FEATURE MATCHING ENHANCEMENT OF UAV IMAGES USING GEOMETRIC CONSTRAINTS

H. M. Mohammed* and N. El-Sheimy

Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive, NW, Calgary, Alberta, Canada T2N 1N4- (hmmohamm, elsheimy)@ucalgary.ca

Commission I, ICWG I/II

KEY WORDS: Feature Matching, Fundamental Matrix, Homography, Template Matching, 3D Reconstruction, UAV

ABSTRACT:

Preliminary matching of image features is based on the distance between their descriptors. Matches are further filtered using RANSAC, or a similar method that fits the matches to a model; usually the fundamental matrix and rejects matches not belonging to that model. There are a few issues with this scheme. First, mismatches are no longer considered after RANSAC rejection. Second, RANSAC might fail to detect an accurate model if the number of outliers is significant. Third, a fundamental matrix model could be degenerate even if the matches are all inliers. To address these issues, a new method is proposed that relies on the prior knowledge of the images' geometry, which can be obtained from the orientation sensors or a set of initial matches. Using a set of initial matches, a fundamental matrix and a global homography can be estimated. These two entities are then used with a detect-and-match strategy to gain more accurate matches. Features are detected in one image, then the locations of their correspondences in the other image are predicted using the epipolar constraints and the global homography. The feature correspondences are then corrected with template matching. Since global homography is only valid with a plane-to-plane mapping, discrepancy vectors are introduced to represent an alternative to local homographies. The method was tested on Unmanned Aerial Vehicle (UAV) images, where the images are usually taken successively, and differences in scale and orientation are not an issue. The method promises to find a well-distributed set of matches over the scene structure, especially with scenes of multiple depths. Furthermore; the number of outliers is reduced, encouraging to use a least square adjustment instead of RANSAC, to fit a non-degenerate model.

1. INTRODUCTION

Feature matching plays an essential role in photogrammetry, computer vision and remote sensing applications. The applications of feature detection and matching include, but not limited to, visual odometry, image mosaicking, 3D reconstruction, object tracking and many more.

The typical feature matching task comprises features detection, features descriptions, and preliminary matching. Preliminary matching between features in an image pair is usually done on the basis of the Euclidean distance between features. Matches are then filtered using the Random Sample Consensus (RANSAC) or a variant of RANSAC. RANSAC tries to fit a model using a subset of the matched features, then add more matches to the subset if they are consistent with the model. Hence, mismatches are discarded provided that RANSAC catches the correct model. Three problems are facing the block of detection, description and matching. First, descriptor-based matching is a probabilistic procedure, and it usually results in mismatches. Second, these mismatches are usually discarded by RANSAC and cannot be reconsidered for matching after RANSAC. Finally, even in cases of perfect inliers, RANSAC might lead to a degenerate solution to the fundamental matrix. The correspondences are termed degenerate when they are all inlier but still do not admit to a unique solution (Torr et al., 1998). We can conclude that the combination of descriptor-based matching with RANSAC is not the perfect solution for the feature matching problem. Especially in cases in which there is one predominant planar surface with many textures. In such a case, a large number of matches over the predominant surface forces RANSAC to follow a specific model which might not be able to represent other matching points off that plane. This scenario is better represented by a local homography rather than a fundamental matrix model.

It might be clear now that the main reason for the above problems is the separation between the radiometric properties of the features (when matching descriptors) and the geometric modelling (when using RANSAC). Therefore, to solve those problems, a new method is proposed on which geometrically and radiometric -aware matching is performed.

The rest of the paper is organised in the following structure: the next section covers a literature review of the related work followed by the paper contribution.The third section presents the proposed methodology in detail. The results and discussion are in the fifth section, and finally, the conclusion is given in the last section.

2. RELATED WORK AND PAPER CONTRIBUTION

Feature detection goes back to the early 80th. When Moravec published his work "Rover visual obstacle avoidance," introducing the corner detector named the interest operator (Moravec, 1981). In 1988, a modified version of the interest operator was introduced by Harris and Stephen, and it was called Harris corner detector (Harris and Stephens, 1988). Harris corner is robust to the rotation and small intensity changes but cannot handle the changes in scale.

Schmid and Mohr used the local invariant features to find matches between an image and a database of images for image recognition (Schmid and Mohr, 1997). They introduced rotation local invariant descriptors with Harris detector. So, their algorithm handled the rotation changes between image pairs, but could not handle the difference in scale.

Later in 2004, David Lowe introduced the Scale Invariant Feature Transform (SIFT): a detector and a descriptor that are rotation and scale invariant (Lowe, 2004) that. SIFT provides distinctive features that are invariant under both scale and rotation. Other scale-invariant features include the work of Mikolajczyk and Schmid (Mikolajczyk and Schmid, 2001). The

* Corresponding author

authors used the determinant of the Hessian matrix to select the location of a keypoint and used the Laplacian of Gaussian (LoG) to determine its scale. However; SIFT approximates LoG by the Difference of Gaussian (DoG) to speed up the process with no extra cost. At the time when SIFT was introduced, it was better than the other methods in terms of speed, repeatability, and stability. However; SIFT requires a 128-dimensional descriptor vector for each feature; which is memory and time consuming, especially when matching via Euclidean distance. Ke and Sukthankar (Yan Ke and Sukthankar, 2004) offered the PCA-SIFT to deal with the large number of dimensions of SIFT. The vector dimensions decreased from 128 to 36.

A move towards more efficient features was made by Bay et al. (Bay et al. , 2008). The authors proposed a detector and descriptor coined SURF. They exploited the integral images and approximated DoG by a box filter to detect the keypoints. That reduced the time as the same Gaussian filter was used without scaling the image. Haar wavelet responses in horizontal and vertical directions are computed to provide a 64-dimensional descriptor vector, that is half the size of the SIFT descriptor.

In 2010 a new keypoint descriptor: Binary Robust Independent Elementary Features (BRIEF) was introduced by Calonder et al. (Calonder et al. , 2010), to address the problem of the descriptor's size. BRIEF used only 256 bits or even 128 bits to store the information about the keypoints. That was not only memory efficient, but also speeded up the matching procedure. However; these descriptors failed with rotated images. In 2011, Leutenegger et al. tackled the problem of variation of scale and rotation by introducing the Binary Robust Invariant Scalable Keypoints (BRISK) (Leutenegger et al. , 2011), a detector that has a degree of modularity allowing it to be combined with other descriptors.

Another contribution was made by Rublee et al. (Rublee et al., 2011) to handle the scale and roatation of images and preserve the time and memory efficiency of BRIEF. Oriented FAST and Rotated BRIEF (ORB) was introduced by the authors as an alternative to FAST and BRIEF. Their method was based on combining the time efficient detector FAST (Rosten et al., 2010) and BRIEF as a binary descriptor. But after modifying them to cope with rotation and scale changes.

To filter the correct matches, Zhang et al. utilised the epipolar geometry as a constraint (Zhang et al., 1995). The authors' approach was to estimate the fundamental matrix and reject the outliers at the same time. Torr et al. proposed a similar approach. They used RANSAC to clear the outliers and estimate the fundamental matrix simultaneously (Torr and Murray, 1993). In the last two approaches, the model fitting and the outlier removal were performed after the preliminary matching. Moreover; a single model consensus was used as a filtering criterion.

Isack et al. [5] performed matching based on both the geometry and appearance of the features. However; the filtering is typically done after the preliminary matching. The authors' method starts with an initial set of matched SIFT features and uses it to refine the overall matching.

The proposed method starts with a set of matches, which are weak in terms of their number, accuracy and distribution over the scene structure (depths). From this set, a more extensive set of accurate and well-distributed matches is obtained. From the initial matches, the fundamental matrix and a global homography are estimated. It should be mentioned here that the global homography is not a good representation of the features correspondence. However, it will be used just as an approximation, and modifications will be performed later. After the estimation of these two entities, a detect-and-match strategy is employed, in which for every newly detected feature in one image (the left image) a corresponding window or region of interest (ROI) is determined in the other image (the right image). This ROI is predicted using the fundamental matrix (epipolar constraints) and the local homographies induced by the global homography and some discrepancy vectors. As the fundamental matrix can be used as a point-to-line mapping, and the homography can be used as a one-to-one mapping, but cannot be reliable; they can be used together as a complementary solution to limit the search space of the feature points in the right image. It is obvious that the prediction of the corresponding feature's location in the right image is just an approximation. However, this prediction procedure has a significant impact on limiting the ROI size.

After limiting the ROI size, template matching is employed to detect an accurate correspondence (correction). The template size can be chosen as a few pixels around the left image's feature point, and the source image is chosen as the ROI around the corresponding feature in the right image.

The global homography to be estimated from the initial set of matches is roughly representing the matches in this set. In fact, there is a discrepancy between the point correspondence obtained from the homography and the point correspondence in the initial matches for the same feature in the left image. This discrepancy is computed as a vector in homogeneous coordinates and then used for all neighbouring feature points as an alternative to the local homography. So, instead of limiting the method to work with a rough global homography, local homographies are implicitly used via the discrepancy vectors.

The method works recursively, such that more matches are added, and the geometric entities are recomputed.

3. METHODOLOGY

The proposed method has two phases; the first is the geometric prediction using epipolar constraints and local homography. The detailed description of this phase is shown in Figure 1. In this phase, we start with an initial set of matches, which we call the seed. These can be obtained either manually or using one of the common methods, such as SIFT, SURF, or ORB. In the tests performed in this paper, ORB was used to provide the set of initial matches, as ORB is known to provide less accurate and smaller number of matches compared to those of SIFT and SURF. Therefore, using ORB to generate the seed of matches implies that the proposed method can provide a more accurate set of matches even when weak matches were initially used. Additionally, the feature detection is done using the Features from Accelerated Segment Test (FAST) detector.

From the seed of matches, the homography and fundamental matrix are computed. The homography is then used with the initial matches to provide an alternative for a local homography that is called the discrepancy vectors. For each feature associated with a discrepancy vector, a kD-tree search runs to find neighbouring features that are assigned the same discrepancy vector. Using the general homography and the discrepancy vectors, feature correspondence can be estimated in the right image. To further enhance the estimation of the features' locations, the features are being projected on their corresponding epipolar lines.

The second phase is the correction phase via template matching. Figure 2 shows a description of the combined geometric prediction and the correction by template matching. A region of interest around the predicted feature position is assumed to be a search image for the template. Then, using the template matching, a correction is added to the initial prediction. Each newly predicted and corrected feature correspondence is added to the seed, and then the geometric entities are recomputed. The two phases are recursively repeated until all the detected features are either matched or rejected.

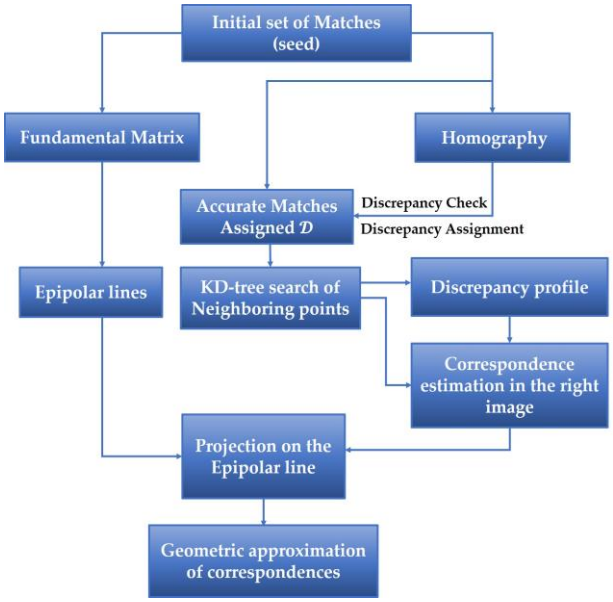


Figure 1. Geometric approximation (prediction) of the corresponding feature in the right image.

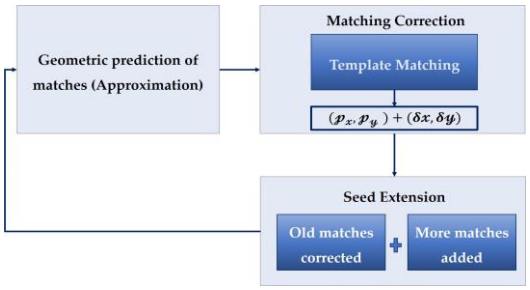


Figure 2. The recursive method: the geometric prediction of point correspondence and the correction by template matching

3.1 Correspondence prediction using geometric constraints

Epipolar geometry can be used to estimate approximate locations of the correspondences. The following notations are used: $\mathcal{I}_\ell$ and $\mathcal{I}_r$ are the left and right images respectively, $\boldsymbol{q}$ and $\boldsymbol{p}$ denote features in $\mathcal{I}_\ell$ and $\mathcal{I}_r$ respectively. Considering an image with no distortion, which is a very ideal situation, the epipolar constraint can be written as:

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{q} = 0 \qquad (1)$$

where $\mathcal{F}$ is the fundamental matrix.
Equation (1) states that the for a feature point $\boldsymbol{q}$ in the left image, there is a corresponding feature point $\boldsymbol{p}$ in the right image that lies on the corresponding epipolar line $\mathcal{F}\boldsymbol{q}$. That is, equation (1) defines a point-to-line mapping. A more general form of equation (1) is:

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{q} = \varepsilon \qquad (2)$$

where ε is a scaler denotes the errors (in pixels) caused by noise, lens distortion or other factors. Equation (2) states that the feature $\boldsymbol{p}$ lies approximately on the estimated epipolar line $\mathcal{F}\boldsymbol{q}$.
If the fundamental matrix is somehow known with reasonable accuracy, we could map any feature in one image to a line that is close enough to the corresponding feature in the other image.

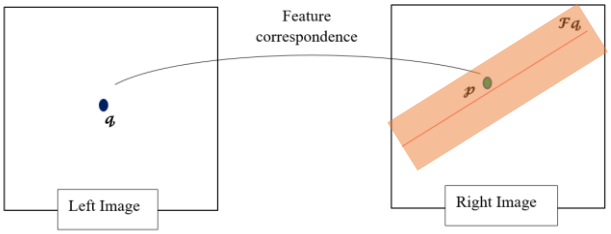


Figure 3. Epipolar constraints in the presence of errors. A feature point in the left image is mapped to the estimated epipolar line which lies close to the actual corresponding feature point in the right image.

Now, if a technique like Normalized Cross Correlation (NCC) is employed, the accurate location of the feature $\boldsymbol{p}$ can be found. But that requires sliding a window over a rectangular area around the epipolar line, and this could be an expensive procedure, especially when performed for several feature points.
To limit this rectangular area to a small ROI, we use an alternative to the local homography associated with the pair $(\boldsymbol{q}, \boldsymbol{p})$.
Suppose we have an initial set of matches, which we call the seed of matches. This seed can be used to estimate the fundamental matrix $\mathcal{F}$ and a global homography $\mathcal{H}$. The global homography does not help in finding correspondences, unless all the corresponding points are confined on a plane, since homography defines a plane-to-plane mapping. If we need to use this homography to map a feature point in the left image to its corresponding point in the right image, a discrepancy vector $\mathcal{D}$ must be added to compensate for the error caused by mapping using global homography instead of local homography. If we map with a local homography $\mathcal{h}$ that defines a plane-to-plane mapping, we can write:

$$\boldsymbol{p} = \mathcal{h}\, \boldsymbol{q} \qquad (3)$$

However; if we use the global homography $\mathcal{H}$, an error term must be added, as the mapping with $\mathcal{H}$ will shift the point $\boldsymbol{p}$ from the true correspondence position.

$$\boldsymbol{p} = \mathcal{H}\boldsymbol{q} + \mathcal{D} \qquad (4)$$

The vector $\mathcal{D}$ which we call the Discrepancy vector is a compensation for the local homography $\mathcal{h}$. From equation (3) and equation (4), we have:

$$\mathcal{h}\, \boldsymbol{q} = \mathcal{H}\boldsymbol{q} + \mathcal{D} \qquad (5)$$

Which means that mapping by the global homography differs by a shift, in x and y, from the more accurate mapping by a local homography. It is intuitive to think that each local homography in the pair of images is corresponding to a vector $\mathcal{D}$, if the global homography $\mathcal{H}$ is known. That is, there is a one-to-one correspondence between the local homography $\mathcal{h}$ and the discrepancy vector $\mathcal{D}$. In fact, to get the point correspondence for any point $\boldsymbol{q_i}$, it is only required to find the global homography $\mathcal{H}$ and a discrepancy vector $\mathcal{D}_i$, where the vector $\mathcal{D}_i$ is a unique vector for all the points on the local plane on which $\boldsymbol{q_i}$ is confined. That is, all the points in the neighbourhood of $\boldsymbol{q_i}$ and confined on the same plane as $\boldsymbol{q_i}$ are associated with a single discrepancy vector $\mathcal{D}_i$.
Now, suppose a new feature $\boldsymbol{q_j}$ is detected in the left image, if an associated discrepancy vector is known, then the position of the feature $\boldsymbol{p_j}$ can be approximately estimated from equation (4). So,

the problem is now reduced to finding a set of vectors $\mathcal{D}_i$, representing the local homography implicitly, and the global homography $\mathcal{H}$.
The procedure to assign the discrepancy vectors $\mathcal{D}$ to newly detected features q is as follows:

1. From the seed set of initial matches find the global homography $\mathcal{H}$.
2. Use the matched pairs in the seed to find a discrepancy vector $\mathcal{D}_i$ for each pair (q_i, p_i) from equation (4).
3. For each feature point q_i in the seed perform a kD-tree search to find a set of neighbouring features $\mathcal{N}_i$.
4. Assign the vector $\mathcal{D}_i$ to all the feature points $q_j \in \mathcal{N}_i$.

A note about the third point is mentioned here. Instead of finding $\mathcal{D}_i$ to newly detected points, it is less expensive to associate $\mathcal{D}_i$ with q_i and find more feature points in the neighbourhood $\mathcal{N}_i$.
Now, since newly added point are associated with some vectors $\mathcal{D}_i$ and since the global homography is estimated, one can use equation (4) to obtain the correspondence p_j for the feature point q_j. We shall denote the feature estimated from local homography by p_{jh}. It should be noted that the estimation of p_j is expected to be inaccurate. However; the estimated feature position is normally within a small distance from the accurate feature's position. Therefore, we can define a ROI as a small window around the estimated feature q_j.
To further enhance the accuracy of the estimated feature's position, the estimated feature is projected on the epipolar line $\ell_r = \mathcal{F}q$.
Let the epipolar line be expressed in the homogeneous coordinates as $\ell_r = [a_r, b_r, c_r]^T$, then equation (2) becomes:

$$a_r p_x + b_r p_y + c_r = \mathcal{E} \tag{6}$$

Where the feature point p **is** represented in homogeneous coordinates as: $p = [p_x, p_y, 1]$.
Suppose now we have estimated the feature p_{jh} using the local homography, which is represented as: $p_{jh} = [p_x^{jh}, p_y^{jh}, 1]$, then instead of relying on the estimation of the pair (p_x^{jh}, p_y^{jh}), we can use the epipolar constraint to project the point on the epipolar line. Using equation (6), we can project the estimated feature point on the epipolar line:

$$p_y^{j\mathcal{F}} = \frac{\mathcal{E} - a_r p_y^{jh} - c_r}{b_r} \tag{7}$$

where the term $p_y^{j\mathcal{F}}$ denotes the y's coordinate of the feature point p_j after the epipolar projection. The projected feature is $p_{j\mathcal{F}} = [p_y^{jh}, p_y^{j\mathcal{F}}, 1]$.
Figure 4 shows how local homography is used to find an estimated feature position p_{jh} corresponding to the feature point q_j. The estimated feature's position is different from the accurate one which is denoted p_j, the projected feature point is $p_{j\mathcal{F}}$.
The accurate feature p_j can be found by employing a template matching over a ROI around the estimated feature $p_{j\mathcal{F}}$.

3.2 Template matching using NCC

Template matching was defined in (Goshtasby, 1985) as the process of determining the position of a sub-image called the template $t(x, y)$ inside a larger image called the search image $s(x, y)$. Given the predicted position of the feature point p_{jh}, the search image is limited to the ROI around that point. The template slides over the search image, as in the example in Figure 5. Then a matching score is calculated. This score is maximum at the point of concurrence. The common score used in template matching are either the summation of absolute difference (SAD), the summation of squared difference (SSD) or the normalized cross-correlation (NCC). NCC is being used here as it is illumination invariant. NCC is defined as (Briechle and Hanebeck, 2001):

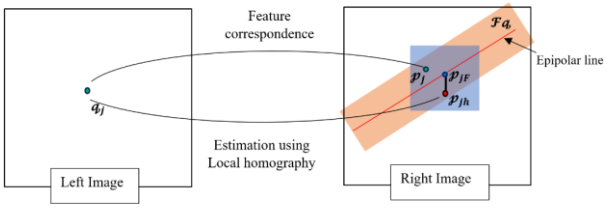


Figure 4: Using the local homography and epipolar projection to estimate the corresponding feature's location.

$$\gamma_{u,v} = \frac{\sum_{x,y}(s(x,y) - \bar{s}_{u,v})(t(x-u, y-v) - \bar{t})}{\sqrt{\sum_{x,y}(s(x,y) - s_{u,v})^2 \sum_{x,y}(t(x-u, y-v) - \bar{t})^2}} \tag{8}$$

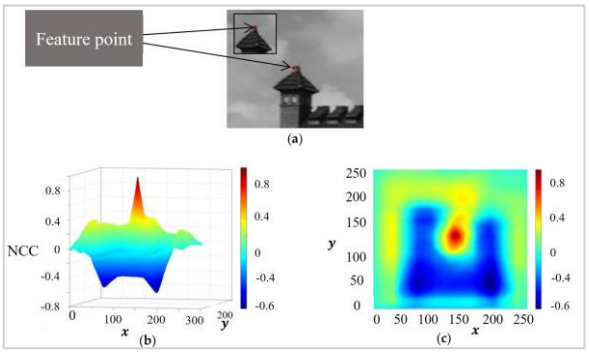


Figure 5. (a) the template window around the feature q_j slides over the search images (a larger image around the estimated feature $p_{j\mathcal{F}}$). (b) The NCC score is maximum at the concurrence point. (c) a planar view of the NCC score.

$\bar{s}_{u,v}$ and $\bar{t}$ denote the mean of s and t within the area of the template respectively.
There are some motivations for choosing template matching over descriptors. First, usually, NCC does not include mismatches, since the NCC is maximum at the exact point of concurrence if the score is less than a certain threshold (0.92 for example) the feature point can be excluded. Therefore, NCC is a more deterministic than the descriptors which tend to be more probabilistic. Second, the processing time of template matching is reduced since the search image is limited to the ROI around the predicted feature position. Matching time could be further reduced if image pyramids are exploited.
Fast NCC was proposed by J. P. Lewis (Lewis, 1995) to make template matching even faster. It was proposed to use recursive summation tables of the image function to find an approximation of the numerator and denominator of $\gamma_{u,v}$.
J.P. Lewis showed that the fast NCC could reduce the time of matching dramatically. Moreover, if fast NCC is considered over a small ROI, just like in the proposed method, it is expected to give satisfactory results in a short time.

3.2.1 Window size optimisation

The size of the template can be chosen as a few numbers of pixels around the detected feature $\boldsymbol{q_j}$, for example, a template of 9×9 or 15×15 could be suitable depending on the actual image size and resolution. However; the most important is the size of the search image, or the sub-image defined by the window around the predicted feature position $\boldsymbol{p_{jF}}$.

The size of the search image should be chosen such that we trade off the matching time efficiency with the matching accuracy. If the search sub-image's size is too small, the true corresponding feature might be located outside it. On the other hand, if the sub-image's size is too large, template matching will take much more time to work on the whole image. Therefore, an efficient approach is proposed here. First, note that the vector $\boldsymbol{\mathcal{D}_i}$ indicates the deviation of the feature $\boldsymbol{p_j}$ from the estimated value $\boldsymbol{p_{jh}}$. It also reflects the uncertainty region in which the estimated feature position is in the vicinity of the exact feature point $\boldsymbol{p_j}$.Therefore, the size of the sub-image can be chosen to be $(\mathcal{D}_x \times \mathcal{D}_y)$.

To further optimise the NCC time, instead of sliding the template over the whole source image, only small windows over the regions containing features are being used as search images. Figure 6 Shows that a search image may contain a large area of no features. Consequently, it is more efficient not to include the whole source image when performing NCC.

Figure 6. Instead of sliding the template over the whole ROI, it is more efficient to create smaller sub-windows around the possible detected features in the ROI.

4. RESULTS AND DISCUSSION

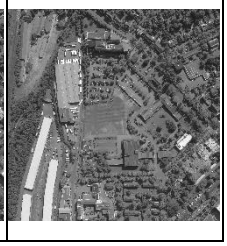
4.1 Dataset

Two datasets each were provided by the International Society of Photogrammetry and Remote Sensing (ISPRS) through the ISPRS and EuroSDR benchmark on multi-platform photogrammetry (Nex et al., 2015). The first and second pairs of images are for the Zollern Colliery (Industrial Museum) in Dortmund, Germany. These two pairs of images were taken by a UAV in a close-range manner. The third pair of images is taken over part of the city taken in a long-range manner.

4.2 Evaluation Criteria

The feature matching tests were performed on the two datasets, and the proposed methodology was compared with three state of the art methods, SIFT, SURF and ORB. All these methods were combined with RANSAC, while the model in the proposed method is estimated using Least Squares (LS). LS cannot be used efficiently with the other methods since a large number of outliers and the lack of geometric awareness associated with the feature points could lead to a highly inaccurate fundamental matrix model.

Table 1. Dataset from ISPRS and EuroSDR benchmark

Image pair properties	Left Image	Right Image
Close-range Successive images Multiple-depths Size: 3000×2000		
Close-range Non-Successive images Multiple-depths Size: 3000×2000		
Airborne Long-range size: 2500×2500		

To judge on the matching quality, two evaluation criteria are used. The first is the matching precision (MP), which was adopted by Chen et al. (Chen et al., 2017) and is defined as the percentage of the correct matches (NCM) to the number of matches (M). The second measure is the percentage of the correct matches to the number of detected features (PCMF): $PCMF = (NCM/F) \times 100\%$, where F is the number of detected features. For convenience, the number of detected features was set to be relatively the same for all the methods to neutralize the number of features when evaluating the matching process. To find the correct matches, a fundamental matrix is estimated from the data points and outliers are rejected if they exceed certain error threshold, based on equation (2).

4.3 Close-Range tests

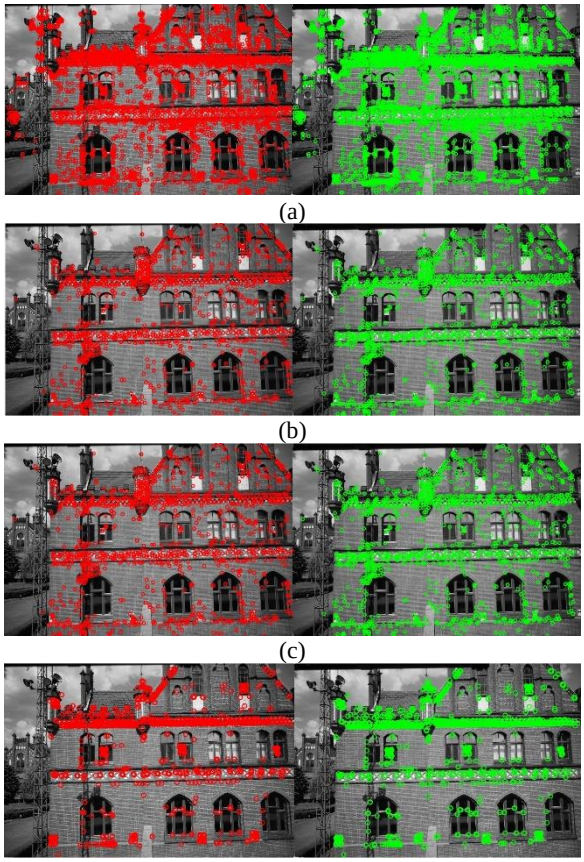
The first pair of images contains two images that are taken one after another. Therefore, this pair of images is characterised by a large overlap. However; this pair includes a large predominant surface. That is, a surface covering most of the overlapping area and containing many textures.

Despite the large overlap between the images, the computed fundamental matrix, based on RANSAC like methods, is most probably degenerate. The reason is that RANSAC tends to fit the model that best describes the distributed feature points. Since most of the points are on confined on the main building façade, RANSAC will reject other points in the background. This behaviour is obvious in Figure 7-b, 7-c, and 7-d, in which SIFT, SURF and ORB were combined with RANSAC.

On the other hand, the proposed method guarantees distribution of feature correspondence through different depth, as feature points can be seen not only on the building façade but also on the background. The visual results were also supported by the numerical results listed in Table 2.

It can be seen the proposed method outperforms the state of the art methods in terms of the number of correct matches, the matching precision, and the percentage of correct matches to the detected features. In fact, the matching precision of the proposed method is one of the motivations to use LS instead of RANSAC,

as the number of outliers is relatively small and can be rejected by LS.



(a) (b) (c) (d)

Figure 7. Feature correspondence for the first image pair. Results obtained using (a) the proposed method, (b) SIFT+RANSAC, (c) SURF+RANSAC, and (d) ORB+RANSAC.

Table 2. Matching measures for the first image pair when using SIFT, SURF, ORB and our proposed method.

Matching Measures	Matching Methods			
	SIFT	SURF	ORB	OURS
F	6559	6220	6500	6479
M	2906	3456	3039	4496
NCM	2340	2811	2162	4158
PCMF	35.68%	45.19%	33.26%	64.18%
MP	80.5%	81.34%	70.11%	92.48%

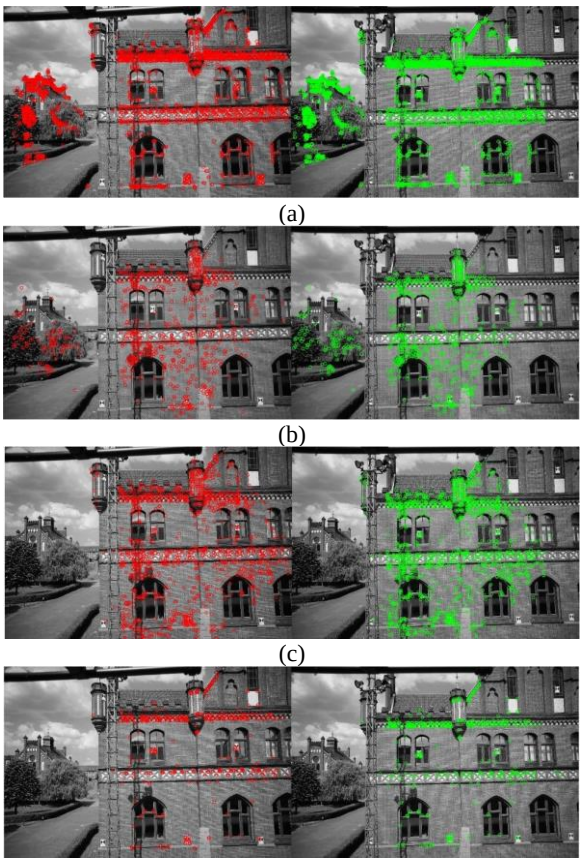
The second test is performed on the second image pair. This image pair is characterised by a relatively larger baseline, as the two images are not taken successively, but there are two images between them. The images contain a building façade with other buildings and trees in the background. Each object in the scene is characterised by a local homography.

Similar results to the first image pair are obtained with the second pair, as depicted visually in Figure 7. SURF and ORB, combined with RANSAC, failed to produce a well-distributed correspondence over the scene structure. The main cause of this issue is the weak distribution of the preliminary matched features, which is based on descriptor distance comparison. On the other hand, both SIFT and the proposed method could find a well-distributed set of matches. However; the proposed method has the highest matching precision and the highest correct matches to detected features ratio, as shown in Table 3.

Table 3. Matching measures for the second image pair when using SIFT, SURF, ORB and our proposed method.

Matching Measures	Matching Methods			
	SIFT	SURF	ORB	OURS
F	7280	7239	7300	7255
M	1921	2765	1472	3229
NCM	867	1303	666	3067
PCMF	11.91%	18.00%	9.12%	42.27%
MP	45.13%	47.12%	45.24%	94.98%

It should be mentioned here that well-distributed matches are the matches that exist over several depths or local homographies in the scene structure, and not confined on a single planar surface. The impact of the distribution of matches is on the 3D reconstruction is highlighted in subsection (4.5).



(a) (b) (c) (d)

Figure 8. Feature correspondence for the second image pair. Results obtained using (a) the proposed method, (b) SIFT+RANSAC, (c) SURF+RANSAC, and (d) ORB+RANSAC.

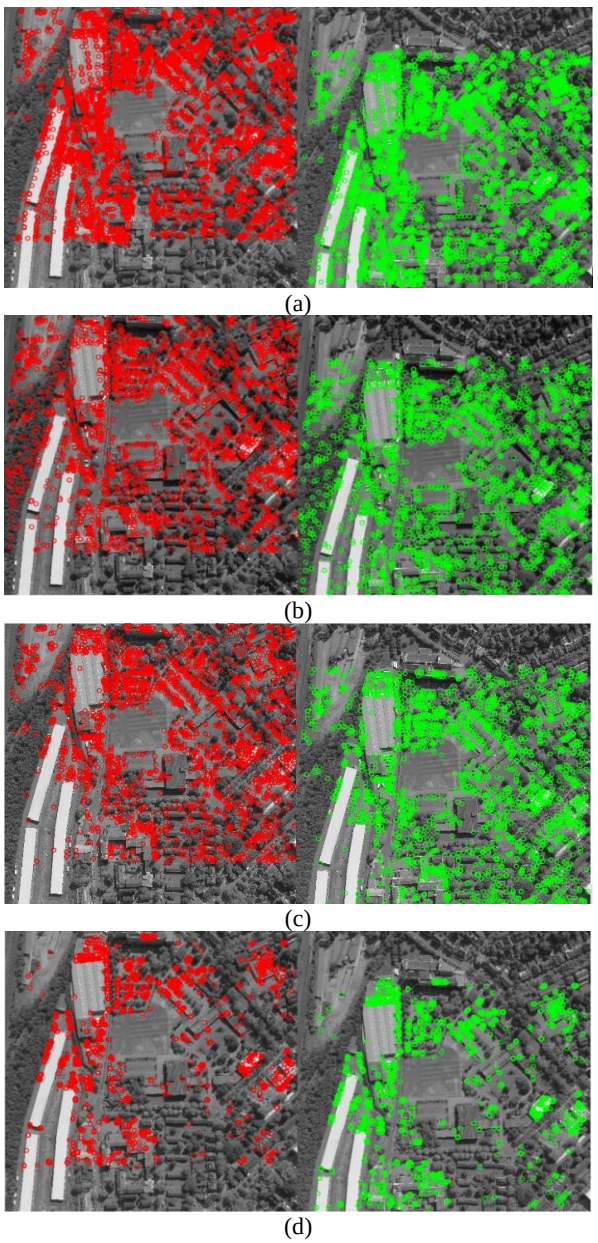
4.4 Airborne (Long-Range) test

The third image pair were taken by a UAV in Nadir view. The image pair contains several objects like building, trees and the ground. That is, several depths and variety of scene structure exist in the image. Therefore, matches features must cover all the depths in the scene to obtain the correct fundamental matrix and the correct rectification parameters. It can be seen that the matched features are distributed over the scene structure in

Figure 9-a and 9-b. Whereas, most of the matched features in Figure 9-c and 9-d are located on the ground. As expected these matches normally result in a flawed rectification.
The matching precision and ratio of correct matches to the detected features for this image pair are listed in Table 4. These numerical indicators emphasise the performance of the proposed method over the other methods.

Table 4. Matching measures for the third image pair when using SIFT, SURF, ORB and our proposed method.

Matching Measures	Matching Methods			
	SIFT	SURF	ORB	Ours
F	14201	14335	14500	14247
M	2792	4636	2466	5598
NCM	2592	2311	2415	5414
PCMF	18.25%	16.12%	16.66%	38.00%
MP	92.83%	49.85%	97.93%	96.71%

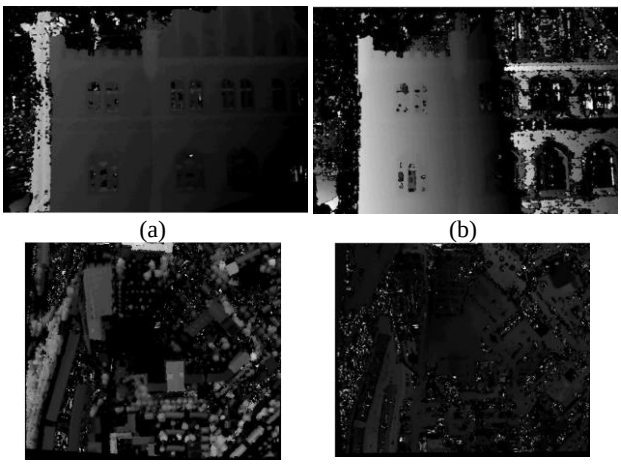


(a) (b) (c) (d)

Figure 9. Feature correspondence for the third image pair (airborne- Nadir view). Results obtained using (a) the proposed method, (b) SIFT+RANSAC, (c) SURF+RANSAC, and (d) ORB+RANSAC.

4.5 Applications to 3D reconstruction

As discussed earlier, the distribution of matched features over different depths (or local homographies) in the scene structure is a factor that has a significant impact on the quality of the 3D reconstruction. Typically, 3D reconstruction is done using dense image matching. The procedure of dense image matching starts with image rectification. Two images are first rectified using the fundamental matrix, and then matching over scanlines are performed. Therefore, errors in dense matching are either due to a flawed rectification or due to radiometric characteristics of the image pair. Consequently; to limit the error in dense matching to the radiometric properties only, a well rectification of the image pair is a must. Rectification parameters are a function of the fundamental matrix, which should be ideal (non-degenerate) for a perfect rectification. As was discussed in earlier subsections, accurate fundamental matrix estimation is dependent on the distribution of the matches over the scene structure.



(a) (b) (c) (d)

Figure 10. Disparity maps obtained with SGM after image rectification. (a) First image pair, rectification using the proposed method. (b) First image pair, rectification using the SURF+RANSAC. (c) Third image pair, rectification using the proposed method. (d) Third image pair, rectification using the SURF+RANSAC.

The impact of the distribution of matches over the scene structure is evident from Figure 10, which includes the disparity maps of the first and third image pairs that were computed using the semi-global dense matching (SGM) technique (Hirschmüller, 2008).
The disparity maps in Figure 10-a and 10-b are produced for the first image pair. Matches were obtained using the proposed method and SURF+RANSAC for the disparity maps in Figure 10-a and 10-b respectively. Similarly, for the third image pair, disparity maps were created after using the proposed method (Figure 10-c) and the SURF+RANSAC (Figure 10-d).
The disparity maps in Figure 10 highlight the performance of the proposed method, especially when dealing with images of multiple depths, or dominant planar surfaces. While the disparity map in Figure 10-a resembles the actual scene structure, except for some noise due to the difference in illumination, the disparity in Figure 10-b has much more noise due to the flawed rectification.
A similar situation occurs in the disparity maps in Figure 10-c and 10-d. In Figure 10-c, the disparity maps match the actual scene. While in Figure 10-d, the disparity map is a false one, in which the depths are reversed. For example, parts of the ground

in the actual scene are of higher disparity values in the disparity map than these of the buildings and trees, which is non-realistic. This behaviour is expected, as most of the matched features were only covering the ground, as depicted in Figure 9-b.

5. CONCLUSION

In this paper, a new method was proposed to find accurate, well-distributed feature correspondences based on an initial set of matches (the seed). The seed is used to estimate an initial fundamental matrix and an initial homography. Discrepancy vectors were introduced as an alternative to the local homographies. These geometric entities are then used to predict the positions of feature correspondences. To find the actual feature correspondence, a correction is performed using template matching with the normalised cross-correlation. The proposed method was tested on images taken by UAV in close range and airborne manner. The performance of the proposed method outperforms the state of the art methods in terms of the number of matches, the matching precision and the distribution of the matches over the scene structure. The distribution of matches over the scene structure is of significant impact on the model degeneracy and the 3D reconstruction. Therefore, the proposed method is a move towards the elimination of model degeneracy and flawless scene reconstruction. The method is mainly used with UAV images that are taken successively. This assumption limits the method to images of similar scale and orientation. Otherwise, scale and orientation estimation must be included in the method.

ACKNOWLEDGEMENTS

This work was supported by Dr. Naser El-Sheimy research funds from NSERC and Canada Research Chairs programs.

The authors would like to acknowledge the provision of the datasets by ISPRS and EuroSDR, released in conjunction with the ISPRS scientific initiative 2014 and 2015, led by ISPRS ICWG I/II

REFERENCES

Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* 2008, 110, 346–359.

Briechle, K.; Hanebeck, U.D. Template matching using fast normalized cross correlation. In *Proceedings SPIE 4387, Optical Pattern Recognition XII; SPIE*: Bellingham, WA, USA, 2001; pp. 95–102, doi:10.1117/12.42, 1129.

Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision*, Heraklion, Crete, Greece, 5–11 September 2010; Lecture Notes in Computer Science. Springer: Berlin, Germany, 2010; Volume 6314, pp. 778–792.

Chen, M.; Habib, A.; He, H.; Zhu, Q.; Zhang, W. Robust Feature Matching Method for SAR and Optical Images by Using Gaussian-Gamma-Shaped Bi-Windows-Based Descriptor and Geometric Constraint. *Remote Sens.* 2017, 9, 882.

Goshtasby, A. (1985). Template matching in rotated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (3), 338-344.

Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In *Proceedings of the Alvey Vision Conference*, Manchester, UK, 31 August–1 September 1988; pp. 147–152.

Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* 2008, 30, 328–341.

Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004*, Washington, DC, USA, 27 June–2 July 2004; Volume 2, pp. 506–513.

Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary Robust invariant scalable keypoints. In *Proceedings of the IEEE International Conference on Computer Vision*, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.

Lewis, J.P. Fast Template Matching. *Pattern Recognit.* 1995, 10, 120–123.

Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* 2004, 60, 91–110.

Mikolajczyk, K.; Schmid, C. Indexing based on scale invariant interest points. In *Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001*, Vancouver, BC, Canada, 7–14 July 2001; Volume 1, pp. 525–531.

Moravec, H.P. Rover Visual Obstacle Avoidance. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, BC, Canada, 24–28 August 1981; Volume 2, pp. 785–790.

Nex, F.; Gerke, M.; Remondino, F.; Przybilla, H.-J.; Bäumker, M.; Zurhorst, A. ISPRS Benchmark for Multi-Platform Photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 2015, II-3/W4, 135–142.

Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2010, 32, 105–119.

Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision*, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.

Schmid, C.; Mohr, R. Local Greyvalue Invariants for Image Retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* 1997, 19, 530–535.

Torr, P.H.S.; Murray, D.W. Outlier detection and motion segmentation. In *Sensor Fusion VI, SPIE 2059; SPIE*: Bellingham, WA, USA, 1993; pp. 432–443.

Torr, P.H.S.; Zisserman, A. MLESAC: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.* 2000, 78, 138–156.

Zhang, Z.; Deriche, R.; Faugeras, O.; Luong, Q.T. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.* 1995, 78, 87–119.