POTENTIAL EXPLORATION OF SEGMENTATION DRIVEN STEREO MATCHING OF VERY HIGH-RESOLUTION SATELLITE IMAGERY

Elisabeth Johanna Dippold [1]*, Fuan Tsai [1,2]

[1] Department of Civil Engineering, National Central University, 300, Zhongda Rd., Zhongli, Taoyuan 32001, Taiwan - (elisabeth.dippold@gmail.com)
[2] Centre for Space and Remote Sensing Research, National Central University, 300, Zhongda Rd., Zhongli, Taoyuan 32001, Taiwan - ftsai@csrsr.ncu.edu.tw

Youth Forum

KEY WORDS: Feature-based Matching Very High-Resolution Satellite Imagery, Normalized Differential Vegetation Index (NDVI), Image Matching, Stereo Matching

ABSTRACT:

The automatic establishment of image relationship between oblique images can be a challenging task. Feature based image matching (FBM) establishes this relationship by detecting and matching corresponding feature points. A robust matching is beneficial for many tasks including reconstruction, mapping and localization. The need of automatic processing of remotely sensed data, like very high-resolution (VHR) satellite imagery, increased over the years. Furthermore, green vegetation and water are changing the physical properties with respect to the amount of light emitted, season and pollution. In addition, with human interaction, the change in appearance increases. The Normalized Differential Vegetation Index (NDVI) is a well-known and studied index in order to detect healthy green vegetation. The Normalized Differential Water Index (NDWI) can help identify water areas in an image. They can be used to preliminarily segment images into different categories for later process. This study proposes a novel framework to explore the potential of feature based stereo matching for very high-resolution satellite imagery with segmentation. The proposed framework will perform first, image segmentation with NDVI and NDWI on stereo VHR satellite image pairs. Then, classification by threshold to detect healthy green vegetation, water and image frame. Features within these three classes are masked out and kept from being processed during the feature matching step. The idea is that, features in these classes are easy to cause miss matching because they are more prone to be affected by different image and environmental conditions in the stereo image pairs. As a result, the amount of miss matches can be reduced on average by 7.5%. Furthermore, the segmentation decreases the total amount of detected features by 13.71%, so that the processing time decreases. This study has successfully proven that segmentation can lead to improved stereo matching. In future, segmentation driven can be utilized by AI matching processes as well as traditionally photogrammetric or computer vision tasks.

1. INTRODUCTION

The scene understanding of remotely sensed data as well as in the field of photogrammetry is crucial for a comprehensive analysis. Image segmentation can helpful to gain a better understanding, especially for monitoring nature i.e. forest structure (Lamonaca et al., 2008), identification and evaluation of waste (Chen et al., 2021) or Land Use and Land Cover (Xu et al., 2021).

The need of monitoring drought in the US lead to the development of Normalized Differential Vegetation Index (NDVI) by (Kogan, 1995; Tarpley et al., 1984). Further, the exploration by comparing satellite bands in order to monitor vegetation set up the fundamentals (Tucker, 1978). Nowadays, are different methods published and reviewed (Haboudane et al., 2004) in order to monitor vegetation (Kavak et al., 2015). In addition, the Normalized Differential Water Index (NDWI) based on the key idea of NDVI (Gao, 1996). In general, minima and maxima of certain bands for different materials can help to classify natural environment. Certain band combinations can detect different types of water and vegetation (Gao, 1996) and as natural occurrence in lakes and rivers (McFeeters, 1996). Further, classification with respect to low resolution like Landsat (Xu, 2006), medium resolution like Spot (Fisher and Danaher, 2013) and VHR like World-View 3 (Niroumand-Jadidi and Vitti, 2017). The process of image matching includes detection, description, matching and, if needed, optimalization. SURF, Speeded Up Robust Features, utilises Hessian Matrix for point detection similar to Harris corner detector (Bay et al., 2006). This detector seems to perform under dynamic of nature less reliable (Yildirim et al., 2019). So that, SURF in a context of an regular structured urban environment without vegetation and water, more reliable. The strategy of matching SURF features based, for this study on FLANN, Fast Approximate Nearest Neighbours (Muja and Lowe, 2009). FLANN applies kmeans clustering instead of, commonly used, Euclidian distance. In general, the evaluation can be conducted by segmentation performance (Chen et al., 2018), by benchmark (Bian et al., 2017) or ground truth (Hadi and Khalaf, 2021). The evaluation of this study, focuses on the overall performance in order to explore the potential of segmentation driven matching of VHR satellite imagery.

The rest of this study is organized as follows: In section 2. study area and materials, followed by the framework design and finally results and conclusion.

2. STUDY AREA AND MATERIALS

2.1 Study Area

The island of Taiwan is located in the West Pacific, South East Asia (figure 1). Pleiades VHR satellite imagery uses UTM and WGS84 for positioning. So that, Taiwan can be found within the UTM system at Zone 51 of direction North.

* Corresponding author

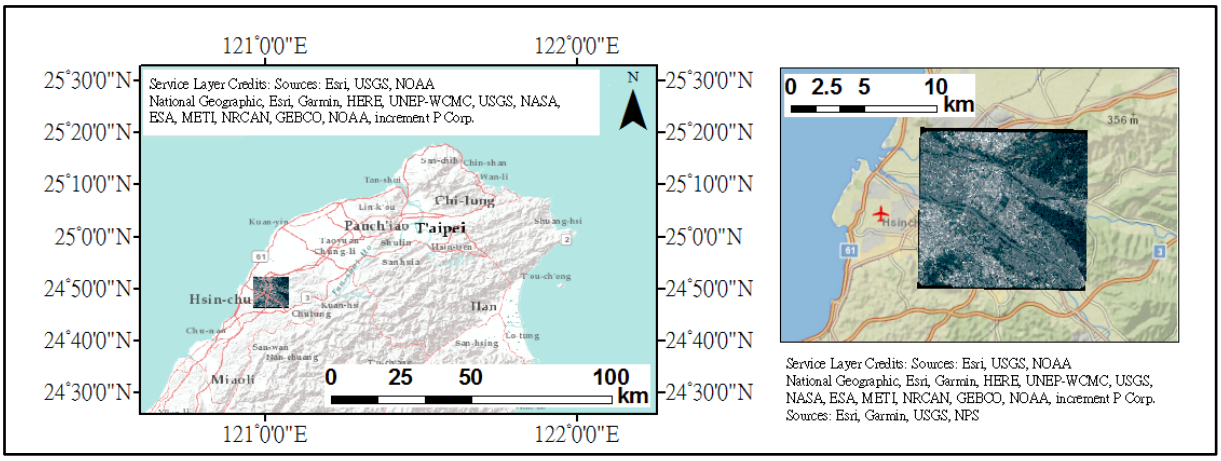


Figure 1. Area covered by VHR satellite imagery.

The study area is located in the north-west part of the island. Pleiades VHR satellite imagery capture parts of the city and county of Hsinchu (figure 1). One Pleiades image can cover 122.091 square km. The dataset: consist of three images (tri-stereo) (table 2). The city of Hsinchu is characterized by water streams, buildings and a dense street network. The urban area is in the east surrounded by a chain of hills and in the west by the sea.

2.2 Pleiades VHR Satellite Images

Satellite based remote sensing is among other sensors able to capture a large area. This study uses Pleiades tri stereo pair from Airbus (figure 2).

Pleiades Tri-Stereo			
Parameter		**Value**	
Res. MS [m]		2.8	
Sequence MS		B, G, R, NIR	
B, Blue λ [nm]		450-530	
G, Green λ [nm]		510-590	
R, Red λ [nm]		620-700	
NIR λ [nm]		775-915	
Img 1 (70)		Nadir	
Img 2 (74)		Forward	
Img 3 (93)		Backward	
	Img 1	**Img 2**	**Img 3**
IAalT	6.98	-10.93	14.17
IAacT	10.55	14.27	9.04
IA [◦]	12.57	17.71	16.62
Size [pixel]	5563 x 5285	5364 x 5228	5499 x 5189

Table 1. Pleiades VHR Satellite imagery, Airbus. With the incidence angle along track (IAalT), incidence angle across track (IAacT), incidence angle. Dataset.

The benefits of tri-stereo pairs are visualized in figure 2. It emphasises that in an urban area the coverage is more reliable. In addition, this effect can be beneficial in mountainous area as well. Further technical data, like incidence angle and image size are summarized in table 1. Pleiades tri stereo capture 4 days off-nadir with a resolution of 0.7 m Panchromatic and 2.8 m of the Multispectral bands (Panagiotakis et al., 2018).

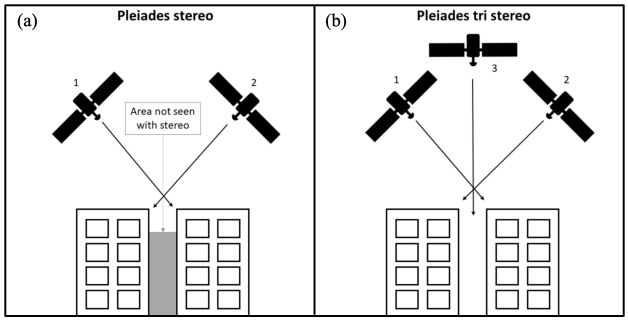


Figure 2. Pleiades VHR (a) stereo and (b) tri stereo satellite imagery (Panagiotakis et al., 2018).

AIRBUS DS GEO, Pleiades 1B	
Nadir	2019-08-27T 09:18:44.807
Forward	2019-08-27T 08:43:17.817
Backward	2019-08-27T 08:47:06.900

Table 2. Reference and production date of applied Pleiades Dataset.

3. FRAMEWORK DESIGN

The proposed framework consists of four major steps. Firstly, the segmentation with respect to the reflection of light. Secondly, the classification with respect to the material. For this study, green vegetation, water and the image frame. Thirdly, mask out the previous dynamic and unstable areas. Finale step, establishing the image correspondence based on SURF features and matching with FLANN.

3.1 Workflow

The material, introduced in the previous section, consists of the multispectral bands including the red, green, blue, and near-infrared band (figure 3).

The first step, segmentation, divides the image pixelwise with the focus of the indexes on vegetation and water. So that, the within the classification step vegetation, water and the image frame can be singled out. These three classes are added of, to one class characterised as unstable and dynamic area and mask out. The finale step is the actual stereo matching detection and description of SURF features including removal by class membership, followed by FLANN based matching.

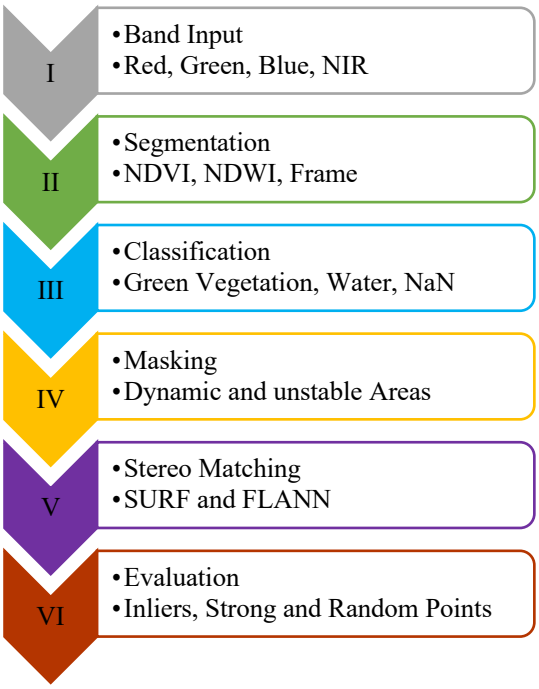


Figure 3. Workflow segmentation driven stereo matching.

The evaluation examines three images which generate six stereopairs. Firstly, case study assessment of the three classes generated in step II, frame, vegetation and water. It is based on the match or miss match of the 50 strongest points. Secondly, the deduction of the points caused by segmentation & classification. Thirdly and finally, the number of good matches, inliers and miss matches originally (no segmentation) in contrast to segmentation driven stereo matching.

3.2 Segmentation

The segmentation utilizes the red, green and NIR band of Pleiades VHR satellite imagery (table 1). The red band, as presented in figure 4, show a small response on the green vegetation graph. In contrast, the green vegetation graphs show a great positive respond in the NIR band. On this fundamental base Normalized Differential Vegetation Index (NDVI) (Kogan, 1995; Tarpley et al., 1984) can be describe like in (1).

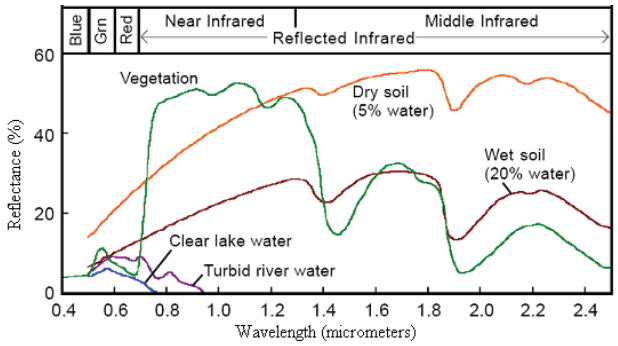


Figure 4. Surface Reflection of Light with respect to natural environment (Mercan and Alam, 2011).

Like NDVI, Normalized Differential Water Index (NDWI) was derived by Gao (1996) It focuses on the water content of vegetation where as NDVI concentrates on Chlorophyll content in vegetation (Gao, 1996). In contrast, McFeeters (1996) utilizes the green band (2) instead of SWIR and detect with this changes water bodies for Landsat satellite (McFeeters, 1996). In addition, if the SWIR band is available Xu (2006) published a modified NDWI which obtain a better classification result (Xu, 2006). For this study NDVI and NDWI are obtained by the following equation's:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2)$$

where NIR = near Infrared band
Red = red band
Green = green band

As a result, NDVI can be used to identify healthy green vegetation and NDWI clean, clear lake water.

3.3 Classification

The previous step performed a segmentation focusing on vegetation and water. In order to classify vegetation and water are thresholds within the range of NDVI and NDWI applied. The values of both, NDVI and NDWI range between -1 and 1. Firstly, mark out NaN regions, within the NDVI result. There are characterized as a lack of data, here, as expected, the image frame. Then, NDVI >= 0.72 are consider vegetated regions (Kavak et al., 2015; Weier and Herring, 2000).

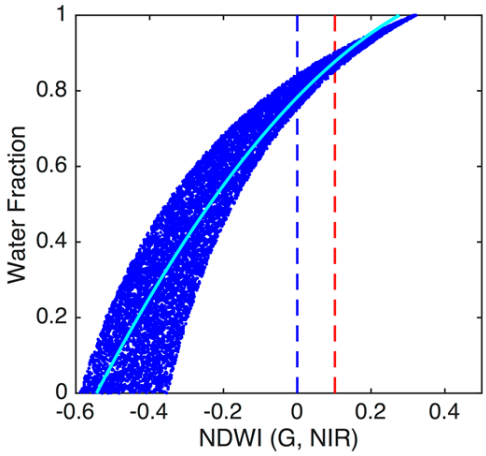
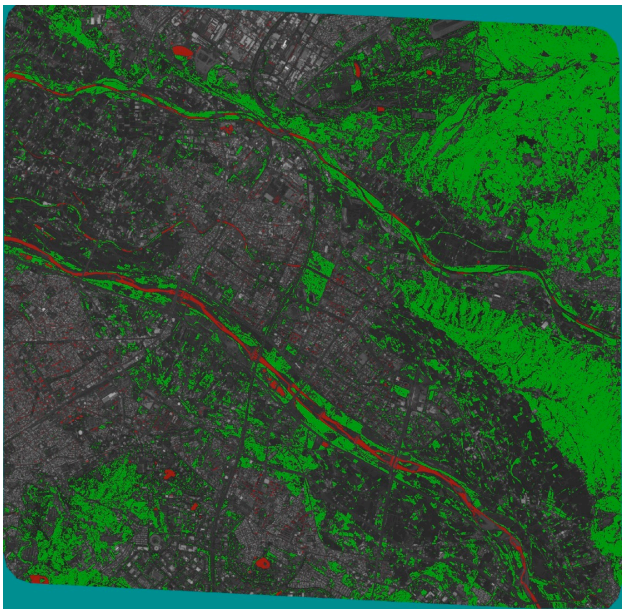


Figure 5. NDWI regression of the band combination: Green and near infrared. Zero threshold (blue dash) and Otsu's threshold (red dash) illustrated (Niroumand-Jadidi and Vitti, 2017).

The classification equation (2) of NDWI utilizes, for this study, the green band and NIR band of Pleiades VHR satellite. The suitable threshold of NDWI, is determine by resolution, target and study area including some experimental fine tuning so that, NDWI >= 0.15 (figure 5).

The finale classification results (figure 6) consist, as described earlier, of three classes. The majority of marked out pixels belong to the image frame and vegetation class. Water can be described as minor.



Legend: Image Frame, Vegetation, Water

Figure 6. Pleiades VHR satellite imagery, Gray RGB with marked out Frame, Vegetation and Water.

3.4 Masking

This step carries out the masking of the detected features based on their class. If the NDVI or NDWI value meet the criteria of the threshold applied, then SURF feature point was removed. The deduction of the feature point is described within the result section, table 2.

3.5 Stereo Matching

Image matching, the process of relating two or more images to each other. This is for many applications essential i.e., Structure from Motion, Image Stitching, SLAM etc. (Bian et al., 2017).

Figure 7. Image Matching Process.

The fundamental steps of Feature-Based Matching (FBM) are summarized in figure 7. Features are points of interest, edges or regions. Feature points are categorized in local and global. This study focuses on local feature points (Hadi and Khalaf, 2021). Therefore, SURF (Speeded Up Robust Features) has been selected for feature extraction and description (Bay et al., 2006) and FLANN (Fast Approximate Nearest Neighbours) as matching algorithm (Muja and Lowe, 2009).

The feature extraction with SURF feature detector applies the determinant of Hessian matrix in order to determine location and scale. Further, the convolution with the box filter (figure 8) can be calculated fast due to parallel computing. The feature description is accomplished on Haar wavelet descriptor of the intensity values of the feature point and their neighbourhood.

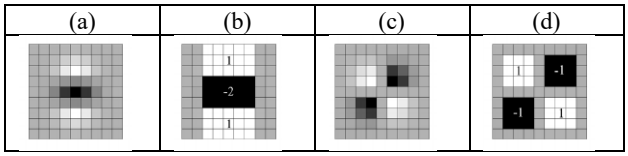


Figure 8. Example of Laplacian of Gaussian (LoG) box filter (Bay et al., 2006).

Afterwards, matching with FLANN, which applies kmeans instead of Euclidian distance or any kind of distance calculation. The clustering of a hierarchical kmeans is more suitable for extensive amount of data, like in this study.

4. RESULT

The evaluation is separated into three parts. Firstly, a comparison, with and without segmentation, over the number of the detected features. Secondly, the case study about the influence of the segmentation process over the three classes. Thirdly, the matching result with good matches, miss matches and inliers.

The total amount of features (TF) detected in the original images varies around 110’183 points (table 3). In contrast, the number of features decreases on average to 95’078 points. As a result, the segmentation on average decreases the number of points detected by 13.71 %.

	Img1	Img2	Img3	AVG
TF	**117’342**	**103’108**	**110’098**	**110’183**
SF	**94’643**	**90’006**	**100’585**	**95’078**
Diff%	**-19.34**	**-12.71**	**-8.64**	**-13.71**

Table 3. The features detection with and without segmentation. The number of features detected by SURF initially (total amount of features, TF) and after segmentation (SF). AVG, average. Diff%, difference in percent.

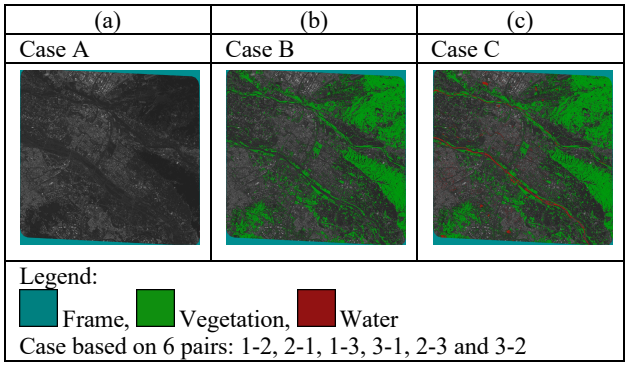


Figure 9. Case Study. Case A, frame. Case B, frame and vegetation. Case C, frame, vegetation and water.

The case study, consists, as described, of three cases (figure 9). In addition, every case consists of six stereo pairs and the finale outcome as presented in table 4, averaged (AVG). Case A, is the standard case and the reference for a comprehensive comparison. It demonstrates that roughly 50 % of the points got matched and that 25 % of these matches are miss matches. That indicates a need of adjustment, optimization and change.
Case B matches 58 % with a miss match rate of 16.2 %. Longer processing time with a lower miss match rate.
In contrast, case C reaches the same number of miss matches as case A, but with over 20 % less matches (less processing time).

Further evaluation of Case A in comparison to Case C is summarized in table 4. Firstly, shown the number of matches and the number of miss matches. In addition, the number of inliers based on Homography matrix estimation summarized (table 5).

The first column of table 4 show that the segmentation deducts the number of possible matches roughly by 15 %. In addition, the original case achieves ~3% more inliers as the segmented case. However, the segmented case achieves a lower miss match ratio of 7.5 % (table 5). The miss matches are manually and visible evaluated based on 20 random selected matches.

	Case A	Case B	Case C
AVG Match	**26.6 (53.3)**	**29 (58)**	**15.4 (30.8)**
AVG MMatch	**5.2 (24.4)**	**3.8 (16.2)**	**3.2 (24.4)**

Table 4. Average (Avg) of matches from 50 strongest points and averaged (AVG) of the miss matches (MMatch). Absolute value with (relative value, %) for each case respectively.

	Match	Inlier	MMatch
Ori	**110'182**	**46'491**	**11**
TF ratio%	**100**	**42.19**	**55**
Seg	**95'078**	**37'555**	**9.5**
SF ratio %	**86.29**	**39.5**	**47.5**

Table 5. Matching result of the Segmented (Seg) case and Original (Ori). Match, total number of matches possible. Inlier, based on Homography matrix estimation (incl. Ransac). MMatch, number of miss matches based on 20 random selected matches. TF ratio and SF ratio are based on AVG of TF (total amount of features) and SF (number of features after segmentation).

5. CONCLUSION

This study has successfully explored the potential of segmentation driven stereo matching for VHR satellite imagery. Firstly, the case study (Cases A, B, C) from 6 stereo pairs of Pleiades VRH image pairs were performed using the proposed method and clearly proofs that the segmentation of the image has an postive influence on the matching results. In addition, the reduction of features due to segmentation (14%) decreases processing time. Finally, the number of miss-matches can be reduced by 7.5 % as well.

The second part of the evaluation applied the 50 strongest points, whereas the third part applied 20 random selected matches. Ground truth of any kind would make the findings more reliable. Further, the study area is in the subtropics, so that a seasonal influence can be consider as minor. In addition, the images chosen due to the selected satellite, on the same day, roughly at same time, can be consider as less challenging as well.

Furthermore, NDWI rely on clear lake water. As soon, as the water is polluted, i.e. waste or algae, NDWI will fail to detect the water body. In addition, the speed of river flow, occlusions from bridges or and the occupation as infrastructure by ship etc. will lead to misclassification.
Moreover, NDVI rely on the chlorophyl of the vegetation, the greenness and healthiness of vegetation. As soon as vegetation is i.e. blooming or suffering of drought, NDVI will fail to detect those vegetation.

REFERENCES

Bay, H., Tuytelaars, T., Gool, L.V., 2006. SURF: Speeded Up Robust Features, in: Zurich, E., Leuven, K.U. (Eds.), switzerland.

Bian, J., Zhang, L., Liu, Y., Lin, W.-Y., Cheng, M.-M., Reid, I.D., 2017. Image Matching: An Application-oriented Benchmark. arXiv: Computer Vision and Pattern Recognition.

Chen, Q., Cheng, Q., Wang, J., Du, M., Zhou, L., Liu, Y., 2021. Identification and Evaluation of Urban Construction Waste with VHR Remote Sensing Using Multi-Feature Analysis and a Hierarchical Segmentation Method. Remote Sensing 13.

Chen, Y., Ming, D., Zhao, L., Lv, B., Zhou, K., Qing, Y., 2018. Review on High Spatial Resolution Remote Sensing Image Segmentation Evaluation. Photogrammetric Engineering & Remote Sensing.

Fisher, A., Danaher, T., 2013. A Water Index for SPOT5 HRG Satellite Imagery, New South Wales, Australia, Determined by Linear Discriminant Analysis. Remote Sensing 5, 5907-5925.

Gao, B.-c., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58, 257-266.

Haboudane, D., Miller, J., Pattey, E., Zarco-Tejada, P., Strachan, I., 2004. Hyperspectral vegetation indices and Novel Algorithms for Predicting Green LAI of crop canopies: Modeling and Validation in the Context of Precision Agriculture. Remote Sensing of Environment 90, 337-352.

Hadi, A., Khalaf, A., 2021. Evaluation of Stereo Images Matching. E3S Web of Conferences 318 (2021), 11.

Kavak, M., Karadoğan, S., Ozdemir, G., 2015. Hevsel Bahçelerinin NDVI Değerlerinin Uzaktan Algılama Teknikleri Kullanarak Uzun Dönem İçin İncelenmesi (A long Term NDVI Investigation of Hevsel Gardens Using Remote Sensing Techniques), p. 7.

Kogan, F.N., 1995. Droughts of the Late 1980s in the United States as Derived from NOAA Polar-Orbiting Satellite Data. Bulletin of the American Meteorological Society 76, 655-668.

Lamonaca, A., Corona, P., Barbati, A., 2008. Exploring forest structural complexity by multi-scale segmentation of VHR imagery. Remote Sensing of Environment 112, 2839-2849.

McFeeters, S., 1996. The Use of Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. International Journal of Remote Sensing - INT J REMOTE SENS 17, 1425-1432.

Mercan, S., Alam, M., 2011. Anomaly detection in hyperspectral imagery using Stable Distribution. Proceedings of SPIE - The International Society for Optical Engineering 8049.

Muja, M., Lowe, D., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration, Proceedings of the Fourth International Conference on Computer Vision Theory and Applications,, Lisboa, Portugal, pp. 331-340.

Niroumand-Jadidi, M., Vitti, A., 2017. Reconstruction of River Boundaries at Sub-Pixel Resolution: Estimation and Spatial

Allocation of Water Fractions. ISPRS International Journal of Geo-Information 6, 383.

Panagiotakis, E., Chrysoulakis, N., Charalampopoulou, V., Poursanidis, D., 2018. Validation of Pleiades Tri-Stereo DSM in Urban Areas. International Journal of Geo-Information 7.

Tarpley, J.D., Schneider, S.R., Money, R., 1984. Global vegetation indices from the NOAA-7 meteorological satellite.

Tucker, C.J., 1978. A comparison of satellite sensor bands for vegetation monitoring. Photogrammetric Engineering & Remote Sensing v. 44, 1369 - 1380.

Weier, J., Herring, D., 2000. Measuring Vegetation (NDVI & EVI) : Feature Articles.

Xu, H., 2006. Modification of Normalized Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. International Journal of Remote Sensing 27, 3025–3033.

Xu, Z., Su, C., Zhang, X., 2021. A semantic segmentation method with category boundary for Land Use and Land Cover (LULC) mapping of Very-High Resolution (VHR) remote sensing image. International journal of remote sensing 2021 v.42 no.8, pp. 3146-3165.

Yildirim, I., Demirtas, F., Gulmez, B., Leloglu, U., Yaman, M., Güneyi, E., 2019. COMPARISON OF IMAGE MATCHING ALGORITHMS ON SATELLITE IMAGES TAKEN IN DIFFERENT SEASONS, Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birligi X, Turkey.