AUTOMATED PHOTOGRAMMETRIC IMAGE MATCHING WITH SIFT ALGORITHM AND DELAUNAY TRIANGULATION

Georgios Karagiannis, Francesc Antón Castro, Darka Mioc

National Space Institute, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
giorgkarag@gmail.com, francescantoncastro@me.com, mioc@live.com

Commission II WG II/2

KEY WORDS: Automated image matching, SIFT algorithm, Delaunay triangulation, graph isomorphism, multi-sensor image matching, multi-temporal image matching.

ABSTRACT:

An algorithm for image matching of multi-sensor and multi-temporal satellite images is developed. The method is based on the SIFT feature detector proposed by Lowe in (Lowe, 1999). First, SIFT feature points are detected independently in two images (reference and sensed image). The features detected are invariant to image rotations, translations, scaling and also to changes in illumination, brightness and 3-dimensional viewpoint. Afterwards, each feature of the reference image is matched with one in the sensed image if, and only if, the distance between them multiplied by a threshold is shorter than the distances between the point and all the other points in the sensed image. Then, the matched features are used to compute the parameters of the homography that transforms the coordinate system of the sensed image to the coordinate system of the reference image. The Delaunay triangulations of each feature set for each image are computed. The isomorphism of the Delaunay triangulations is determined to guarantee the quality of the image matching. The algorithm is implemented in Matlab and tested on World-View 2, SPOT6 and TerraSAR-X image patches.

1 INTRODUCTION

Most of the older and recent researches (Lowe, 2004, Lowe, 1999, Feng et al., 2008, Yang and Kurita, 2013, Harris and Stephens, 1988, Moravec, 1981, Shi and Tomasi, 1994, Zhao and Ngo, 2013, Harris, 1993) on image matching and registration are based on the concept of detecting feature points in the reference image and then matching them to the corresponding feature points in the other image. This is highly challenging considering that the only information available for a point is its reflectivity in a certain portion of the EM spectrum. Certainly, by combining the spectral information of surrounding points, geometrical and topological information can be derived.

In order to solve this problem, local interest points with, as much as possible, unique geometrical, topological and spectral characteristics have to be detected. These points should be highly distinctive in the sense that they can be identified successfully against a large database of other points. This uniqueness of feature points is necessary in image matching because in most of the cases in real life, images taken at different dates or/and from different sensors are at the same time rotated, translated and different in scale and illumination. The problem of matching becomes even more complicated by accounting the local and global distortion in both reference and sensed images. In addition, satellite images are even more demanding because they cover very large areas that can confuse the algorithm.

Furthermore, the ground-breaking work of Lowe in 1999 (Lowe, 1999) extended the local feature based previous approaches even more by proposing a scale invariant method, the prominent Scale Invariant Feature Transform (SIFT) method. Even though it is not actually a transform, it is called transform in the sense of transforming image data into scale-invariant coordinates(Lowe, 2004). This method is invariant not only in scale but also in rotations, translations and, partially, in illumination changes. A scale space is created by smoothing the images with a Gaussian filter and then sub-sampling them, creating a pyramid structure in which the levels are actually smoothed versions of the original images. Then, the neighboring pyramid layers are subtracted producing the Difference of Gaussian (DoG) images. Afterwards, local extrema points are detected in the DoG images that represent the candidate feature points. Consequently, feature descriptors are created by assigning an orientation to each feature point using 36 bins covering the 360° of a full circle. Finally, feature points of the reference image are matched with their corresponding feature points in the sensed image by a nearest neighbor criterion.

Finally, in 2008, Bay et al. (Bay et al., 2008) proposed a feature detection method that, in certain applications, approximates or even outperforms other feature detectors such as Harris or SIFT. The Speeded-Up Robust Features (SURF) method relies on the use of integral images that result in a notable reduction of the number of operations. Each entry in the integral image is defined as the summary of the corresponding point in the original image with the neighboring pixels of a specified square neighborhood. The SURF method is conceptually similar to the SIFT one, with the main difference relying on the scale space construction. Using integral images instead of the original ones enables the scale space construction through box filters of any size at exactly the same speed directly to the original image and even simultaneously. In this way, instead of iteratively sub-sampling the original image, the box filter is up-scaled. This difference reduces drastically the number of operations and thus, the required computational time.

There have been many research works on image quality but surprisingly very few on the specific problem of image matching quality. Such quality measures compare either a mesh with the original image (Fogtmann and Larsen, 2013) or the objects (targets) in two images (Cao and Duan, 2011). The former work focuses on image-mesh matching and therefore, it is not applicable to our problem since we need either to compare two images or their meshes. The later work uses classical linear parametric statistics techniques, that assume a priori that data (images) obey some probability distribution function. To the best of our knowledge, there are no image matching quality measures based on non-linear, non-parametric statistical techniques, which only assume the local smoothness of the data. By opposition to these research works, the present research work focuses on a deterministic image matching quality measure: the percentage of edges in the subgraph isomorphism between Delaunay graphs (the dual graph of the Voronoi diagram or Dirichlet tessellation of the feature points, which captures the geometric topology of the objects). If the image matching is perfect from the point of view of geometric topology, the two Delaunay graphs are isomorphic and the image matching quality measure is 100 %.

To the best of our knowledge, there has not been any research work using Delaunay triangulation for the automated check of SIFT based image matching. This paper is organised as follows. Section 2 introduces the SIFT method for image matching. Section 3 shows the results of the automated quality control of the SIFT method based on Delaunay graph isomorphism. Finally, we conclude this paper in Section 4.

2 SIFT-BASED IMAGE MATCHING USING HOMOGRAPHS

The approach that was followed can be divided into three main steps, each one described in separate sections of this chapter. These steps are:

1. **SIFT feature extraction:** Detection of stable feature points from both the reference and the sensed images for an accurate match. This is performed by implementing the SIFT feature detection method as described by Lowe in (Lowe, 2004) and (Lowe, 1999). All the three steps are very important but the quality of this one is the most crucial for the quality of the final accuracy. Any inaccuracy will be compounded until the end influencing all the following processes.

2. **Feature matching:** After the independent extraction of the SIFT feature points from both images, features that represent the same point in both images are matched. Each feature point of the reference image is matched with its corresponding feature point of the sensed image by computing the Euclidean distance between that feature point and all the feature points in the sensed image. Then, the nearest neighbor is considered a candidate for matching. In order to avoid including false positive matches (points that erroneously have been matched) and discarding false negative matches (matches that mistakenly have not been included), the distance to the nearest neighbor and the distance to the second closest neighbor is also computed. This is based on the assumption that the distance to the nearest neighbor will be much shorter compared to the one to the second closest.

3. **Homographic transformation:** Finally, after the detection of pairs of matched points with known image coordinates in both images, the parameters of the image matching will be computed accurately. These parameters take into consideration any variation in translations and rotations in addition to scaling and skewness between the two image coordinate systems and form a transformation. Specifically, they form a homographic transformation whose projection in two dimensions corresponds to a similarity.

The implementation can be represented by four main stages (Lowe, 2004):

1. **Scale-space construction and space extremum point detection:** The algorithm searches all image scales and locations by computing the Laplacians of Gaussians (LoG) for the image with various σ values. The different σ values act like a scale parameter and in this way, feature points that are -potentially in this stage- invariant to scale and rotations are detected. Difference of Gaussians (DoG) is the difference of two blurred versions of the original image. The two blurred versions occur by applying a Gaussian filter with different σ in the original image(Lowe, 2004).

2. **Key-point localization:** For each candidate point from the previous stage, a fit to the nearby data for location, scale and ratio of principal curvatures is performed. Points that are sensitive to noise (have low contrast) or are not well localized on an edge are discarded. In this way, the 2-dimensional translation and scale invariance is reached.

3. **Orientation assignment:** In this stage, the remaining points after the previous stage are assigned with one or more consistent orientations based on the average direction of the gradient in the vicinity of the point. In this way, the invariance to image rotation is achieved.

4. **Key-point descriptor:** In the previous two stages, the rotation, scale and 2-dimensional translation invariance was ensured. The goal of this stage is to attain the invariance in illumination and 3-dimensional viewpoint of the features. For this purpose, a local image descriptor incorporates the magnitude of the regional gradient for each feature point at selected scale.

These points will be used to compute the parameters that can allow the computation of the image coordinates of a point on the second image when its image coordinates on the first image are known. These parameters will include the rotations, the translations and the scaling that has to be applied on the coordinate system of the second image in order to *transform* it to the coordinate system of the first image. These parameters are the parameters of the *homographic transformation* and are the elements of the *homographic* matrix H (Equation 1).

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{1}$$

It is noted that the homography assumes that the images follow the pinhole camera model (the aperture is almost zero and all lenses have negligible width). Now, let $X_i^{'}$ be the vector coordinates of a point in the first image, X_i be the corresponding coordinates of the point in the second image and H be the homographic matrix. Then, the relationship of the two points is shown in Equation 2 which is known as the *homography equation*.

$$X_i^{'} = HX_i \tag{2}$$

where $X_i^{'}$ and X_i are in homogeneous coordinates:

$$X_i^{'} = \begin{bmatrix} x_i^{'} \\ y_i^{'} \\ 1 \end{bmatrix} \tag{3a}$$

$$X_i = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{3b}$$

Therefore, the image coordinate vector of the first image becomes:

$$X_i^{'} = \begin{bmatrix} \rho_i^{'} x_i^{'} \\ \rho_i^{'} y_i^{'} \\ \rho_i^{'} \end{bmatrix} \tag{4a}$$

$$\rho_i^{'} = -\frac{Z_i^{'}}{f} \tag{4b}$$

where $Z_i^{'}$ is the distance between the optical center of the camera and the object in the real world in meters and f is the focal length of the camera.

From Equations 1 and 4a, the homography (Equation 2) can be expanded(Kheng, 2012):

$$\rho_i^{'} x_i^{'} = h_{11} x_i + h_{12} y_i + h_{13} \tag{5a}$$

$$\rho_i^{'} y_i^{'} = h_{21} x_i + h_{22} y_i + h_{23} \tag{5b}$$

$$\rho_i^{'} = h_{31} x_i + h_{32} y_i + h_{33} \tag{5c}$$

In addition, the homography is defined in unspecified scale since scaling H by scale s does not change the homography equation(Kheng, 2012):

$$(sH)X_i = sX_i^{'} = X_i^{'} \tag{6}$$

Therefore, h_{33} can be set $h_{33} = 1$ and by substituting h_{33} and $\rho_i^{'}$ from 5c to 5a and 5b we get:

$$x_i^{'} = h_{11} x_i + h_{12} y_i + h_{13} - h_{31} x_i x_i^{'} - h_{32} y_i x_i^{'} \tag{7a}$$

$$y_i^{'} = h_{21} x_i + h_{22} y_i + h_{23} - h_{31} x_i y_i^{'} - h_{32} y_i y_i^{'} \tag{7b}$$

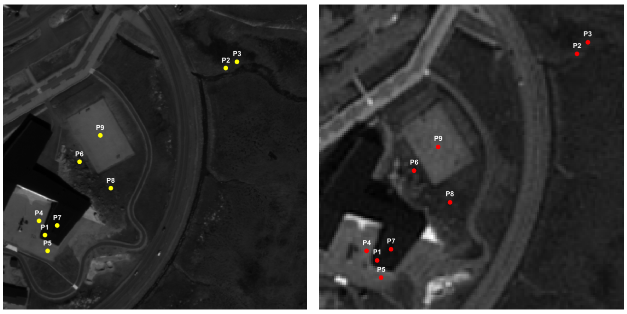
For many points, Equations 7a and 7b yields a system of equations:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x_1^{'} & -y_1 x_1^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n x_n^{'} & -y_n x_n^{'} \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y_1^{'} & -y_1 y_1^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_n y_n^{'} & -y_n y_n^{'} \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x_1^{'} \\ \vdots \\ x_n^{'} \\ y_1^{'} \\ \vdots \\ y_n^{'} \end{bmatrix} \tag{8}$$

Equation 8 is a linear system of equations with eight unknowns (the elements of the homographic matrix). Therefore, four unique pairs of points with known image coordinates in both images are enough to solve it. If more observations are available the error of the computation can be minimized by using least squares in 3 dimensional affine coordinates rather than homogeneous coordinates. In practice, the transformation parameters computed using six correct points result in sub-pixel accuracy in this research.

3 QUALITY CHECK OF SIFT ALGORITHM WITH DELAUNAY TRIANGULATION

Figure 1 shows the matched points obtained as a result of the aforementioned processing. Figure 1a shows the points in the World View patch and Figure 1b shows those in the SPOT6 patch. It can be observed that there are no false positives among the points and that they are nine in total, a number that is sufficient for the computation of the parameters. The nine pairs of points (observations) are then used to compute the transformation parameters via Equation 8 as described in Section 2. The computed parameters can transform image coordinates from the patch of the World View image to image coordinates of the patch of SPOT6 with a mean accuracy of 0.43 pixels in x-coordinate and 0.51 pixels in y-coordinate. This accuracy is satisfactory since it is below one pixel.



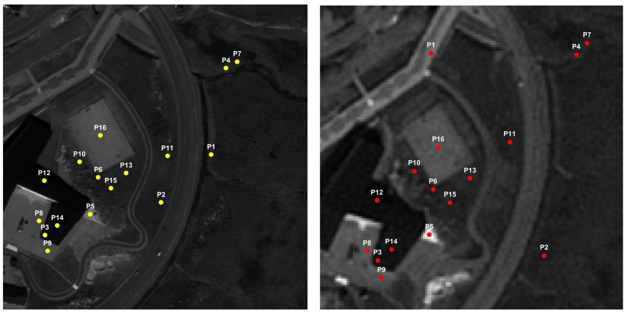
(a) Matched feature points (yellow dots) in patch number 7 in World View image

(b) Matched feature points (red dots) in patch number 7 in SPOT6 image

Figure 1: Matched feature points in patch number 7 in World View (1a, yellow dots) and SPOT6 image (1b, red dots). Note the fact that there are no false positive matches.

Figure 2 shows the result of the matching process in the same pair of patches like in Figure 1 but with a slightly looser threshold. The result is that seven more matches were detected, two of which were false positives. It is interesting to see how these two false positive observations will influence the accuracy on the computations of the transformation parameters. In this case, the image coordinates of the second image were computed with an accuracy of 3.27 pixels for the x-coordinate and 3.84 pixels for the y-coordinate. Certainly, these values cannot be considered poor but the increase of the error with just two false positives is significant.

Figures 3 and 4 show the Delaunay triangulations of the matched points on patch number 7 in both images for the two different distance thresholds. The labels that start with a "V" denote a vertex of the triangulation and those that start with a "T" represent a triangle. Moreover, the red polygons show the convex hull of each triangulation. In Figure 3, the two triangulations are almost identical. Most of the triangle corners are equal and only few are almost equal. An important observation in this figure is the size and the shape of the convex hull, which in this case is big



(a) Matched feature points (yellow dots) in patch number 7 in World View image with distance threshold at 2.5

(b) Matched feature points (red dots) in patch number 7 in SPOT6 image with distance threshold at 2.5

Figure 2: Matched feature points in patch number 7 in World View (1a, yellow dots) and SPOT6 image (1b, red dots) with distance threshold at 2.5. This change in the threshold was enough to result in 7 more matched points (16 in total) with the cost of 12.5% commission error (2 out of 16 are false positives).

enough but narrow. The size and the shape of the convex hull of the Delaunay triangulation is an indication of the distribution of the points in an image. A small sized convex hull means that the points are all located in a small region of the image. In addition, a narrow convex hull occurs when there is good distribution in one direction but not in the other. Particularly, the two convex hulls show a good distribution in the y-direction but a poor one in the x-direction. Ideally, some points would exist in the vicinity of the left and right boundaries of the images. The good distribution of the points is desired in order to minimize the error from image distortion when computing the transformation parameters. However, even though it didn't influence the accuracy in this case.

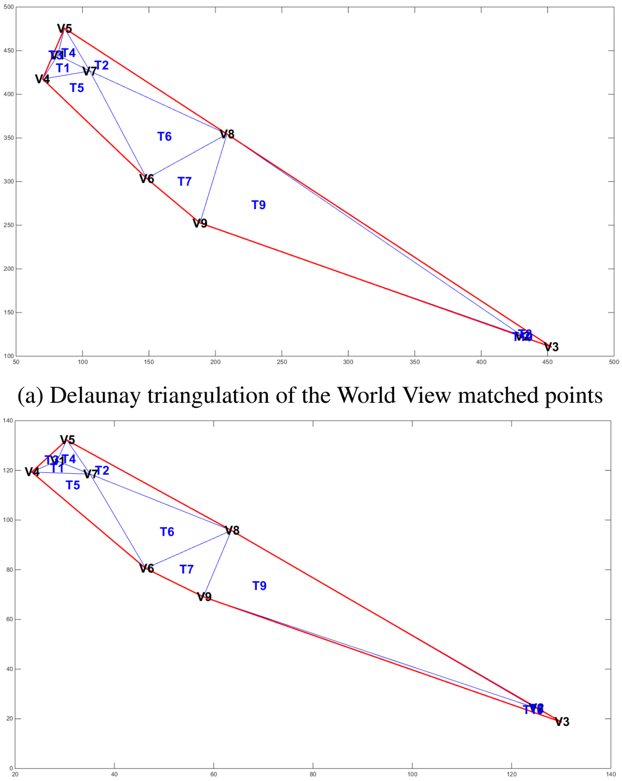
The influence of the false positives on the Delaunay triangulation can be observed in Figures 5, and 6. These figures are plots of the graph isomorphism of the triangulations shown in Figures 3 and 4. In Figure 5 the graphs that correspond to the two triangulations of the generators (feature points) shown in Figure 3 are shown. It can be observed that the two graphs are identical, indicating that each vertex is connected with the same vertices in both triangulations. In contrast, the applications with a looser distance threshold are different, as it can be seen in the corresponding graphs on Figure 6. Figures 7 and 8 show the minimum spanning trees of the graphs shown in Figures 5 and 6. A minimum spanning tree shows the minimum way of connecting the nodes of a graph. It can be observed that a single false positive can change significantly the connections among the detected features which implies a reduced accuracy.

4 CONCLUSIONS

In this paper, we have presented a novel quality control technique based on Delaunay triangulation isomorphism (or subgraph isomorphism) to assess SIFT-based image matching. We have applied this technique to multi-sensor, multi-temporal image matching. Further work will present a matching measure based on Delaunay triangulation subgraph isomorphism.

REFERENCES

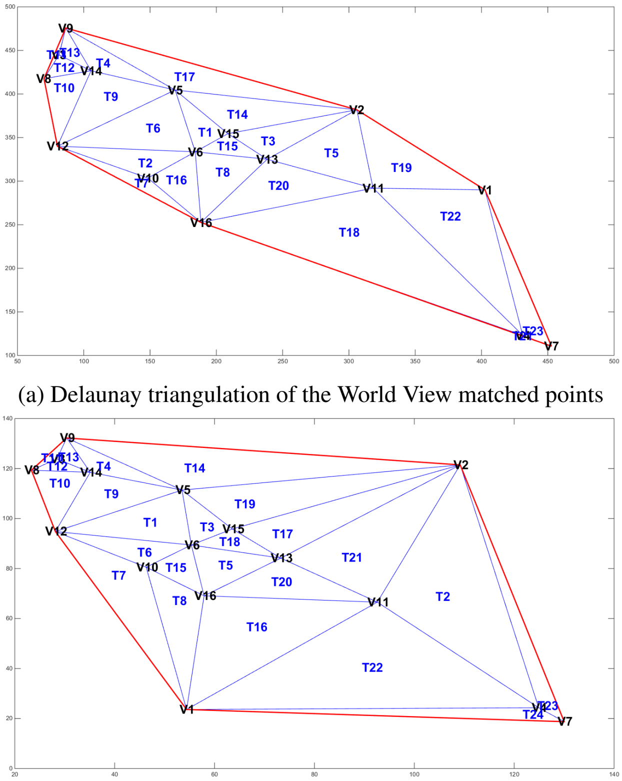
Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V., 2008. Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), pp. 346 – 359. Similarity Matching in Computer Vision and Multimedia.



(a) Delaunay triangulation of the World View matched points

(b) Delaunay triangulation of the SPOT6 matched points

Figure 3: Delaunay triangulations of the matched points on both World View and SPOT6 patch number seven for distance threshold of 3. Labels with letter T in front represent number of triangle and those with letter V in front represent a vertex. The red polygon represents the convex hull of the triangulation in each case. The axes correspond to number of pixels.



(a) Delaunay triangulation of the World View matched points

(b) Delaunay triangulation of the SPOT6 matched points

Figure 4: Delaunay triangulations of the matched points on both World View and SPOT6 patch number seven for distance threshold of 2.5. The labels are shown in the same way like in Figure 3. Note how different look the two triangulations with only two false positives out of 16 points in total.

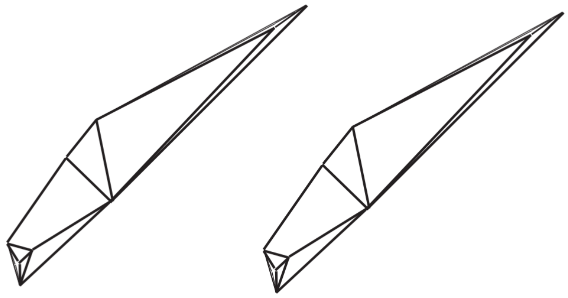


Figure 5: Graph isomorphism of the Delaunay graphs of the matched points on both World View and SPOT6 patch number seven for distance threshold of 3. The two graphs, if assumed without their geometric embeddings, are identical proving that the points are distributed in the same way in the two images.

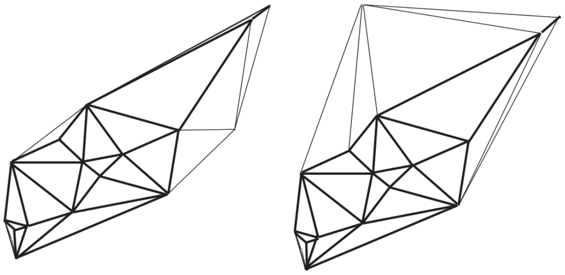


Figure 6: Graph isomorphism of the Delaunay graphs of the matched points on both World View and SPOT6 patch number seven for distance threshold of 2.5. The Delaunay edges with the highest width are isomorphic. The edges with the lowest width are not isomorphic. The subgraphs corresponding to the Delaunay edges of highest width are isomorphic. The percentage of subgraph isomorphism is $\frac{33}{40}$=82.5 %

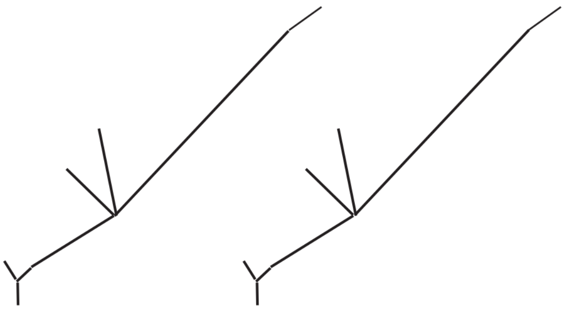


Figure 7: Minimum spanning tree of graph in Figure 5

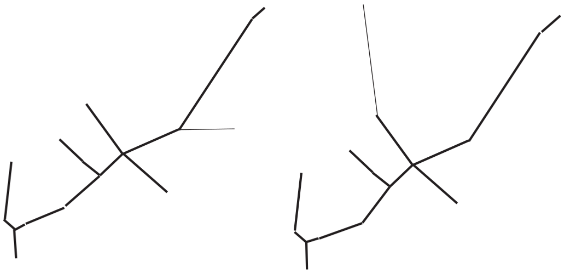


Figure 8: Minimum spanning tree of graph in Figure 6

Cao, Z. and Duan, X., 2011. Object Matching Task-oriented Image Quality Assessment. In: Cao, Z and Fenster, A and Nyul, LG and Cai, C (ed.), MIPPR 2011: MULTISPECTRAL IMAGE ACQUISITION, PROCESSING, AND ANALYSIS, Proceedings of SPIE, Vol. 8002, Huazhong Univ Sci & Technol; Natl Key Lab Sci & Technol Multi-Spectral Informat Proc; Guilin Univ Elect Technol; SPIE. 7th Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR) - Multispectral Image Acquisition, Processing, and Analysis, Guilin, PEOPLES R CHINA, NOV 04-06, 2011.

Feng, H., Li, E., Chen, Y. and Zhang, Y., 2008. Parallelization and characterization of sift on multi-core systems. In: Workload Characterization, 2008. IISWC 2008. IEEE International Symposium on, pp. 14–23.

Fogtmann, M. and Larsen, R., 2013. ADAPTIVE MESH GENERATION FOR IMAGE REGISTRATION AND SEGMENTATION. In: 2013 20TH IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP 2013), IEEE International Conference on Image Processing ICIP, Inst Elect & Elect Engineers; IEEE Signal Proc Soc, pp. 757–760. 20th IEEE International Conference on Image Processing (ICIP), Melbourne, AUSTRALIA, SEP 15-18, 2013.

Harris, C., 1993. Active vision. In: A. Blake and A. Yuille (eds), Active vision, MIT Press, Cambridge, MA, USA, chapter Geometry from Visual Motion, pp. 263–284.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: In Proc. of Fourth Alvey Vision Conference, pp. 147–151.

Kheng, L. W., 2012. Camera models and imaging. `http://www.comp.nus.edu.sg/~cs4243/`.

Lowe, D., 1999. Object recognition from local scale-invariant features. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, Vol. 2, pp. 1150–1157 vol.2.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), pp. 91–110.

Moravec, H. P., 1981. Rover visual obstacle avoidance. In: P. J. Hayes (ed.), IJCAI, William Kaufmann, pp. 785–790.

Shi, J. and Tomasi, C., 1994. Good features to track. In: 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), pp. 593 – 600.

Yang, Z. and Kurita, T., 2013. Improvements to the descriptor of sift by bof approaches. In: Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on, pp. 95–99.

Zhao, W.-L. and Ngo, C.-W., 2013. Flip-invariant sift for copy and object detection. Image Processing, IEEE Transactions on 22(3), pp. 980–991.