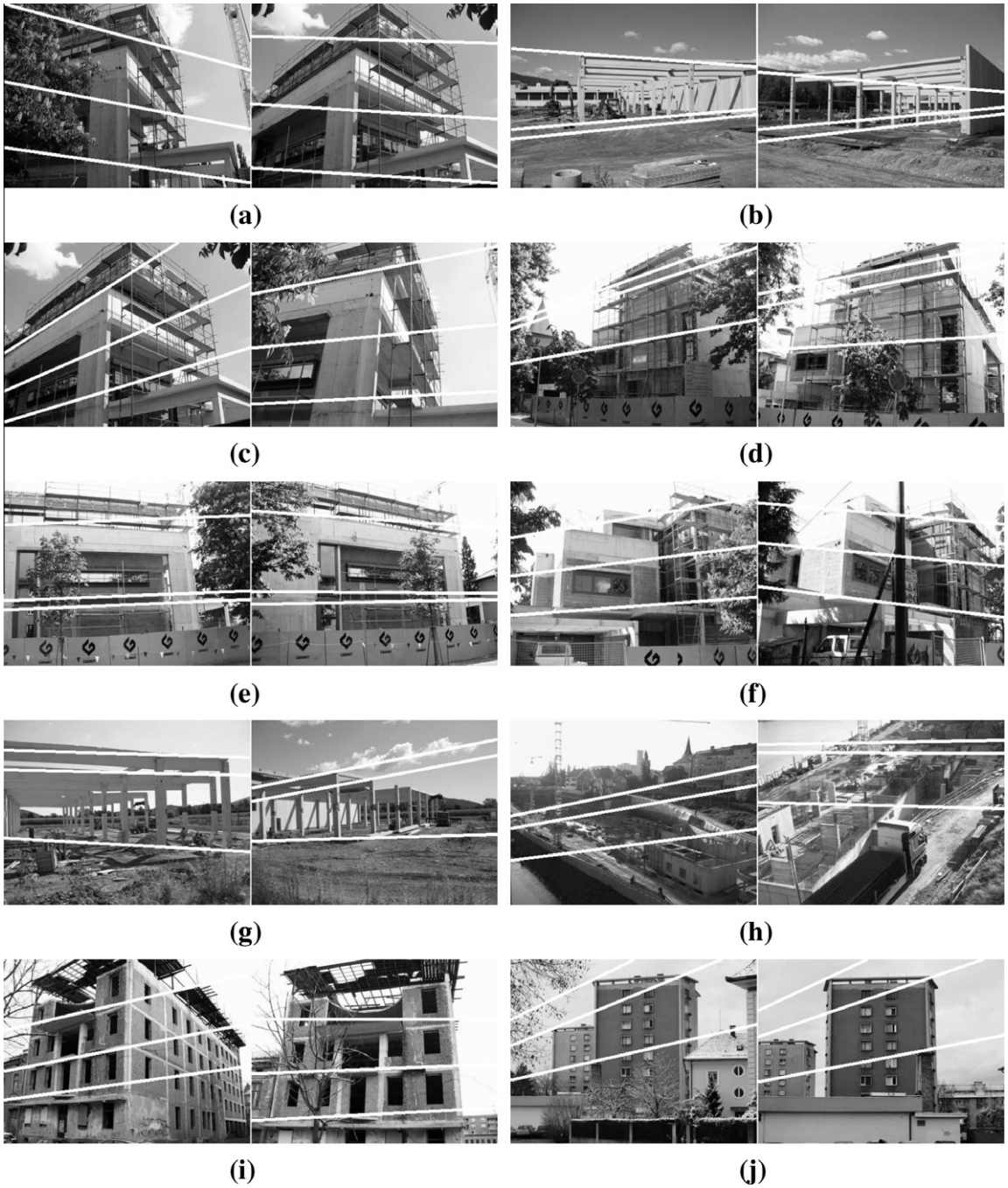

Fig. 3. Testing image pairs: (a) Measurements 1, (b) Measurements 2, (c) Measurements 3, (d) Measurements 4, (e) Measurements 5, (f) Measurements 6, (g) Measurements 7, (h) Measurements 8, (i) Measurements 9, and (j) Measurements 10. The first camera's view is on the left, while the second camera's view is on the right of each image pair. The original image's size was 2670×2003 pixels. White lines represent samples of the epipolar lines.

(d) for the selected point $\boldsymbol{p}_i$ from the first image, the actual corresponding point $\boldsymbol{p}'_i$ within the second image was determined as the average reading from all three observers;
(e) a set of "testing" corresponding point pairs $(\boldsymbol{p}_i, \boldsymbol{p}'_i), i = 1, \ldots, 100$, was considered as the Ground truth during our experiments, and the evaluation procedure.

In these cases, selected points were usually determined on the corners or any other areas where human vision recognises whether the corresponding points are accurate enough. It should be noted that the calculated intra-observer variability for the construction of the Ground truth was 1.31 pixels (measured as absolute), with a standard deviation of 1.01.

2. The above-described procedure was very time consuming. Therefore, for the next 40 tested pairs of images, the initial corresponding points were determined by using the ASIFT method. Generally, the ASIFT method ensures very accurate results [19], but, as already explained, so called dense matching cannot be provided using this method. Because of the high accuracy of the ASIFT method, the corresponding points generated using this method can be used as "testing" corresponding point pairs $(\boldsymbol{p}_i, \boldsymbol{p}'_i)$, i.e. as the Ground truth. All the outliers from this set were manually eliminated. In this way, the rest of the corresponding points in the Ground truth, were established with complete accuracy.

The Self-adaptive ASIFT-SH method's accuracy was measured as a discrepancy (error) between $\boldsymbol{p}^*_i$ and $\boldsymbol{p}'_i$, i.e. as a distance between $\boldsymbol{p}^*_i$ and $\boldsymbol{p}'_i$ ($\mathrm{d}(\boldsymbol{p}^*_i, \boldsymbol{p}'_i)$), where $\boldsymbol{p}^*_i$ was a calculated corresponding point using the Self-adaptive ASIFT-SH method, and $\boldsymbol{p}'_i$ was the actual corresponding point (i.e. ground truth) for point $\boldsymbol{p}_i$ from the first image.

4.2. Qualitative and quantitative results

For a selected point $\boldsymbol{p}_i$ from the testing set (i.e. Ground truth), a corresponding point $\boldsymbol{p}^*_i$ on the second image was calculated by using our method. A discrepancy d, defined as the Euclidean distance between $\boldsymbol{p}^*_i$ (i.e. calculated point) and $\boldsymbol{p}'_i$ (point from Ground truth), was measured in pixel unit. Tables 1–3 present the obtained results. Table 1 presents those results from the first 15 measurements, where the corresponding points in the ground truth were established manually. The other two tables, i.e. Tables 2 and 3, present the obtained results from the next 40 measurements. In this case, the corresponding points in the Ground truth were determined automatically by using the ASIFT method, as described in Section 4.1.

All three tables consisted of two parts. In the first part, the results for the basic version of the ASIFT-SH method are presented, whilst the results from the Self-adaptive ASIFT-SH method are collected in the second part. Each part within three columns. The first and second columns, i.e. columns $\bar{\mathrm{d}}$ and σ, present the mean and standard deviation of the discrepancy (error). The third column, i.e. column ε, presents the percentage of initial corresponding points for which correspondence could not be calculated, because the matrix $\mathcal{H}$ did not exist for the region they belonged to (see the third step of the ASIFT-SH method). The last column in these tables, i.e. column n, presents the number of corresponding pairs within the Ground truth.

Table 1
Mean discrepancy ($\bar{\mathrm{d}}$) and standard deviation (σ) determined between corresponding points calculated by Basic or Self-adaptive ASIFT-SH method and set of "testing" corresponding point pairs (i.e. Ground truth). Column n presents a number of corresponding point pairs, while ε denotes a percentage of corresponding points for which the correspondence could not be calculated. Results for the first 15 measurements are shown.

Measurement	Basic ASIFT-SH			Self-adaptive ASIFT-SH			
	$\bar{\mathrm{d}}$	σ	ε (%)	$\bar{\mathrm{d}}$	σ	ε (%)	n
1	7.53	39.75	1	4.94	15.94	2	100
2	10.69	22.68	0	4.93	16.65	0	100
3	4.6	11.77	0	2.28	6.42	2	100
4	5.94	17.75	1	5.49	29.81	2	100
5	57.85	91.77	0	56.93	121.72	0	100
6	33.89	39.44	0	25.55	132.56	1	100
7	6.63	14.34	0	4.11	11.14	1	100
8	2.63	2.25	0	2.12	1.21	0	100
9	3.34	6.44	0	2.14	5.12	1	100
10	24.69	224.66	0	3.65	36.77	0	100
11	6.96	14.97	0	3.18	12.34	2	100
12	1.64	1.55	0	1.62	1.61	0	100
13	3.94	7.39	0	2.33	6.5	2	100
14	6.44	3.55	0	4.22	2.21	0	100
15	9.94	7.39	0	4.13	6.35	1	100

5. Discussion

This discussion begins with a short analysis of the basic version of the ASIFT-SH method [7]. Two types of problems were identified by studying the basic methods: (i) potentially mis-segmented regions (this problem is substantiated by the example in "Measurement 10"), and (ii) potentially mis-matched initial corresponding points (this problem is substantiated by the examples in "Measurement 5" and "Measurement 6"). Let us explain both problems in detail.

5.1. Shortcomings of the basic ASIFT-SH method

Image mis-segmentation: The quality and accuracy of image segmentation is crucial for the Basic ASIFT-SH method. The segmentation procedure, as proposed by [28], was employed in our solution. In most cases, this segmentation works very well; however, there are some situations and image parts, where the image is segmented erroneously. An incorrectly segmented image, of course, leads to inaccuracy within the ASIFT-SH method. Such an example is "Measurement 10" (see also Table 1 for quantitative results), where the images are apparently mis-segmented. Incorrect segmentation leads to several inconveniences: (i) segmented region $\mathcal{R}_j$ contains non-coplanar points and, thus, an incorrect matrix $\mathcal{H}_j$ is calculated; (ii) the arbitrarily-selected point $\boldsymbol{p}_i$ can belong to the wrong region and, thus, an incorrect matrix $\mathcal{H}$ could be used; (iii) the arbitrarily-selected point $\boldsymbol{p}_i$ can belong to the correct region; however, matrix $\mathcal{H}$ is inaccurately calculated (non-coplanar points) and, consequently, the corresponding point is completely miscalculated.

A measurement with very high mean and standard deviation—see Table 1—is "Measurement 10" (see Fig. 3j). The image in this measurement is accurately segmented almost everywhere except in those areas around the chimney, as depicted in Fig. 4b. It can be clearly seen that almost the whole chimney is mis-segmented as part of one side of the residential building. The matrix $\mathcal{H}$, calculated by using points from this region, is certainly wrong and, of course, leads to significant errors during correspondence estimation (e.g. a very expressive parallax [27]).

Mis-matched initial corresponding points: A completely different problem was found in "Measurement 5" and "Measurement 6", although the mean $\bar{\mathrm{d}}$ and standard deviation σ for both measurements were relatively high. In these examples, the main reason is not only in the segmentation method but, above all, in the ASIFT method used to determine the initial corresponding points. On both images, the ASIFT method mis-matched some of the initial corresponding points. Especially problematic was a building railing with too similar black symbols (see also Fig. 3d and e). Moreover, the mis-matched points (i.e., outliers) could not have been eliminated, even by considering epipolar constraint. The mis-matched corresponding points mainly resided very close to the appropriate epipolar lines and were, therefore, not eliminated. Such a situation

Table 2
Mean discrepancy ($\bar{d}$) and standard deviation (σ) determined between a set of "testing" corresponding point pairs (i.e. Ground truth) and corresponding points, calculated by the basic or Self-adaptive ASIFT-SH method. The corresponding points in Ground truth were determined automatically by using ASIFT method. Column n presents a number of "testing" corresponding point pairs, while ε denotes a percentage of corresponding points for which the correspondence could not be calculated. Results for 20 out of 55 measurements are shown.

Measurement	Basic ASIFT-SH			Self-adaptive ASIFT-SH			
	$\bar{d}$	σ	ε (%)	$\bar{d}$	σ	ε (%)	n
16	8.09	56.56	0	3.96	25.25	0.13	3007
17	13.86	37.62	0.02	3.44	16.2	0.07	6066
18	21.06	115.34	0	9.44	39.1	0.17	2999
19	51.63	602.97	0.03	25.98	123.13	0	3130
20	31.67	59.6	0	21.11	45.59	0.05	2180
21	15.75	121.88	0	7.61	13.49	0.14	5090
22	4.17	8.62	0	2.61	7.31	0.01	29,328
23	19.75	28.72	0.06	8.51	25.69	0.09	3359
24	2.52	6.48	0	1.82	4.22	0	53,281
25	27.81	55.19	0	15.63	49.06	0.04	2426
26	12.86	25.15	0	8.04	16.75	0.05	8200
27	20.51	413.98	0	9.06	35.78	0.01	12,409
28	207.26	12369.66	0	77.32	2777.31	0.02	9072
29	17.36	27.17	0.01	12.09	303.41	0.03	7500
30	9.92	11.22	0	3.42	6.38	0.02	13,471
31	15.23	24.97	0	6.79	13.17	0.05	8529
32	24.74	34.3	0	14.97	44.28	0.25	2795
33	30.76	68.06	0	28.09	97.4	1.65	846
34	17.9	33.23	0	7.11	21.37	0	27,965
35	18.73	21.85	0	6.97	16.43	0	32,838

Table 3
Mean discrepancy ($\bar{d}$) and standard deviation (σ) determined between a set of "testing" corresponding point pairs (i.e. Ground truth) and corresponding points calculated by the basic or Self-adaptive ASIFT-SH method. The corresponding points in Ground truth were determined automatically by using ASIFT method. Column n presents a number of "testing" corresponding point pairs, while ε denotes a percentage of corresponding points for which the correspondence could not be calculated. Results for the last 20 out of 55 measurements are shown.

Measurement	Basic ASIFT-SH			Self-adaptive ASIFT-SH			
	$\bar{d}$	σ	ε (%)	$\bar{d}$	σ	ε (%)	n
36	25.18	32.62	0	11.69	25.44	0	23,816
37	6.44	13.58	0	3.02	7.75	0	68,283
38	52.57	43.98	0	31.21	51.51	0	5395
39	29.5	26.02	0	22.46	26.35	0	36,257
40	48.38	55.31	0	26.09	38.47	0	20,273
41	10.12	23.66	0.01	3.96	16.2	0	12,654
42	30.2	51.73	0.01	9.56	21.9	0.03	7330
43	11.47	109.8	0	6.84	20.89	0.02	5972
44	11.49	18.99	0	5.78	12.41	0.05	6642
45	12.3	23.38	0	7.64	21.37	0.03	3156
46	10.73	24.55	0	7.08	24.66	0.08	3842
47	8.69	18.48	0	6.53	15.77	0	28,344
48	9.87	19.46	0	4.51	9.5	0.01	29,785
49	8.29	16.81	0	4.22	9.66	0	35,084
50	107.69	309.27	1	80.7	366.84	1.06	567
51	41.14	532	1.16	33	373.07	1.16	688
52	25.97	225.65	0	7.55	68.17	0.66	1362
53	10.39	39.47	1	8.7	36.85	1.13	1537
54	4.24	33	1.23	3.1	33.07	1.23	6438
55	22.47	25.45	0	7.1	28.17	0.56	2342

is depicted in Fig. 5, where elimination of the outliers was unsuccessful.

The column ε from Tables 1–3 denotes the percentage of points for which the correspondence could not be calculated. There are two main reasons that correspondence could not be determined: (i) segmented region $\mathcal{R}_j$ contained less than four initial corresponding points, and (ii) the arbitrarily-selected point did not actually have correspondence within the second image (i.e., this part of the object was hidden or invisible).

All the mentioned problems were more thoroughly discussed and some solutions proposed by Podbreznik and Potočnik [7].

5.2. Self-adaptive ASIFT-SH method

The Self-adaptive ASIFT-SH method addresses all the problems of the basic ASIFT-SH method. Our proposed method introduces a procedure for the 3D reconstruction of corresponding points, and a procedure when searching for a minimal number of planes within the 3D space to which these points belong. The results from the Self-adaptive ASIFT-SH method are presented in the second parts of Tables 1–3. For almost all the measurements, the mean discrepancy $\bar{d}$ and the standard deviation of discrepancies σ decreased when compared with the basic version of the ASIFT-SH method (see the results from the first parts of Tables 1–3). This improvement is a consequence of reshaping those segmented regions, identified as non-coplanar.

An example of such reshaping for "Measurement 10", is depicted in Fig. 6. The segmentation result from the basic version of the ASIFT-SH method, is shown in Fig. 6a. It can be seen that the areas around the chimney were segmented inaccurately. The result from the reshaping procedure carried out in the Self-adaptive ASIFT-SH method, is clearly presented in Fig. 6b. It is

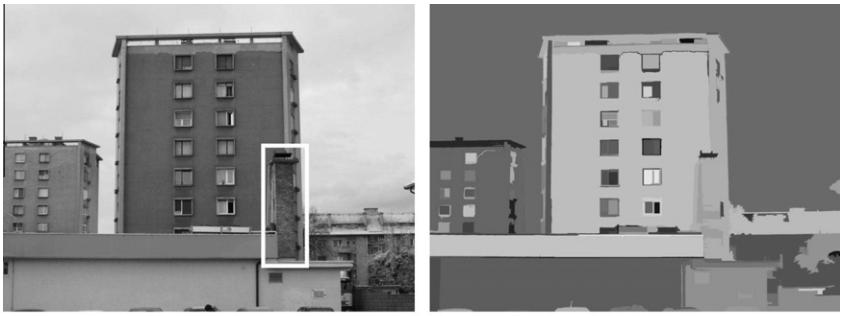

Fig. 4. "Measurement 10": (a) original image and (b) segmented image. Mis-segmented area around chimney is highlighted with a white rectangular border on the left image.

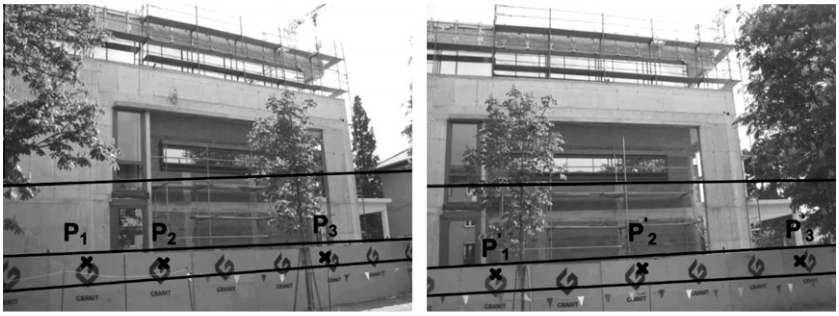

Fig. 5. "Measurement 5": (a) the first camera view and (b) the second camera view. Black crosses represent mis-matched initial corresponding points, given by the ASIFT method. Only 3 out of 12 mis-matched corresponding point pairs are depicted.

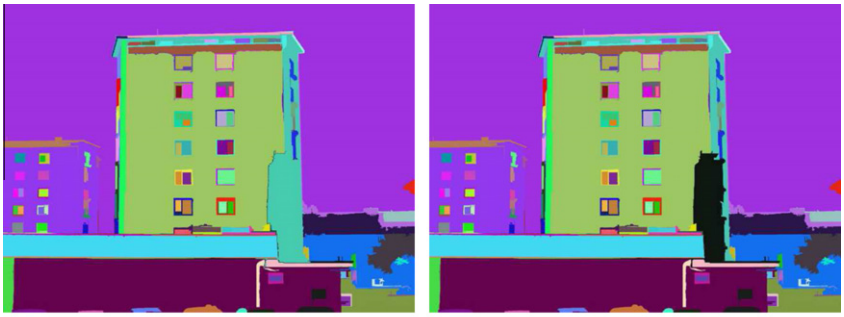

Fig. 6. Example of reshaping for "Measurement 10": (a) segmentation result of Basic ASIFT-SH method and (b) reshaped segmented region with Self-adaptive ASIFT-SH method (see a region around chimney).

evident that a non-coplanar segmented region (i.e., area around the chimney) had been divided into two new regions. Finally, these new regions better met the coplanarity criterion.

A potential problem of such reshaping could be the insufficient number of initial corresponding points within the particular region. In this case, we were unable to calculate the local homographies for such segmented regions. Consequently, it was impossible to determine correspondence for an arbitrarily-selected point from this region. In our solution, we have therefore integrated a security mechanism that prevents a newly-created region having less than 4 initial corresponding points (see subSection 3.2). The calculated mean ($\bar{d}$) and standard deviation (σ) of the discrepancies from Tables 2 and 3 are slightly greater when compared to the results from Table 1. The reason for this phenomenon lies in the imaging material. Namely, the distances between the camera and observed object were, for the first 15 measurements, significantly greater than for the other 40 measurements. Such an example is depicted in Fig. 7. In the left image (see Fig. 7a), an observed object was located around 200 m from the camera, whilst in the second image (see Fig. 7b) the distance between the camera and the observed object was only around 5 m. Of course, we perceived a major difference between the pixel sizes in both images. It is easy to calculate that a pixel size in Fig. 7b is 40 times smaller than in Fig. 7a. It should be stressed that such a major difference in distances between the observed objects and the camera was just an extreme case within our testing set. Generally, during the first 15 measurements, the mean distances between the camera and the observed objects are occasionally greater than during the other 40 measurements. Therefore, the results from Tables 2 and 3 have to be interpreted appropriately.

Extreme deviations were only noted in "Measurement 23" from Table 2. The main reason for such deviations lies in an unsuccessful

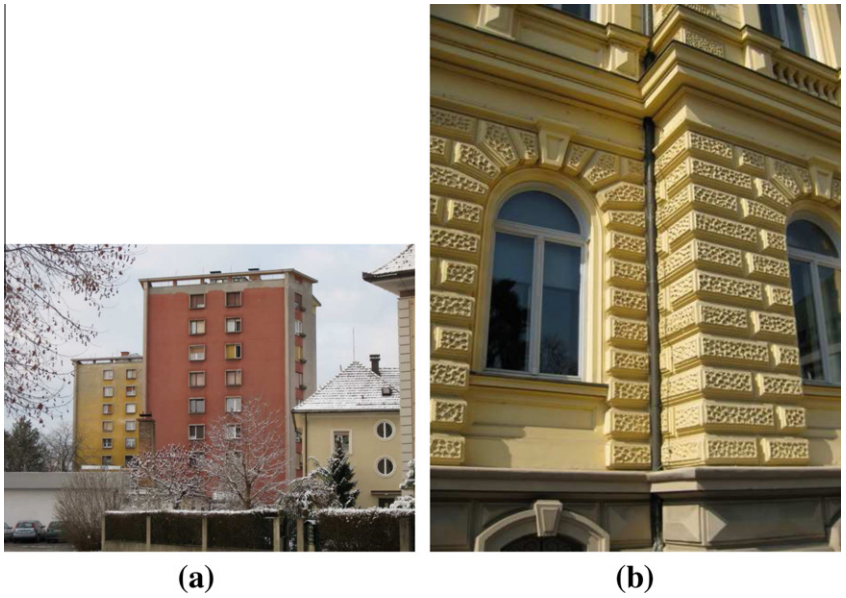

Fig. 7. Acquisition of imaging material from different distances of observed object from camera: (a) big distance (distance for "Measurement 10" is around 200 m) and (b) small distance (distance for "Measurement 27" is around 5 m).

segmentation procedure. An original image from the first camera is depicted in Fig. 8a, whilst Fig. 8b presents the segmentation result. Let us explain this problem. It can be seen from Fig. 8b that the segmented image contains very few regions. "Segment 1" (see Fig. 8c) covers practically the entire image, whereas the remaining area is distributed among a few other segments. For this reason, difficul-

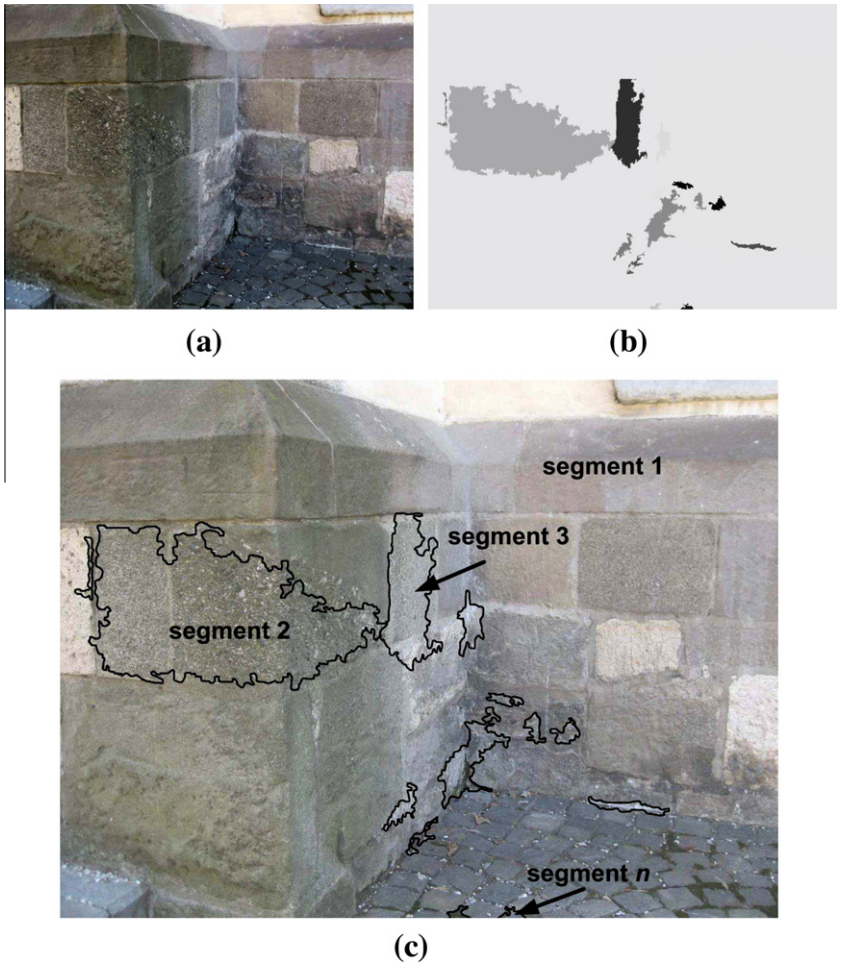

Fig. 8. Image from the first camera of "Measurement 23": (a) an original image, (b) unsuccessfully segmented image, and (c) an original image with overlaid segmentation result.

ties arise within the step of reshaping a non-coplanar segmented region (see sub Section 3.2). The 3D reconstructed corresponding points from "segment 1" lay in too many planes. The criterion defined by Eq. (4) proved to be insufficient during this measurement, resulting in an unsuccessful reshaping of the original set of initial corresponding points for "segment 1", thus provoking an extreme mean discrepancy $\bar{d}$, and standard deviation σ.

The second parts of Tables 1–3 present the improvements from the Self-adaptive ASIFT-SH method compared to those results obtained by the basic version of the ASIFT-SH method. A decrease in mean discrepancy $\bar{d}$ was noticed for all 40 measurements in Tables 2 and 3. The standard deviation σ also decreased in almost all the measurements. Based on the obtained results, we have concluded that the mean discrepancy $\bar{d}$ calculated over all 40 measurements decreased, on average, by 1.59 times, whilst the standard deviation σ decreased, on average, by 2.47 times with respect to the results obtained by using the Basic ASIFT-SH method. A similar trend is also noticeable in Table 1.

Special attention was given to assessing the applicability of our Self-adaptive ASIFT-SH method on uncompleted building structures (see Fig. 3). Measurements 1–9 from Table 1 present the mean discrepancy ($\bar{d}$) and the standard deviation (σ) for uncompleted building structures, whilst the Measurements 10–15 show the results for completed building structures. An average mean discrepancy and average standard deviation for the first nine measurements (without Measurements 5 and 6) were 3.71 and 12.3 pixels, respectively whilst for the completed building structures (i.e., Measurements 10–15) these values were 3.19 and 10.9 pixels. For Measurements 5 and 6 our method still had a problem with outliers (see Fig. 5 and the corresponding text). We believe that such situations are exceptions with very small probabilities of appearing in the future. For this reason the Measurements 5 and 6 were excluded from this calculation. The above results confirmed that our method is suitable either for either completed or uncompleted building structures.

The last columns of Tables 1–3, i.e. column ε, represents the percentage of those points from the ground truth for which calculation of the corresponding points on the second image was impossible. This percentage is slightly higher for the Self-adaptive ASIFT-SH method compared to the basic version of the ASIFT-SH method. The average ε for the Basic ASIFT-SH method was only 0.15%, whilst for the Self-adaptive ASIFT-SH method this percentage was slightly higher, namely 0.31%. The main reason for this increase in ε is in the procedure for reshaping the non-coplanar segmented regions. After each of the reshaping iterations, the non-coplanar segmented region is divided into two new regions. One of them is, of course, the coplanar whilst for the other region, the coplanarity criterion has to be verified during the next iteration. Due to a numerical error during the individual steps of our algorithm, this approach allows for certain deviations of the reconstructed corresponding points from the calculated plane Π. Those reconstructed corresponding points that do not meet the prescribed criteria form a new subset (and then a new segmented region within the existing one), for which the coplanarity conditions still need to be verified in the next iteration. After a few reshaping iterations, small set of points may still exist, which do not fulfil the criterion of coplanarity. It is a known fact that local homography cannot be calculated for a segmented region with less than four initial corresponding-points. Newly-segmented regions with less than four initial corresponding points actually contribute to an increase in the percentage of ε.

6. Conclusions

Using the adaptive adjustment of the ASIFT-SH method, segmented regions can be remodelled so that they better meet the required coplanarity criterion, as introduced in this article. The remodelling process is based on a 3D reconstruction of the initial corresponding points and a search for the minimal number of planes within the 3D space to which these initial corresponding points best-fit with respect to the Euclidean distance measurement. Those points that belong to a particular plane, represent a newly-created subset of the initial corresponding points.

It can be concluded from the obtained results, that the introduced additional step-efficiently adjusts the segmented regions based on 3D reconstruction. The new Self-adaptive ASIFT-SH method produces more accurately calculated corresponding points compared to those results from the basic version of ASIFT-SH, especially in those cases where the images were not segmented that accurately. The results pointed out that the average deviation (error) calculated through all measurements was 1.62 times lower and the standard deviation 2.51 times lower, compared to those results obtained by the basic ASIFT-SH method. On these bases, it can be concluded that the new Self-adaptive ASIFT-SH method enables very precise calculation of correspondence for arbitrarily-selected points within two-widely separated views and thus can efficiently deal with large viewpoint changes.

The percentage of points for which a correspondence on the second image could not be calculated was slightly higher for the Self-adaptive ASIFT-SH method than for its basic version. This problem could be mitigated by merging some smaller adjacent segmented regions.

The applicative value of this research can be seen from our research work in the field of building-site monitoring automation. Our subsystem for automatic building elements' recognition from site images, based on a 4D model, is still under development. By applying the introduced Self-adaptive ASIFT-SH method, the correspondence for an arbitrarily-selected point from the first image can be established more accurately compared with the previous version of the ASIFT-SH method. This has a strong impact on the accuracy of our subsystem for building elements' recognition.

References

[1] M. Golparvar-Fard, F. Peña Mora, S. Savarese, D4AR A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication, ITcon 14 (2009) 129–153.
[2] Y.M. Ibrahim, T.C. Lukins, X. Zhang, E. Trucco, A.P. Kaka, Towards automated progress assessment of workpackage components in construction projects using computer vision, Advanced Engineering Informatics 23 (2009) 93–103.
[3] S. Kiziltas, B. Akinci, E. Ergen, P. Tang, C. Gordon, Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management, ITcon 13 (2008) 134–154.
[4] E. Trucco, A.P. Kaka, A framework for automatic progress assessment on construction sites using computer vision, International Journal of IT in Architecture, Engineering and Construction 2 (2004) 147–164.
[5] I. Brilakis, M.-W. Park, G. Jog, Automated vision tracking of project related entities, Advanced Engineering Informatics 25 (2011) 713–724.
[6] M.-W. Park, A. Makhmalbaf, I. Brilakis, Comparative study of vision tracking methods for tracking of construction site resources, Automation in Construction 20 (2011) 905–915.
[7] P. Podbreznik, B. Potočnik, Estimating correspondence between arbitrarily selected points in two widely-separated views, Advanced Engineering Informatics 24 (2010) 367–376.
[8] P. Podbreznik, D. Rebolj, Automatic comparison of site images and the 4D model of the building, in: R.J. Scherer, P. Katranuschkov, S.-E. Schapke (Eds.), CIB W78 22nd Conference on Information Technology in Construction, Institute for Construction Informatics and Technische Universitat and Dresden, Dresden, Germany, 2005, pp. 235–239.
[9] N. Čuš Babič, P. Podbreznik, D. Rebolj, Integrating resource production and construction using BIM, Automation in Construction 19 (2010) 539–543.
[10] D. Rebolj, N. Čuš Babič, A. Magdič, P. Podbreznik, M. Pšunder, Automated construction activity monitoring system, Advanced Engineering Informatics 22 (2008) 493–503.
[11] D. Nashwan, M. Zaki, Construction workspace planning: assignment and analysis utilizing 4D visualization technologies, Computer Aided Civil and Infrastructure Engineering 21 (2006) 498–513.
[12] A. Baumberg, Reliable feature matching across widely separated views, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 774–781.
[13] E. Delponte, F. Isgrò, F. Odone, A. Verri, SVD-matching using SIFT features, Graphical Models 68 (2006) 415–431.

[14] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.
[15] P. Pritchett, A. Zisserman, Matching and reconstruction from widely separated views, in: 3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506, Springer-Verlag, 1998, pp. 78–92.
[16] P. Pritchett, A. Zisserman, Wide baseline stereo matching, in: Proceedings of the International Conference on Computer Vision (ICCV), pp. 754–760.
[17] C. Schmid, R. Mohr, Local greyvalue invariants for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 872–877.
[18] P.H.S. Torr, C. Davidson, IMPSAC: synthesis of importance sampling and random sample consensus, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2003) 354–364.
[19] J.-M. Morel, G. Yu, ASIFT: a new framework for fully affine invariant image comparison, SIAM Journal on Imaging Sciences 2 (2009) 438–469.
[20] G. Yu, J.-M. Morel, ASIFT: an algorithm for fully affine invariant comparison, Image Processing on Line (2011).
[21] D.C. Hauagge, N. Snavely, Image matching using local symmetry features, in: Computer Vision and Pattern Recognition (CVPR), pp. 206–213.
[22] Y. Li, N. Snavely, D.P. Huttenlocher, Location recognition using prioritized feature matching, in: Proceedings of the 11th European Conference on Computer Vision: Part II, Springer-Verlag, Heraklion, Crete, Greece, 2010, pp. 791–804.
[23] R. Garg, D. Ramanan, S.M. Seitz, N. Snavely, Where's Waldo: matching people in images of crowds, in: Computer Vision and Pattern Recognition (CVPR), pp. 1793–1800.
[24] D. Crandall, A. Owens, N. Snavely, D.P. Huttenlocher, Discrete-continuous optimization for large-scale structure from motion, in: Computer Vision and Pattern Recognition (CVPR), pp. 3000–3008.
[25] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S.M. Seitz, R. Szeliski, Building Rome in a day, Communications of the ACM 54 (2011) 105–112.
[26] G. Schindler, F. Dellaert, 4D cities: analyzing, visualizing, and interacting with historical urban photo collections, Journal of Multimedia 7 (2012) 124–131.
[27] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, 2004.
[28] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, International Journal of Computer Vision 59 (2004) 109–131.
[29] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.
[30] B. Potočnik, D. Zazula, Automated analysis of a sequence of ovarian ultrasound images. Part 1: segmentation of single 2D image, Image and Vision Computing 20 (2002) 217–225.
[31] Z. Zhang, Determining the epipolar geometry and its uncertainty: a review, International Journal of Computer Vision 27 (1998) 161–198.