

A STEREO LINE MATCHING TECHNIQUE FOR AERIAL IMAGES BASED ON A PAIR-WISE RELATION APPROACH

A. O. Ok ^a*, J. D. Wegner ^b, C. Heipke ^b, F. Rottensteiner ^b, U. Soergel ^b, V. Toprak ^a

^a Dept. of Geodetic and Geographic Information Tech., Middle East Technical University, 06531 Ankara, Turkey
- (oozgun, toprak)@metu.edu.tr

^b Institute of Photogrammetry and Geoinformation, University of Hannover, 30167 Hannover, Germany
- (wegner, heipke, rottensteiner, soergel)@ipi.uni-hannover.de

Commission I, WG I/4

KEY WORDS: pair-wise line matching, post-processing, line extraction, stereo aerial images

ABSTRACT:

In this study, we developed a new pair-wise relation based approach for the matching of line features from stereo aerial images. To solve the final matching inconsistencies, we propose an iterative pair based post-processing algorithm in which the matching inconsistencies are eliminated using three novel measures and a final similarity voting scheme. The approach is tested over four urban test sites with various built-up characteristics, and for all test sites, we achieved a stereo line matching performance of 98%. The overall results indicate that the proposed approach is highly robust for the line features extracted in (very) dense urban areas.

1. INTRODUCTION

Corresponding lines in overlapping aerial images can be used for different purposes such as 3D object extraction, improving the automated triangulation, image registration, motion analysis etc. However, line matching in ultra high resolution (6–8 cm) stereo aerial images is a very challenging task due to various reasons; substantial change in viewpoints, inconsistency of line endpoint locations, the limitations of the geometric constraints imposed, lack of rich textures in line local neighbourhood, repetitive patterns etc. Up to now, a significant number of research papers have been published in this field; however, in a stereo environment, the ambiguity problem of line matching is an issue that remain unsolved. The major problem in line matching arises from the lack of measure(s) and/or constraint(s) for line features that are invariant under different viewing conditions. Existing geometric attributes for line matching in the stereo geometry is strictly limited. For example, Zhang (2005) mentioned the major problems of the available geometric constraints for line features, and finally, utilized only the orientation of the line segments as a single geometric constraint. The information around the line local neighbourhood is also well issued by most of the researchers; for example, Schmid and Zisserman (1997) proposed a direct and warped correlation measures computed around the line neighbourhoods observed by short and long range motions, respectively. The color and chromatic information within the local neighbourhood was also well issued in elsewhere (Scholze et. al., 2000; Zhang and Baltsavias, 2000, Herbert et. al., 2005). Recently, Wang et. al. (2009) proposed a new measure that takes into account the information based on the gradient orientation around the line local neighbourhoods and presented some good results for a number of close range datasets. However, all those measures are almost non-discriminative by their own for the aerial image case and suffer from the same problem, repetitive patterns, where the information extracted from local neighbourhoods of very different lines has similar information (Fig. 1). Therefore, for example, the strength of the work of Schmid and Zisserman (1997) relies to the post-processing stage where the epipolar ordering constraint is forced over line features for the disambiguation. However, the ordering constraint has some critical drawbacks as well; (i) some certain lines especially belonging to thin objects (mostly details of buildings) will be unquestionable lost, (ii) the matching failures especially in

occluded areas that do not violate the ordering constraint could not be detected and eliminated. To be specific, almost all the previous work related to line matching relies on various descriptors specialized for one to one line matching in which the relations between the line features are not taken into account. However, the integration of those line relations during matching expose new constraints and further possibilities to improve the quality and the performance of the line matching.

2. METHODOLOGY

2.1 Line Extraction and Pair-wise Line Matching

In a recent work, we have presented a novel approach for the pair-wise matching of line features (Ok et. al., 2010). In this paper, we only briefly summarize the algorithm and refer the reader to the reference for further details. The algorithm consists of two main steps; (i) straight line extraction and (ii) the stereo matching of the extracted lines with a pair-wise approach. During the first step, in order to maximize the performance of the line detection, existing multispectral information in aerial images was fully utilized throughout the steps of pre-processing and edge detection. To accurately describe the straight edge segments, a principal component analysis technique was adapted and the extracted segments were converted to their line counterparts using an iterative Ransac algorithm. To establish the pair-wise line correspondences between the stereo images,

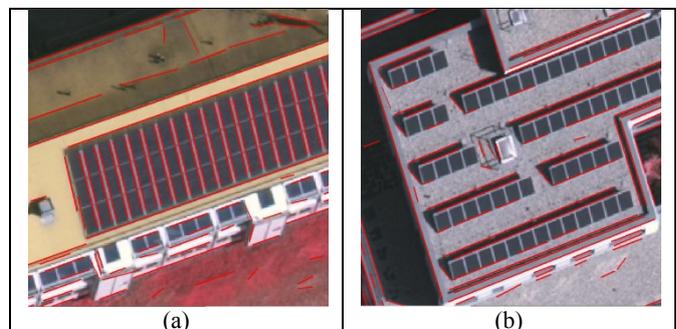


Figure 1. Two examples of line segments commonly observed in repetitive patterns and their very similar local neighbourhoods.

a new pair-wise stereo line matching approach was developed. For each pair in the left image, the best candidate line pair in the right image was assigned after a weighted pair-wise matching similarity score which was computed over a total of eight measures; an epipolar, three geometric, two photometric, a correlation and a spatiogram constraint. All the measures were normalized from 0 to 1 prior to the calculation, and the total similarity result was computed as the average of all similarities.

The main problem of the pair-wise matching presented is that it does not always guarantee one to one matches for each line. Based on our experiences, after the pair-wise matching, the ambiguities mostly occur for the lines that are adjacently located within a very short perpendicular distance. A typical example is given in Fig. 2. This is mainly due to two explicit reasons; (i) the lines that are very close to each other that belong to the same object (building, road etc.) reveal similar pair-wise characteristics and (ii) since we apply relaxed thresholds during pair-wise matching (especially for the epipolar intersection), very close lines are mostly susceptible to satisfy those thresholds. Therefore, in this paper, a great care has been devoted to the post-processing stage and a new iterative disambiguation algorithm is developed. For this purpose, we combined three novel measures during the selection of the best line correspondences. (i) The first measure relies on the gradient orientation information in the local neighbourhood of lines which is computed using a recently proposed dense matching measure, Daisy (Tola et al., 2010). Since the original Daisy measure is point based, in this study, the measure is extended and adapted to fulfil the requirements of the linear features and their local neighbourhood. (ii) The second measure, the Redundancy, is computed from the entire pair-wise matches based on the fact that a single line is allowed to have a part in different pair combinations. Thus, after the pair-wise matching, there is a quite large number of matching redundancy available for most of the line correspondences. By this way, the redundancy measure gives a possibility to understand and integrate a local matching support for lines during the disambiguation process. (iii) The third measure is computed from the results of each individual pair-wise matching. Since

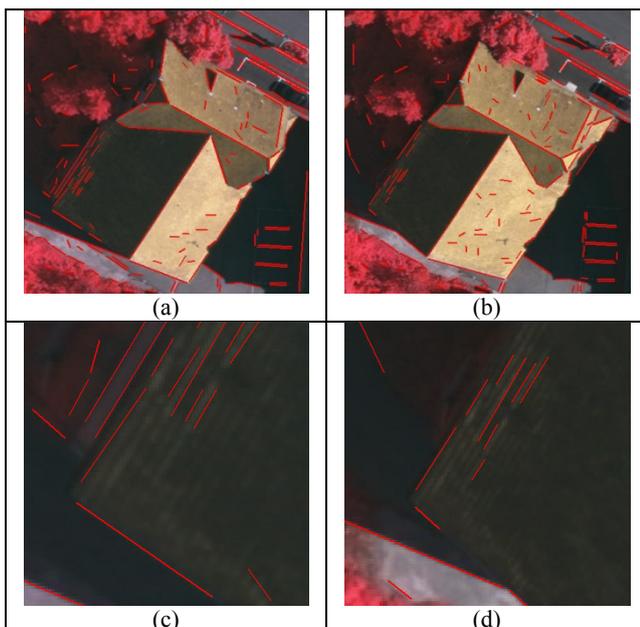


Figure 2. The matching ambiguities after pair-wise matching. (a, b) The stereo images and the extracted lines, (c, d) potential ambiguities after the pair-wise matching.

we assigned the best pair using a pair-wise matching similarity score, this information can also be utilized during the post-processing, since the quality of the pair matches inherently determined by the quality of the line correspondences in each pair. We integrated those three measures for the final disambiguation process in an exclusively developed iterative way, in which the matching inconsistencies are eliminated using nearest/next ratios and a final similarity voting scheme.

2.2 Measures Utilized During Post-Processing

2.2.1 Daisy Measure

In recent years, the gradient orientation histograms has proven to be robust to distortions (up to a level) and found to be successful in terms of point matching when compared to the classical pixel-based measures such as cross-correlation and pixel differencing. Some good examples can be found in (Lowe, 2004; Mikołajczyk and Schmid, 2005; Bay et al., 2006). More recently, Tola et al. (2010) proposed a dense matcher, Daisy, which is also proven to be much more efficient during the computation of the gradient orientation histograms. In the line matching context, up to our knowledge, the only study that takes into account the gradient orientation around the line local neighbourhood was proposed by Wang et al. (2009). However, as we already mentioned in the first section, for aerial images, the final decisions of the line matching that are only based on the information obtained from the line local neighbourhoods could be ambiguous. However, the information within those neighbourhoods may reveal some hints and may provide opportunities to indicate and eliminate the indisputably wrong matches.

In this study, we selected the Daisy as a fundamental local neighbourhood measure for the post-processing due to two explicit reasons; (i) great efficiency and speed during the computation of the gradient orientation histograms, (ii) its circular, symmetric shape and isotropic kernel structure turns out in a small overhead during the computation of the measure for different line orientations. Here, first, we only briefly review the original point-based Daisy measure and refer the reader to the reference for further details. Thereafter, we will introduce new adaptations for the Daisy and present how efficiently the measure could be utilized for capturing the line local neighbourhoods.

The Daisy descriptor is given in Fig. 3 (Tola et al., 2010). In the descriptor, each circle represents a region where the radius is proportional to the standard deviations of the Gaussian kernels and the “+” sign represents the pixel locations of the convolved orientation map centers where the descriptor is computed. Daisy is controlled by a total of 4 parameters; where R is the distance from the center pixel to the outer most grid point, Q is the number of convolved orientation levels, T is the number of histograms at a single layer, and H is the number of bins in the histogram. For a given input image, first, depending on the number of bins H , orientation maps are computed. Each orientation map is then incrementally convolved with Gaussian kernels of different sigma values to obtain convolved orientation maps. At each pixel location illustrated in Fig. 3, a vector made of values from the convolved orientation maps are computed. Let $\mathbf{h}_{\Sigma}(u,v)$ represent the vector made of the values at location (u,v) in the orientation maps after convolution by a Gaussian kernel of standard deviation Σ , and let Q represents the number of different circular layers, then the Daisy descriptor $D(u_0,v_0)$ for location (u_0,v_0) is defined as:

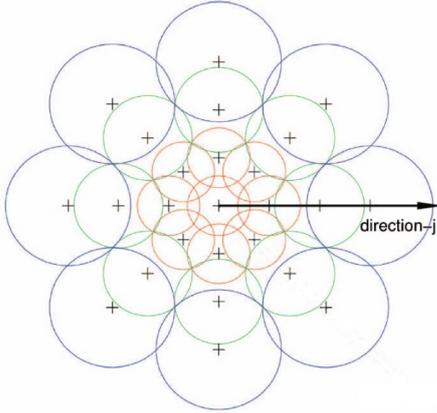


Figure 3. The Daisy descriptor

$$\mathcal{D}(u_0, v_0) = \begin{bmatrix} \tilde{\mathbf{h}}_{\Sigma_1}^\top(u_0, v_0), \\ \tilde{\mathbf{h}}_{\Sigma_1}^\top(\mathbf{l}_1(u_0, v_0, R_1)), \dots, \tilde{\mathbf{h}}_{\Sigma_1}^\top(\mathbf{l}_T(u_0, v_0, R_1)), \\ \tilde{\mathbf{h}}_{\Sigma_2}^\top(\mathbf{l}_1(u_0, v_0, R_2)), \dots, \tilde{\mathbf{h}}_{\Sigma_2}^\top(\mathbf{l}_T(u_0, v_0, R_2)), \\ \dots \\ \tilde{\mathbf{h}}_{\Sigma_Q}^\top(\mathbf{l}_1(u_0, v_0, R_Q)), \dots, \tilde{\mathbf{h}}_{\Sigma_Q}^\top(\mathbf{l}_T(u_0, v_0, R_Q)) \end{bmatrix}^\top,$$

where $\mathbf{l}_j(u, v, R)$ is the location with distance R from (u, v) in the direction given by j when the directions are quantized into the T values (Tola et. al., 2010).

Since the Daisy descriptor is point-based (it belongs to the center grid point), in this study, the measure is extended and adapted to fulfil the requirements of the line features and their local neighbourhood. First, we centralize the center grid point of the descriptor to the center of the overlapping parts of the line segments which are defined by point to point correspondence (Ok et. al., 2010). Next, to achieve rotation invariance over gradient vectors, we rotate the Daisy grid and align the direction vector (Fig. 3) of the descriptor with the orientation of each line. Since the amount of rotation must be adjusted for all lines based on their angle values in image space, during this procedure, we fully utilize one of the main advantages of the Daisy in which we only circularly shift the final orientation histograms to compute the descriptor. To achieve invariance to perspective distortion exactly on the line segments, for each line, we utilize adaptive R values for the Daisy grid (distance from the center pixel to the outer most grid point). The original Daisy measure has a specific constant R value; however, adaptive R values for line segments could be computed with the knowledge of the overlapping parts after imposing point to point correspondence. Since we apply this correspondence during the initial pair-wise matching, it does not bring any further overhead during the computation of the measure. In addition, it is apparent that we don't have any knowledge about the surfaces attached to the lines in their neighbourhoods; thus, we further utilize the adaptively computed R values for entire Daisy grid points. After these adaptations, for the computation of the similarities, we divide the Daisy grid points into two separate classes and produce two constant grid binary masks $\{M_m(x)\}$ for each line; the grid points that are located (i) above the line, and (ii) below the line. Thus, we perform the similarity computations independently for each grid class. Moreover, we also mask out the vector made values, $\mathbf{h}_\Sigma(u, v)$, from the descriptor matrix $\mathcal{D}(u_0, v_0)$ whose grid locations are exactly on the line. This is due

to the reason that if one of the sides of the lines is occluded, then the histograms computed for the points that are exactly on the lines have no reason to resemble each other. Therefore, we exclude those pixel locations and their histograms from the Daisy measure. For the computation of the dissimilarities between two Daisy descriptors, Tola et. al. (2010) proposed a Euclidean difference metric;

$$D = \frac{1}{\sum_{q=1}^S M^{[q]}} \sum_{k=1}^S M^{[k]} \left\| D_i^{[k]}(x) - D_j^{[k]}(x) \right\|_2, \quad (2)$$

where S is the number of grid points, $M^{[k]}$ is the k th element of the binary mask M , and $D_i^{[k]}$ is the k th histogram \mathbf{h} in $\mathcal{D}(x)$ computed image i . However, we observed that, although the metric is successful in most of the cases, it completely ignores the cross-correlation between the two descriptors, D_i and D_j . Thus, we define a modified-similarity (M_S) metric that can be jointly utilized with the cross-correlation:

$$M_S = \frac{1}{1 + \left(\sum_{k=1}^S M^{[k]} \left\| D_i^{[k]}(x) - D_j^{[k]}(x) \right\|_2 \right)^2} \quad (3)$$

First, the normalization coefficient in Eq. 2 is not necessary any longer since our binary masks have constant number of points for each side of the lines. After the modification, the similarity metric produces values between 0 and 1, and in order to be more discriminative, we take the square of the total dissimilarity, thus, we further penalize the higher dissimilarities values ($D > 1$) and give more weight on the lower ones. We also define the cross-correlation-similarity (C_S) between two descriptors as:

$$C_S = \left(\frac{\sum_{k=1}^S (M^{[k]} D_i^{[k]}(x) - \mu_{(MD_i)}) (M^{[k]} D_j^{[k]}(x) - \mu_{(MD_j)})}{s(MD_i) s(MD_j)} \right)^2 \quad (4)$$

where $\mu(\cdot)$ and $s(\cdot)$ operators denote the mean and standard deviations, respectively. Note that, in Eq. 4, similar to in Eq. 3, the correlation is also squared in order to give more weight to high similarity values, and to be even more discriminative. Finally, since the similarities in Eq. 3 and 4 are computed independently for both sides of lines (for the above and below grid points); we propose our final Daisy similarity (Sim_D) for line matching as:

$$Sim_D = \left((M_S^{Above}, M_S^{Below})^+, (C_S^{Above}, C_S^{Below})^+ \right)^- \quad (5)$$

where $(\cdot)^+$ and $(\cdot)^-$ denote $\max(a, b)$ and $\min(a, b)$, respectively. The final $(\cdot)^-$ operator in Eq. 5 ensures that the final Daisy similarity metric (Sim_D) should be high for both of the similarity metrics, M_S and C_S .

2.2.2 Redundancy Measure

In this paper, to solve the matching ambiguities, we propose a new measure, Redundancy, which is computed from the entire pair-wise matches based on the fact that a single line is allowed to have a part in different pair combinations. Thus, after the pair-wise matching, for each line, there is a number of matching redundancy that could be efficiently utilized for the disambiguation. The idea is illustrated in Fig. 5. Assume that Fig. 5 shows the line segments extracted from two stereo images, and Table 1a gives the results of the pair-wise matches for those line segments that only had a pair-wise relation with

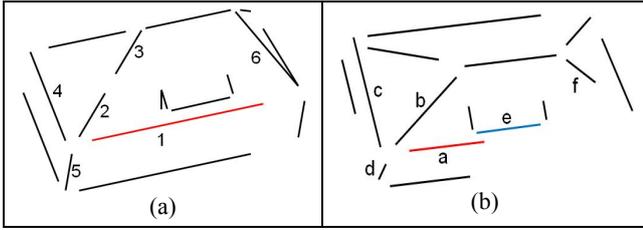


Figure 5. Line segments extracted from two stereo images.

Pair-wise Matches		Line Matches	
Left	Right	Left	Right
1 - 2	a - b	1	a - e
1 - 3	a - b	2 - 3	b
1 - 4	a - c	4	c
1 - 5	e - d	5	d
1 - 6	a - f	6	f

Table 1. The results of the (a) pair-wise matching and (b) inferred line matches from the pair relations.

segment #1. In Table 1a, the left column represents the search pairs generated from the left image, and the right column represents the best pair matches assigned after the pair-wise relational matching. If we look at the results of the pair-wise matches in detail, the segment #1 had a total of five pair-wise relations with other line segments within the pre-defined proximity (Ok et. al., 2010) in image space. One to one line matches inferred from the pair-wise relations are given in Table 1b. Based on the uniqueness constraint, a single line segment from the left image has, at most, one corresponding line segment in the right image (Suveg and Vosselman, 2004). However, during the extraction of line segments, a single segment is often fragmented into several shorter segments, such as in the case for segments #2 and #3. In this case, the uniqueness constraint must be handled carefully by taking into account the colinearity of the fragmented ones (segments #2 and #3). It is also obvious from the Table 1b is that we have two candidate line matches over the segment #1, however, visually, it is clear that, segment #1 in the left image corresponds to the segment #a in the right image. To solve the ambiguity, we evaluate the redundancy within the pair-wise matches (Table 1a). To be specific, we search the total number of occurrences of each one to one relation within the entire pair-wise relations. This gives us quite powerful unique information, since most of the ambiguities occur due to accidental alignments and has a very limited chance to occur in multiple times. For example, in Table 1a, among the total of five pair-wise relations that involve segment #1, four out of five corresponds to the segment #a in the right image. Only, one pair-wise relation indicates the correspondence with segment #1 and #e for the left and right images, respectively. Thus, the number of occurrences computed from the pair-wise relations over segment #1 reveals the segment #a as the correct match.

In (Ok et. al., 2010), we applied this redundancy measure with a single threshold, and for most of the cases, the ambiguities were successfully solved and provided good results. However, due to the perspective distortion and the relaxed thresholds applied during pair-wise matching, in some cases, the number of pair-wise occurrences of wrong candidate may exceed the number of occurrence of the correct match. One good example for this problem is illustrated in Fig. 6. In the figure, the left column (Fig. 6a-c) and right column (Fig. 6b-d) belongs to the left and right stereo images, respectively. It is shown that, after the pair-wise matching, in the left column, the line represented with red

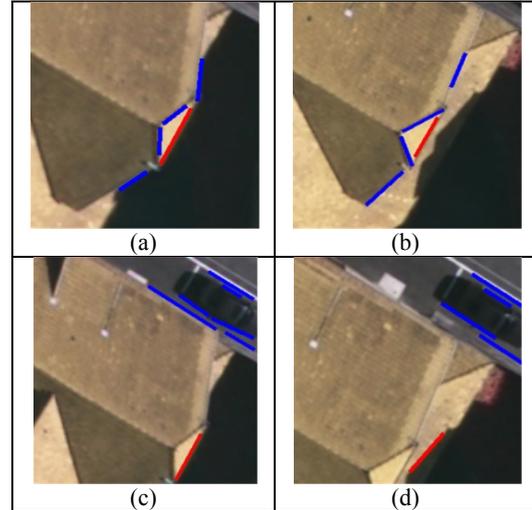


Figure 6. A problematic case of the Redundancy measure.

have two candidates for matching (two different red lines in right column). The blue lines demonstrate the lines that assist the pair relations for both candidates. As expected, the correct match (Fig. 6a-b) had successfully paired with a total of four lines that belong to the surrounding boundaries of the building roof. However, surprisingly, the wrong candidate (Fig. 6c-d) paired with a total of six lines (some of them are multiple matches) extracted from the boundaries of a car parked on the nearby street. Thus, for this example, blindly counting the number of occurrences may lead the redundancy measure to a wrong match (red lines in Fig. 6c-d). Therefore, we weight all pair relations proportional to their within pair minimum distances. By this way, the redundancy measure provides a possibility to understand and integrate a local matching support for lines. It is clear from Fig. 6a-b that the minimum distances between the lines in pair relations that belong to the correct match are much shorter than the ones that belong to the wrong match (Fig. 6c-d). Thus, we propose the new redundancy measure (Sim_R) for a line pair as:

$$Sim_R = \sum_{q=1}^N \frac{1}{\sqrt{(d_{ij}^{[L]}, d_{ij}^{[R]})}} \quad (6)$$

provided that the $d_{ij}^{[L]} \neq 0$ and $d_{ij}^{[R]} \neq 0$. In Eq. 6, N is the number of pair relations assist to matching, d_{ij} is the pixel-based minimum 2D Euclidean distance between two lines (l_i and l_j) in a pair, L and R indicates the pair relations in left and right images, respectively.

2.2.3 Pair-wise Quality Measure

During pair-wise matching, the final pair matches are assigned after a weighted pair-wise matching similarity score which is computed over a total of eight measures; an epipolar, three geometric, two photometric, a correlation and a spatiogram measure (Ok et. al., 2010). All the measures are normalized from 0 to 1 prior to the calculation, and the total similarity (Θ_T) result is computed as the average of all similarities. The final pair-wise similarity value (between 0 and 1) for each pair that is computed from those eight measures may give us a hint about the quality of the line matches in that pair. Thus, if a line match is a part of N number of pairs, the pair-wise quality metric (Sim_Q) for that line match is computed as the average of all pair-wise similarities:

$$Sim_Q = \frac{1}{N} \sum_{q=1}^N \theta_T^{[q]} \quad (7)$$

where $\theta_T^{[q]}$ is total pair-wise similarity of the q th pair relation.

2.3 Post-Processing

In this paper, we propose a new post-processing algorithm that mainly relies on the pair-wise matches. We integrated the measures explained above for the final disambiguation process in an exclusively developed iterative pair-wise manner, in which the matching inconsistencies are eliminated using nearest/next distance ratios (NNDR) and a final similarity voting scheme.

Our aim during post-processing is to first eliminate indisputably wrong relations based on a very strict NNDR values forced jointly over the measures Daisy (Sim_D) and Redundancy (Sim_R). NNDR is first introduced by Lowe (2004) based on the fact that the correct matches need to have the closest matching similarity significantly closer than the closest incorrect match to achieve reliable matching. For false matches, there will likely be a number of other false matches within comparable matching similarities. For a large number of datasets, we investigated the NNDR metric in terms of the ratio of closest to second-closest matches of each line, and for line matching problem, we reject all the related matches of a match that has a Daisy dissimilarity ($1 - Sim_D$) ratio lower than $ratio_D$. The threshold is selected in a way that only a very limited number of line matches that have enough confidence has possibility to fulfil this threshold ($ratio_D = 0.1$). Moreover, on the contrary to the other studies that rely on a single measure, we also jointly force the redundancy NNDR metric during this process, thus, we also restrict the elimination of matches which has a redundancy distance ratio lower than $ratio_R$ ($= 0.35$). Thus, by means of this joint restriction, we eliminate all the indisputably wrong relations beforehand without removing any of the correct matches. Thereafter, we delete the line relations indicated by NNDR from the pair-wise matches, and for each match, we update the redundancy and quality metrics. At this point, it should be pointed out that, if a line relation in a pair is found to be wrong, we do not directly delete the pair, since we don't have any inference (correct or wrong) for the other match in the pair. The example given in Table 1a clarifies this fact. For the fourth pair relation (1-5, e-d), assume that we found that the line match (1-e) is wrong. However, we do not have any information about the other match (5-d) in the pair; thus, we cannot directly label the other match as wrong (although it may be in some cases). Therefore, since one of the matches in a pair is labelled as wrong, we update the redundancy (Sim_R) and quality measures (Sim_Q) of the other match in that pair by eliminating the contribution of that pair from its similarity values. By this way, for example, the match (5-d) given in the Table 1a is not directly eliminated, but penalized, due to reason that the relation (1-e) in the pair is labelled as wrong.

Once all the measures are updated, we initiate an iterative matching scheme by starting from the match that has the highest redundancy measure. Subsequently, we select all the potential matching candidates (ambiguities) for that match. Thereafter, for those matches, we compute an overall similarity metric by taking the weighted linear combination of the similarity measures:

$$Sim_T = w_D \cdot Sim_D + w_R \cdot \frac{Sim_R}{(Sim_R)^+} + w_Q \cdot Sim_Q \quad (8)$$

In Eq. 8, for each selected match, we normalize the Redundancy measure (between 0 and 1) with the maximum Redundancy value of the selected matches, so that the contribution of all similarities is consistent for the final voting. Based on our experiments, we found that the redundancy is the most reliable and unique measure among the three measures, thus, in this study, weights of the similarities in Eq. 8 are designed as $\{w_D, w_R, w_Q\} = \{1/4, 1/2, 1/4\}$.

Apparently, among the selected matches, the correct match is the one that maximizes the overall similarity metric (Sim_T). Thereafter, we fix the correct match and check for the matching ambiguities that violate the selected match. At this point, the colinearity of the line segments of the matching violations (if there any) are individually tested with the line segments of the correct match in order to avoid the deletion of the fragmented lines. The ones that are found to be collinear are labelled along with the correct match for the final matching list. The ones that are not collinear are deleted from the pair-wise matches. After the deletion, we apply the same updating strategy as we explained above. Thus, at the end of each iteration, we penalize all related matches in the pairs that are labelled as wrong. Thus, the (updated) measures (Sim_R and Sim_Q) turn out to be more and more reliable after each iteration.

Finally, the iterations stop after there is no ambiguity exists in the final matching list. Like any other system developed so far, when a line segment in the first image has no corresponding line segment in the second image, the system cannot identify the wrong match (if accidentally assigned) since the correct line to be matched is missing. To solve this problem, a final check with a global threshold is required. On the contrary to the previous studies that rely on a single threshold, we propose a new hysteresis like global thresholding to solve the problem and to maximize the performance of the matching. As we penalize the Redundancy measure (Sim_R) for each match after each iteration, once the iterations has stopped, we have a near-perfect final (Sim_R) values for the final matching list. This gives us a unique way to solve the above mentioned problem, in principle; those ill-posed matches have very low Redundancy values when compared to values of the correct matches. Thus, we define a two-level global thresholding:

- (i). $Sim_D \geq Thr_D^1$
- (ii). $Sim_D \geq Thr_D^2 \quad \& \quad Sim_R \geq Thr_R$

From experiments, we have found that a global Daisy threshold of $Thr_D^1 \geq 0.2$ must be independently satisfied by every match. However, due to lack of rich textures in line local neighbourhood, some false matches may easily exceed this threshold. Increasing the threshold may have a possibility to eliminate some of the correct matches as well, thus results in reducing the overall completeness of the matching. So, we propose to utilize a second high Daisy threshold, $Thr_D^2 \geq 0.85$ restricted with a global Redundancy threshold of $Thr_R \geq 0.5$. By this way, compared to case where only a single global threshold is forced, using a two-level thresholding at the same time can eliminate most of the remaining false matches while keeping the matching precision and matching completeness.

3. RESULTS AND DISCUSSION

To test our methodology, we selected four urban test sites with various built-up characteristics (dense-sparse, flat-gable-complex roofs, repetitive patterns etc.) over a built up area of the city of Vaihingen-Germany. The stereo pairs were acquired by the DMC digital camera with 70% forward overlap (Cramer

and Haala, 2009). The focal length of the camera was 120 mm and the flying height was approximately 800 m above the ground level which corresponds to a final ground sampling distance (GSD) of approximately 8 cm.

For all test sites, we applied a 50 m (≈ 162 pixels) search range difference (between the min. and max. heights) along the epipolar lines. The number of correct and false line matches was assessed manually, and first and second columns of Fig. 7 show

the matched lines for the left and right stereo images, respectively. For all test sites, more than 55% of the extracted lines are matched, and of these matches 98% are correct. If the complexities of the test sites are taken into account, this seems to be a very good performance. Furthermore, on the contrary to the most of the previous approaches, we do not impose any external dataset to the matching (third view, DSM etc.), and do not perform any ill-posed constraint, such as epipolar ordering. It should also be emphasized that the curved segments

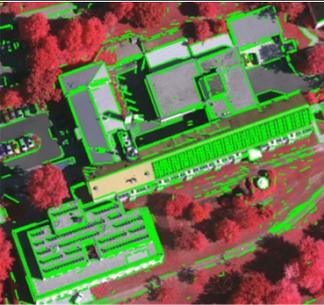
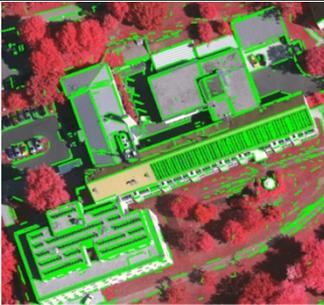
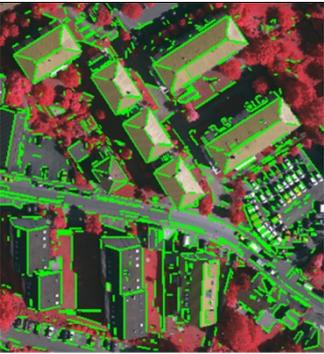
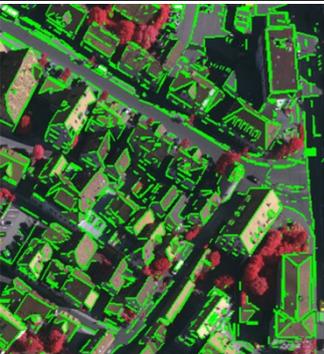
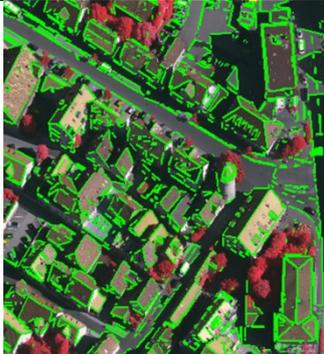
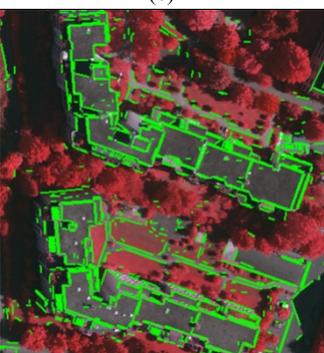
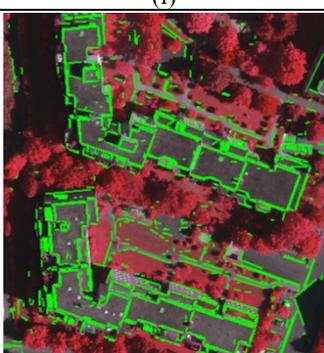
 <p>(a)</p>	 <p>(b)</p>	# of Lines Extracted	Left	1823
			Right	1877
		Matches	Total	1106
			Correct	1088 (98.4%)
			False	18 (1.6%)
 <p>(c)</p>	 <p>(d)</p>	# of Lines Extracted	Left	2145
			Right	2151
		Matches	Total	1207
			Correct	1193 (98.8%)
			False	14 (1.1%)
 <p>(e)</p>	 <p>(f)</p>	# of Lines Extracted	Left	3035
			Right	3175
		Matches	Total	1901
			Correct	1874 (98.6%)
			False	27 (1.4%)
 <p>(g)</p>	 <p>(h)</p>	# of Lines Extracted	Left	1618
			Right	1715
		Matches	Total	900
			Correct	881 (97.8%)
			False	19 (2.1%)

Figure 7. The results of the proposed approach. Matched line segments are shown in green color in the left stereo images (a-c-e-g) and the right stereo images (b-d-f-h) for the selected four test sites.

(especially the ones belong to the road segments) that can be piece-wise linear approximated are also matched successfully. Actually, this is not a surprising fact, since the piece-wise approximated linear segments are also particularly suitable to be matched by the proposed pair-wise approach.

4. CONCLUSIONS

In this study, we developed a new pair-wise relation based approach for the matching of line features from stereo aerial images. To solve the matching inconsistencies, we proposed an iterative pair based post-processing algorithm. The novelty of this study originates from the newly defined measures and the iterative pair-wise elimination in which the nearest/next ratios and a final similarity voting scheme are applied.

Based on the results of the selected test sites, the proposed approach produces accurate and robust results for urban areas even under challenging cases, such as repetitive linear patterns. We would like to stress once more that, we do not impose any external dataset to the matching (third view, DSM etc.), and do not perform any ill-posed constraint, such as epipolar ordering to solve the matching ambiguities. Thus, our first aim for the future work is to adapt the algorithm into a multi-stereo approach where the third image is fully integrated. For sure, the addition of the third image in a multi-stereo approach will boost the performance of our algorithm in all aspects; accuracy, robustness and completeness. Currently, we are also investigating new approaches that take into account the pair relations for the reconstruction of line features from stereo images. The pair-wise approach also provides new opportunities for solving the most problematic case of the reconstruction in which the matched line segments are exactly aligned with the epipolar line (flight direction).

ACKNOWLEDGEMENTS

The Vaihingen dataset was provided by the German Association for Photogrammetry and Remote Sensing (DGPF): <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>

REFERENCES

- Bay, H., Tuytelaars, T., and Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: *European Conf. Computer Vision*.
- Cramer, M., and Haala, N., 2009. DGPF Project: Evaluation of digital photogrammetric aerial based imaging systems – overview and results from the pilot centre. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII (1-4-7/W5)*, CD-ROM.
- Herbert, B., Vittorio, F., and Luc, V.G., 2005. Wide-baseline Stereo Matching with Line Segments, in: *IEEE International Conference on Computer Vision and Pattern Recognition*.
- Mikolajczyk, K., and Schmid, C., 2005. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(10), pp. 1615-1630.
- Lowe, D.G., 2004. Distinctive Image Features from Scale Invariant Keypoints, *International Journal of Computer Vision*, 20(2), pp. 91-110.
- Ok, A. O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., and Toprak, V., 2010. A new straight line reconstruction methodology from multi-spectral stereo aerial images. submitted to Photogrammetric Computer Vision and Image Analysis Conference (PCV'10), accepted.
- Schmid, C., and Zisserman, A., 1997. Automatic Line Matching Across Views. In: *Proceedings of CVPR*, pp. 666–671.
- Scholze, S., Moons, T., Ade, F., and Van Gool, L., 2000. Exploiting Color for Edge Extraction and Line Segment Stereo Matching. In: *IAPRS*, pp. 815-822.
- Suveg, I., and Vosselman, G., 2004. Reconstruction of 3D Building Models from Aerial Images and Maps, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, pp. 202-224.
- Tola E., Lepetit, V., and Fua, P., 2010. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5), pp. 815-830.
- Wang, Z., Wu, F., and Hu, Z., 2009. MSLD: A Robust Descriptor for Line Matching. *Pattern Recognition* 42, pp. 941-953.
- Zhang, C., and Baltsavias, E. P., 2000. Edge matching and 3D road reconstruction using knowledge-based methods, *Schriftenreihe der Fachrichtung Geodäsie*, Darmstadt, Germany, 10, pp. 251-265.
- Zhang, L., 2005. Automatic Digital Surface Model (DSM) Generation from Linear Array Images, PhD Thesis, Swiss Institute of Technology Zurich.