

CO-REGISTRATION OF DSMs GENERATED BY UAV AND TERRESTRIAL LASER SCANNING SYSTEMS

Ravi Ancil Persad*, Costas Armenakis

Geomatics Engineering, GeoICT Lab
Earth and Space Science and Engineering
Lassonde School of Engineering, York University
4700 Keele St., Toronto, Ontario, M3J 1P3 Canada
{ravi071},{armenc}@yorku.ca

Commission I, ICWG I/Vb

KEY WORDS: Co-registration, DSM, Automation, Wavelets, Matching, UAV, Laser Scanning

ABSTRACT:

An approach for the co-registration of Digital Surface Models (DSMs) derived from Unmanned Aerial Vehicles (UAVs) and Terrestrial Laser Scanners (TLS) is proposed. Specifically, a wavelet-based feature descriptor for matching surface keypoints on the 2.5D DSMs is developed. DSMs are useful in wide-scope of various applications such as 3D building modelling and reconstruction, cultural heritage, urban and environmental planning, aircraft navigation/path routing, accident and crime scene reconstruction, mining as well as, topographic map revision and change detection. For these listed applications, it is not uncommon that there will be a need for automatically aligning multi-temporal DSMs which may have been acquired from multiple sensors, with different specifications over a period of time, and may have various overlaps. Terrestrial laser scanners usually capture urban facades in an accurate manner; however this is not the case for building roof structures. On the other hand, vertical photography from UAVs can capture the roofs. Therefore, the automatic fusion of UAV and laser-scanning based DSMs is addressed here as it serves various geospatial applications.

1. INTRODUCTION

Vertical UAV imagery enables us to capture roof details of buildings and other absent structure data which terrestrial laser scanners (TLS) may not be able to capture. However, TLS data provides rich details of 'ground-level' objects such as trees, cars and building facades. Fusion of UAV and TLS data can be used in a variety of geospatial applications such as cultural heritage, urban and environmental planning, aircraft navigation/path routing, accident and crime scene reconstruction, mining, as well as, topographic map revision and change detection. For aligning a DSM pair, corresponding features and the mathematical transformation to map the 'source' DSM to 'target' DSM must be established. In this work, we propose an approach to automatically co-register UAV and TLS DSMs.

The problem of co-registering a pair of DSMs (i.e., a 'source' and 'target' DSM) depends on the type of unknown transformation parameters which has to be retrieved. For instance, if both DSM are of the same scale but are displaced by unknown rotation and translation, global alignment can be achieved by shifting centroids and applying a principal component analysis-based orientation. We concentrate on the most challenging case of 3D rigid alignment whereby both DSMs differ in terms of their scale, rotation, translation and resolution. A feature matching approach is used for this purpose (Fig. 1).

In the first step, an automated point extraction is carried out to establish points of salient morphological significance, i.e.,

keypoints in each dataset. This is done using the Harris corner extraction approach (Harris and Stephens (1988)). Once keypoints are detected, their descriptors (i.e., attributes) are generated afterwards. The descriptors are used to find corresponding keypoints with similar DSM surface characteristics.

Wavelets capture the local patch structure via elevation changes on the 2.5D DSM image along various orientations (e.g., horizontal and vertical). We use wavelet-based descriptors in this work. By applying a log-polar sampling grid to the local DSM regions around the keypoints, the rotational and scale differences between similar source and target DSM surface regions can be expressed in terms of a translational offset in the log-polar space.

We begin by computing directional derivatives of the 2.5D DSM images using a wavelet kernel. Then by applying a log-polar grid and sampling wavelet features, we are able to encode the structural characteristics of local height map DSM 'patches' surrounding the keypoint. This gives us a log-polar based descriptor of our DSM keypoint patch. However, actual corresponding source and target log-polar based descriptors differ by a translation. Therefore, in the final step of the descriptor formation, we apply a fast Fourier transform (FFT) transform on the log-polar descriptors to address the aforementioned translation offset. Hence, the final set of descriptors is in the Fourier domain.

We measure the similarity of keypoints with a cost function

* Corresponding author

(i.e., NNDR: nearest neighbour distance ratio (Lowe, 2004)) by comparing the Euclidean distances for their 1st and 2nd nearest neighbours in the descriptor feature space. If the NNDR tends to have a value of 1 or close to it, then the match is not stable as the difference between the 1st and 2nd nearest neighbour distances is very small. However, a small NNDR represents a potential correspondence. When the NNDR is smaller than a user-defined threshold, a match is established. To refine the NNDR-based matches, the RANSAC method is then applied to remove correspondence outliers. Finally, we compute the 7-parameter 3D conformal transformation, between the source and target DSMs, thus solving for the scale factor, rotational angles and translation.

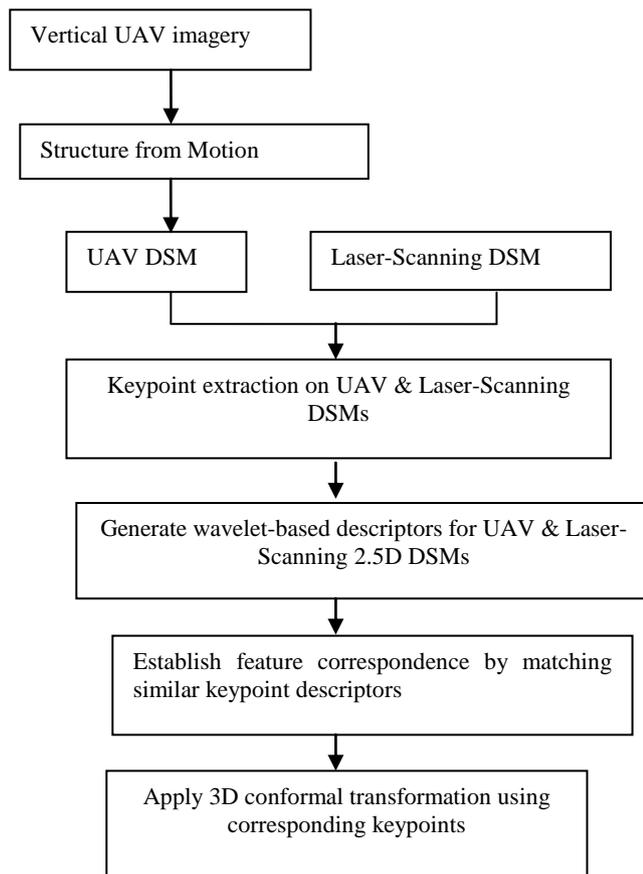


Figure 1. Proposed approach for UAV & TLS point cloud DSM alignment

We present the proposed methodology and experiments on both simulated and actual datasets. To quantify the quality of transformation parameters derived via our automated approach a comparison is made with 'reference' parameters. Reference parameters have been obtained via manual selection of point matches on the DSM pair used for the mapping of the source to the target DSMs.

2. RELATED WORKS

The co-registration of airborne laser scanning (ALS) data with those of TLS has been a recent area of interest. This is important to note since matching and alignment frameworks applied for integration of Airborne Laser Scanner (ALS)/TLS

data can be similarly adopted for point cloud DSMs derived from vertical UAV photography and TLS data. Yang et al. (2015) presented an approach for ALS/TLS alignment of urban areas. Their approach relies on the generation of spatial geometric constraints (such as distances, directions and connectivity) derived from building outlines. Afterwards, they adopt a spectral graph matching approach to find correspondences. The premise of the spectral approach is the utilization of eigen-decomposition (i.e. eigenvectors, eigenvalues) to model the ALS and TLS building outline structural similarity. However, this proposed framework is limited to urban scenes with the necessity to extract building outlines. Urban areas generally comprise of many other rich descriptive details such as road, trees, and vegetation. Therefore, the approach may suffer in datasets which are not predominantly populated by buildings. Additionally, the approach does not seem to handle the possibility of scale differences between ALS and TLS data and only solves for rotation and translation in the 3D space. Their assumption of a known global scale factor greatly simplifies the matching problem between the airborne and terrestrial datasets. For instance, unary constraints such as area, edge lengths and binary constraints such as distances between pairwise edges are strong cues to locate corresponding features.

The co-registration of DSM point cloud datasets with a global scale difference is a challenging problem. Novák and Schindler (2013) presented a method to address the alignment of laser scanning data with photogrammetric point clouds without prior knowledge of rotation, translation and scale. Their framework relies on the generation of height maps by projecting the laser scanning and photogrammetric 3D point clouds onto the planimetric x-y plane. Gradient information and RANSAC (Fischler and Bolles, 1981) on the height maps are then used for matching.

Persad and Armenakis (2015) presented a framework for UAV and TLS alignment. Similar to the work by Novák and Schindler (2013), it is formulated as a height map to height map matching problem. Scale-invariant keypoints are extracted on the UAV and TLS DSM surfaces using curvature information. Afterwards, the SURF (Bay et al. (2008)) descriptor is then used for finding similar keypoints. The use of height map keypoints and their descriptors for matching UAV and TLS point data is not limited to certain scenes (e.g. urban buildings) as with the method presented in Yang et al. (2015). Therefore, in this paper we propose a DSM height map co-registration framework using keypoints as features.

The use of wavelets for descriptors has been applied for various applications. Pun and Lee (2003) proposed a scale and rotation invariant wavelet energy signature based on wavelet transforms and metrics such as entropy, standard deviation and other energy measures. The wavelet energy signatures were used for the classification of textured images. de Ves et al. (2007) also proposed a wavelet-based texture descriptor for an image retrieval system. Even though they also utilize wavelet transforms for descriptor formation, descriptor coefficients were extracted using a gamma distribution probability density function. Amiri and Rabiee (2011) developed a wavelet-descriptor referred to as 'RASIM' (Rotation and Scale Invariant Matching). This approach was developed for keypoint matching on images. It uses a log-polar sampling grid around the keypoints and for each ring on the sampling grid a wavelet transform is applied to form the descriptor.

3. METHOD

3.1 Keypoint extraction on UAV and TLS DSMs

We treat the matching of UAV/TLS height map patches as an image to image matching problem. Identifying corresponding entities in overlapping images has been a difficult problem and an important research effort in the photogrammetry and computer vision communities. The types of entities to be matched must be defined along with the techniques for detecting and matching them. Remondino (2006) provides a review of point feature extraction approaches used in photogrammetry and computer vision applications.

In this work, we utilize the Harris corner detector (Harris and Stephens (1988)) for definition of our interest points. The idea of the Harris detector is to identify keypoints at locations where there is significant change in gradient intensity on the image. The method begins by computation of horizontal and vertical derivatives of the image. Afterwards, the derivatives are used to form a 2x2 Hessian matrix H for each patch (Harris and Stephens (1988)). A measure of corner strength is defined via a ‘corner response’ measure. We use the Noble measure (Noble (1989)) instead of the original Harris metric which requires a user-set parameter. The Noble measure is defined by Eq. (1).

$$\text{Corner response} = \frac{\det(H)}{\text{trace}(H)} \quad (1)$$

We extract keypoints on both the UAV and TLS DSM height maps.

3.2 Keypoint description

Scale, rotation and translation invariance for matching keypoints is critical. In this work we use a wavelet-based descriptor. Our descriptor builds on the work of Tola et al. (2010) and Kokkinos et al. (2012). Our approach comprises of the following steps: i) estimation of wavelet-based directional derivative gradients on the DSM height maps, ii) application of a log-polar transformation gridding around height map keypoints, iii) determination of descriptors based on FFT and iv) matching of descriptor.

As identified in Zokai and Wolberg (2005), the log-polar transform of an image and its scaled and rotated version is the same. Therefore, we utilize this rotation and scale invariant characteristic at each keypoint. The log-polar transform has been widely applied in the well-known Fourier-Mellin Transform (Reddy and Chatterji (1996)) for its transformation invariant properties. However, Fourier-Mellin is used for global registration applications. In our case, the point to point matching is local-based. Nevertheless, the general concept of the log-polar transform of an image is adopted for obtaining scale and rotation invariant representations for local keypoints.

We compute the horizontal and vertical derivatives of the height maps using the Morlet wavelet kernel (Bultheel (2003); Viswanathan et al. (2014)). The Morlet kernel (mk) is shown in Eq. (2). We extract horizontal and vertical structures on the image by applying different θ values. In this work, we use $\theta = 0^\circ$ and 90° . Additionally, we empirically set the size of the wavelet kernel as 30 pixels.

$$mk = e^{-0.5(x^2+y^2)} \cdot e^{2\pi i(x\cos\theta+y\sin\theta)} \quad (2)$$

where, θ specifies the orientation of the wavelet kernel and x, y is the size of the wavelet kernel.

Then we sample a local circular region of the wavelet-generated derivatives around the keypoints using a log-polar grid. The idea is similar to the sampling scheme used in Belongie et al. (2002). The log-polar grid is a series of concentric circles with increasing size which are split into various sectors. In this paper, the maximum radius defining the local region is set to 25% of \mathcal{E} . \mathcal{E} is taken to be the maximum height map dimension (i.e., if the width of the height map is greater than its height then $\mathcal{E} = \text{width}_{\text{height map}}$ and vice versa if the height is larger than the width). We also experimentally define the number of concentric circles as $M = 25$ and the number of bisecting lines to generate the circle sectors as $N = 30$.

The gridded horizontal and vertical derivatives are then binned into log-polar space. Inherent rotation and scale differences around corresponding local regions between the UAV and TLS DSMs become invariant when transformed to the log-polar space. However, the difference in the rotation and scaling manifests as a translation difference between the UAV log-polar bins and TLS log-polar bins. Therefore, to achieve invariance to this shifting, a 2D Fast Fourier Transform (FFT) (Cooley and Tukey (1965)) is applied to the log-polar bins (Fig. 2).

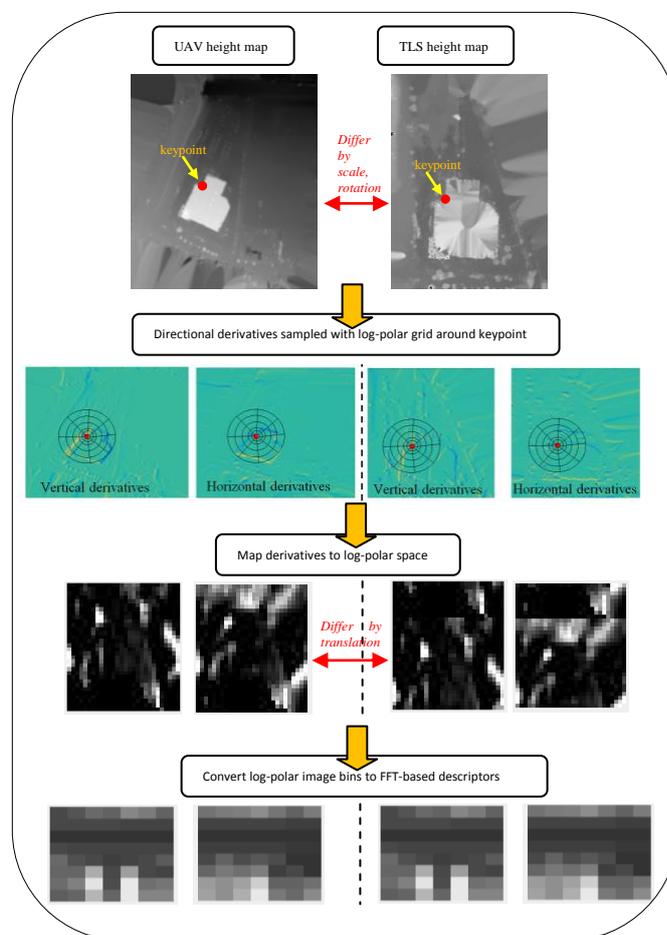


Figure 2. General approach for descriptor generation

The size of one of these Fourier-based descriptors is $2 \times M \times N$ (where is 2 is the number of derivative gradient directions).

3.3 Keypoint matching and RANSAC inlier detection

The match for each keypoint is established by utilizing the Euclidean distances for their 1st and 2nd nearest neighbours in the descriptor feature space. Generally, there are 3 possible cases which exist for feature matching in the descriptor space using nearest neighbours (Szeliski, 2010):

- (a) application of a global, fixed threshold
- (b) keep the closest nearest neighbour as the match
- (c) application of a nearest neighbour distance ratio (*NNDR*)

For case (a) a keypoint K_x with descriptor D_1 can fail to match its true keypoint correspondence with descriptor D_{true} if their relative distance $dist_1$ is greater than a ‘hard’ threshold is set by the user (Fig. 3). Such a threshold can be set as a radius bound defined by a Euclidean distance value. Additionally, multiple matches such as D_{true}^A and D_{true}^B are also possible if they fall within this threshold (Fig. 3).

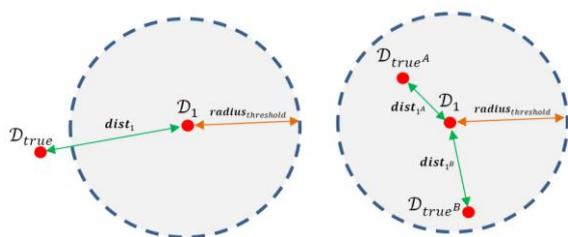


Figure 3. Concept of *NNDR* based matching

We now examine the latter two cases which are independent of the global radius-bound threshold. Visually in Fig. 3, D_{true}^A is closer to D_1 in comparison to D_{true}^B . This is an instance where case (b) can be applied, i.e., using the closest nearest neighbour D_{true}^A with distance $dist_1^A$ as the best match. D_{true}^B is the second closest nearest neighbour. However, situations may arise where $dist_1^A$ and $dist_1^B$ have very similar values. This indicates that the descriptor is unable to clearly distinguish a distinct matching keypoint. In such scenarios, we do not include K_x as part of the final correspondence set due to lack of uniqueness in identifying a clear match. This can be addressed using the *NNDR* approach (Eq. (3)), i.e., case (c) (Lowe, 2004). If *NNDR* is less than a user-defined value τ , then a correspondence is accepted ($\tau=0.3$ is used in all experiments). Afterwards, we apply the RANSAC method (Fischler and Bolles, 1981) to filter outlying keypoint correspondences. We then compute a 3D conformal transformation using inlying keypoint correspondences to estimate the scale, rotation and translation between the UAV and TLS DSMs.

$$NNDR = \frac{dist_1^A}{dist_1^B} \quad (3)$$

4. EXPERIMENTS AND RESULTS

4.1 Study area and datasets

In the following section, we present results from the proposed alignment approach using simulated and actual datasets. For the simulated tests, we analyse how the method performs when there is scale, rotation and translation change between 2 DSMs to be matched. For this test, we use a UAV DSM and apply simulated parameters to get a transformed version of itself. For the tests with actual data, we use the UAV DSM and a TLS DSM. The study area is the same site used in Persad and Armenakis (2015), located in North York, Ontario, Canada. The UAV DSM was generated using Agisoft (Agisoft, 2015) from nadir-looking images collected by the geo-X8000 UAV. The TLS points were collected from an Optech ILRIS long range scanner.

4.2 Results

For simulated test, we have the original UAV DSM and its transformed version. The transformed version was formed by applying artificial transformation parameters to the 3D UAV point clouds. The following parameters were used: scale factor of 0.5, 3D rotation angles ($\omega = 3^\circ, \phi = 5^\circ$ and $\kappa=7^\circ$) and 3D translation ($T_x = 0.4, T_y=0.6, T_z=0.8$). The original UAV point clouds are within a local coordinate system as they are generated from the images via Agisoft. The TLS points are in an orthogonal 3D system (X, Y are the Universal Transverse Mercator reference coordinate system, Z is orthometric height).

Table 1. Transformation parameters –Simulated dataset (Note: translation values are in a local pixel-based coordinate system)

Parameters	Reference	$\pm\sigma$	Automated	$\pm\sigma$
Scale	0.5	-	0.4995	1.2e-15
ω ($^\circ$)	3	-	2.951	0.00013
ϕ ($^\circ$)	5	-	5.047	0.00009
κ ($^\circ$)	7	-	7.337	0.00015
T_x	0.4	-	0.411	0.00001
T_y	0.6	-	0.608	0.00006
T_z	0.8	-	0.789	0.00011

For the simulated data, 24 keypoints were extracted on the original DSM and 19 on the transformed DSM. There were a total of 16 point matches and 2 outliers. Table 1 shows the results with the reference transformation parameters in comparison to the automatically estimated parameters.

Results for the actual co-registration of the UAV and TLS DSMs are shown in Table 2. Table 3 are the reference parameters which were generated by manually selecting 6 point matches. On the UAV DSM, there were 24 keypoints and on the TLS DSM 41 were extracted. After automated matching, there were 8 inlier correspondences. The RMSE of the residuals for the reference parameters is 0.27m and 0.13m for the automatically-derived ones. Fig. 4 shows the UAV and TLS DSM height maps and point matching results. Fig. 5 shows the resulting alignment after applying the 3D conformal transformation and using the point correspondences.

Table 2. Transformation parameters –Manual selection of corresponding points

<i>Parameters (manual)</i>	<i>value</i>	$\pm\sigma$
Scale	44.34	2.9e-08
ω (°)	7.82	0.00008
φ (°)	23.2	0.00043
κ (°)	14.1	0.00007
Tx (m)	-35.36	0.0027
Ty (m)	-39.17	0.0004
Tz (m)	-17.52	0.00014

Table 3. Transformation parameters –Automated selection of corresponding points

<i>Parameters (automated)</i>	<i>value</i>	$\pm\sigma$
Scale	44.82	2.2e-11
ω (°)	7.56	0.00009
φ (°)	23.9	0.00015
κ (°)	14.4	0.00016
Tx (m)	-35.53	0.0010
Ty (m)	-38.37	0.0008
Tz (m)	-16.95	0.00012

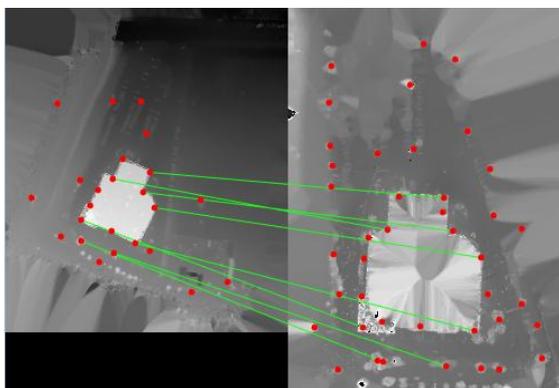


Figure 4. Left image: UAV height map, Right image: TLS height map. Green lines indicate matches for inlying point matches.

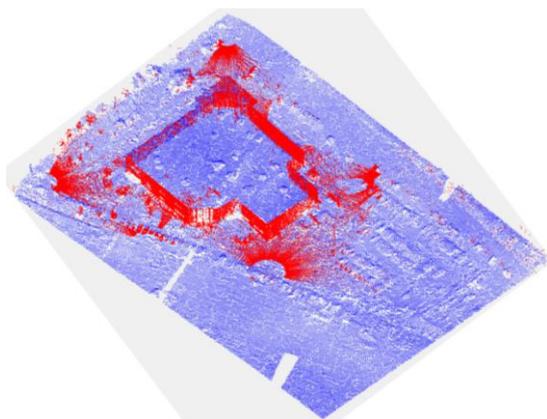


Figure 5. Alignment of UAV and TLS point clouds (Note: the original point clouds have been resampled for computational purposes).

5. CONCLUSIONS AND OUTLOOK

An automatic approach for co-registration of UAV and TLS DSMs has been presented. Salient keypoints are extracted on the DSM height maps using the Harris corner detector. Afterwards, a wavelet-based scale, rotation and translation invariant descriptor is used for characterising and matching the keypoints on the DSM height map pair. The log-polar transform is used to ensure scale and rotation invariance for the descriptors and the FFT is adopted for translation invariance. RANSAC is utilized for retaining point correspondence inliers.

Future work will look at testing on additional UAV/TLS datasets and other multi-sensor datasets. Furthermore, we intend to improve our alignment framework by looking at the development of novel approaches for keypoint detection, matching and outlier removal, as well as, to improve the wavelet-based descriptor approach in terms of accuracy and its efficiency.

ACKNOWLEDGEMENTS

We wish to thank Mike Leslar and Teledyne Optech for providing the data for this work. This work is financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

- Amiri, M. and Rabiee, H.R., 2011. RASIM: a novel rotation and scale invariant matching of local image interest points. *Image Processing, IEEE Transactions on*, 20(12), pp.3580-3591.
- Agisoft, 2015. <http://www.agisoft.com/> (Accessed 15.2.2016).
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. 2008. Speeded-up robust features (SURF). *Computer vision and image understanding* 110, no 3: 346-359.
- Belongie, S., Malik, J. and Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4), pp.509-522.
- Bultheel, A. 2003. "Wavelets with applications in signal and image processing", September 22.
- Cooley, J. W. and Tukey, O. W. 1965. "An Algorithm for the Machine Calculation of Complex Fourier Series." *Math. Comput.* 19, 297-301.
- de Ves, E., Ruedin, A., Acevedo, D., Benavent, X. and Seijas, L., 2007, August. A new wavelet-based texture descriptor for image retrieval. In *Computer Analysis of Images and Patterns* (pp. 895-902). Springer Berlin Heidelberg.
- Fischler, M.A., and Bolles, R.C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24, no 6: 381-395.
- Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, p. 50).
- Kokkinos I., Bronstein M.M., and Yuille A. 2012. "Dense scale invariant descriptors for images and surface," Tech. Rep., INRIA.
- Lowe, D.G. 2004 Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, no 2: 91-110.

- Noble, A. 1989. Descriptions of Image Surfaces (Ph.D.). Department of Engineering Science, Oxford University. p. 45.
- Novak, D. and Schindler, K., 2013. Approximate registration of point clouds with large scale differences. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(2), pp. 211-216.
- Persad, R.A. and Armenakis, C., 2015. Alignment of Point Cloud DSMs from TLS and UAV Platforms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1), pp. 369.
- Pun, C.M. and Lee, M.C., 2003. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5), pp.590-603.
- Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3), pp.49-54.
- Reddy B. and Chatterji B. 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Proc.*, 5(8):1266–1271.
- Szeliski, R. 2010. *Computer vision: algorithms and applications*. Springer.
- Tola, E., Lepetit, V. and Fua, P., 2010. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5), pp.815-830.
- Viswanathan, K., Gioia, P. and Morin, L., 2014, September. Morlet Wavelet transformed holograms for numerical adaptive view-based reconstruction. In *SPIE Optical Engineering+ Applications* (pp. 92160G-92160G). International Society for Optics and Photonics.
- Yang, B., Zang, Y., Dong, Z. and Huang, R., 2015. An automated method to register airborne and terrestrial laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, pp.62-76.
- Zokai, S., and Wolberg, G. 2005. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. *Image Processing, IEEE Transactions on*, 14(10), 1422-1434.