

# EXPERIENCES WITH USING SIFT FOR MULTIPLE IMAGE DOMAIN MATCHING

**Charles K. Toth**, Senior Research Scientist  
**Hui Ju**, Graduate Student  
**Dorota A. Grejner-Brzezinska**, Professor  
The Center for Mapping  
The Ohio State University  
Columbus, OH 43210  
[toth@cfm.ohio-state.edu](mailto:toth@cfm.ohio-state.edu)

## ABSTRACT

The paper reports about investigations into the utilization of the SIFT algorithm to support image matching between different image domains. The Scale-Invariant Feature Transformation, proposed by Lowe in 1999, is a highly robust technique that has been widely used in the computer vision community. Though, SIFT is known in mapping circles, so far its use is rather limited. The objective of our study is to assess the performance of SIFT when it is applied to imagery acquired by different sensors and on different platforms. For testing, four image datasets from different sensors were considered, including airborne and satellite imagery, and LiDAR intensity and elevation. The image co-registration was performed based on SIFT features. The preliminary results indicate mixed but encouraging performance, as the in-domain matching based registration generally works well, while the matching between images coming from different domains (such as airborne and LiDAR intensity) usually produces modest results. In all cases, we used RANSAC to remove outliers. The above behavior could be associated with two facts: first, the style of SIFT features extracted from the imagery is correlated to the image type, and second, the number of actually matched features is significantly lower than that of the in-domain imagery. In summary, our present level of research indicates that the SIFT algorithm has substantial capacity to support image matching between different domains, and thus appears to be an efficient technique for co-registering image data acquired by different sensors.

## INTRODUCTION

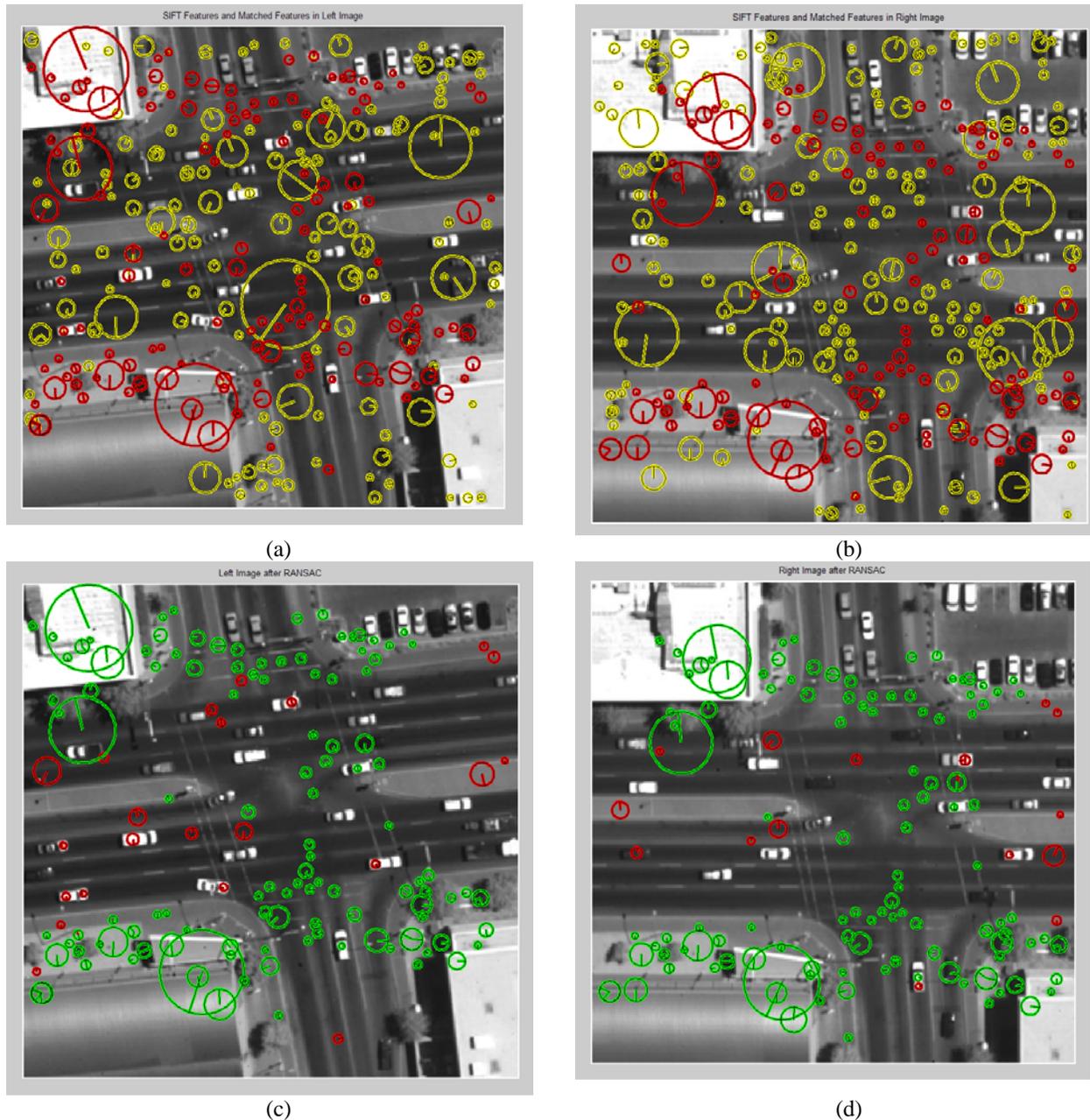
The Scale Invariant Feature Transform (SIFT) matching is a technique to extract highly invariant features from images and to perform reliable matching; a thorough description on SIFT can be found in (Lowe, 2004). To achieve robust performance, the features for image matching should be generally invariant to scale, rotation, affine distortion and intensity changes to handle varying imaging conditions. SIFT algorithm consists of several stages and the computation requirements are quite substantial; in particular, the matching of the extracted features could be a challenge for larger number of features, as it is based on k-D tree structure. There are also some modifications to SIFT to make it more effective: PCA-SIFT (Ke and Sukthankar, 2004), GLOH (Gradient Location-Orientation Histogram) (Mikolajczyk and Schmid, 2005), CSIFT (Abdel-Hakim and Farag, 2006), SR-SIFT (Yi *et al.*, 2008), SURF (Speeded-Up Robust Features) (Bay *et al.*, 2008) and Robust SIFT (Li *et al.*, 2009). In our study, the baseline SIFT implementation was used.

Image matching has been perfected in digital photogrammetry in the past twenty years and state-of-the-art softcopy systems can easily deliver sub-pixel matching accuracy for large-format airborne (vertical) imagery for almost any object space conditions. So the question is why SIFT is of any interest, when SIFT, at best, can only produce few pixels of matching accuracy? The answer is that because of its robustness, SIFT is a potential candidate for matching images of different domains. As sensor technology advances, an increasing volume of multi- and hyperspectral, LiDAR and SAR imagery is acquired from spaceborne, airborne and mobile platforms. In addition to multiple layers of new imagery, there are large volumes of existing geographic data, including orthoimagery, DEMs, etc. Ideally, if all the layers are accurately georeferenced, there should be no difficulty to co-register the various image layers, but, in reality, it is hardly ever the case. Therefore, there is a strong demand for methods to co-register imagery obtained from a variety of sensors (based on different geometrical models) at different times.

This paper investigates the performance of SIFT when applied to airborne, satellite and LiDAR imagery. In-domain and between domain matching are performed for overlapping image pairs. The images were selected to represent typical image characteristics of nowadays airborne surveying practice. There are many aspects of SIFT feature-based matching in terms of parameterization, but in this study, only a limited analysis is provided.

## SIFT MATCHING

In a simplified description, SIFT is based on two main processes: extracting the SIFT features from an image and then matching SIFT features extracted from different images. The SIFT features are described by four parameters, including the two image coordinates, the orientation and strength of the feature, and a 128-dimensional feature descriptor. Figures 1a-b show two images captured over an intersection, where the radius of the circles is proportional to the strength of the feature and the hand in the circle represents the orientation of the feature. The SIFT feature descriptor, not displayed, is used for matching; successfully matched features are marked by red color in the figures. Since there are mismatched SIFT features, blunder detection, based on RANSAC, (RANDOM SAMPLE CONSENSUS) (Fischler and Bolles, 1981), is applied. In Figures 1c-d, red and green mark the removed and kept matched SIFT features, respectively.



**Figure 1.** Aerial image pair with SIFT features extracted (yellow), matched (red), and kept (green).

## MATCHING AERIAL IMAGERY

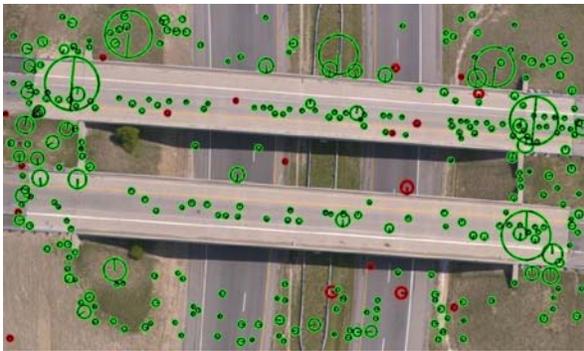
A section of a high-resolution digital camera image pair is shown in Figures 2a-b. The extracted and cleaned matched SIFT features are shown in Figures 2c-d. Figures 2e-f depict the distribution of the errors and its histogram, respectively; note that errors were computed based on a 2D affine transformation. The results, 3.5 pixel matching performance ( $1\sigma_{xy}$ ) could be considered typical. Note that there is a considerable difference between two directions, mostly likely due to the different optical resolution of the image in track and cross-track directions; the errors in the y direction about five times larger than the x direction.



(a)



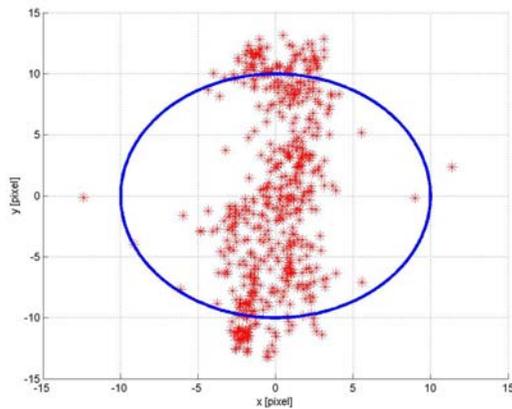
(b)



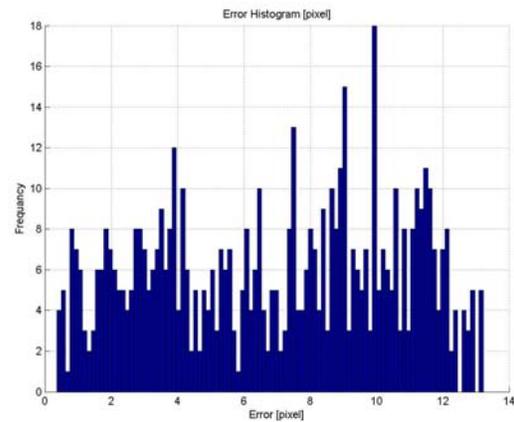
(c)



(d)



(e)

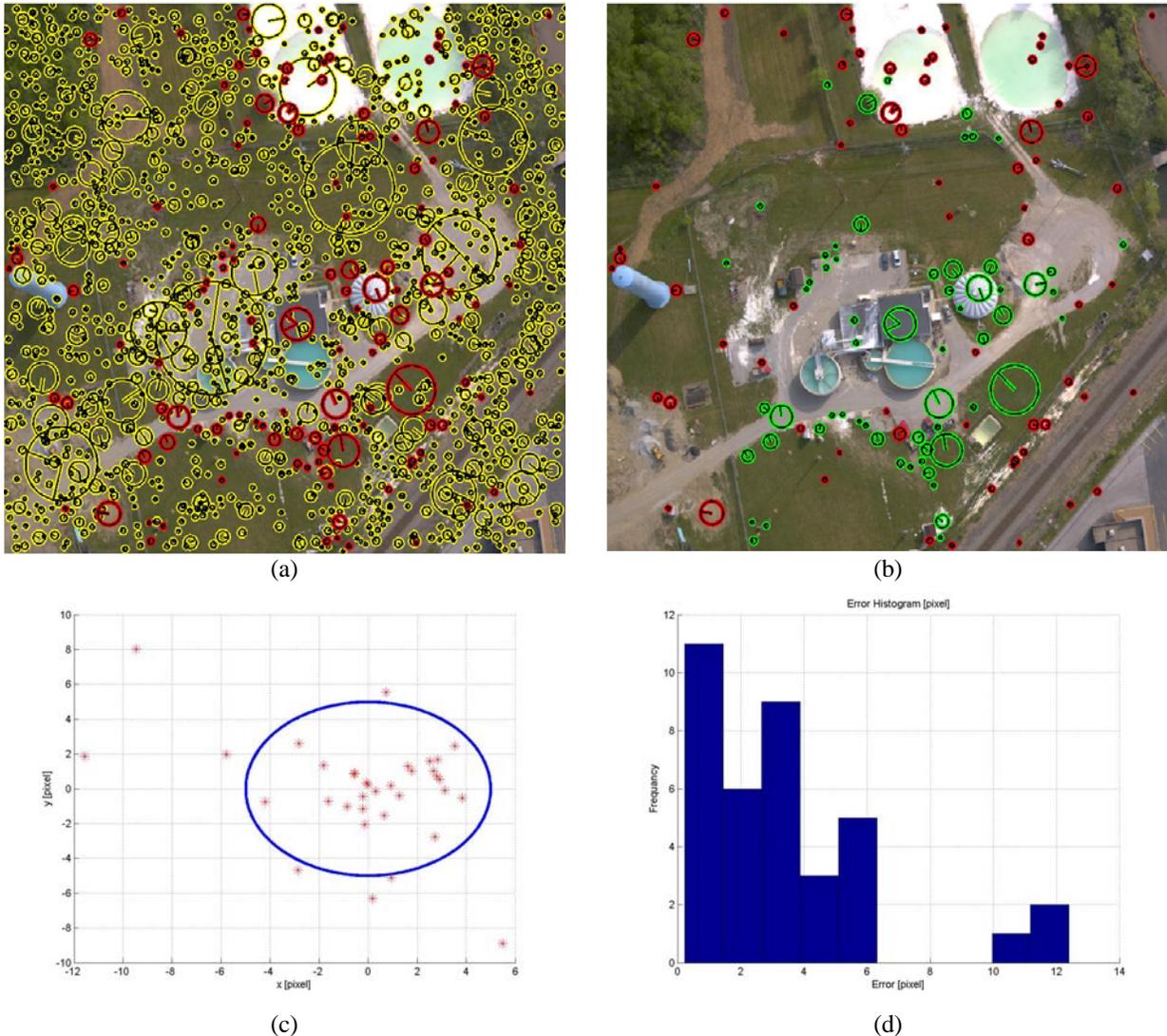


(f)

**Figure 2.** High-resolution digital camera image pair (a-b) with matched SIFT features (red), and kept (green) (c-d), and error distribution (e) and histogram (f).

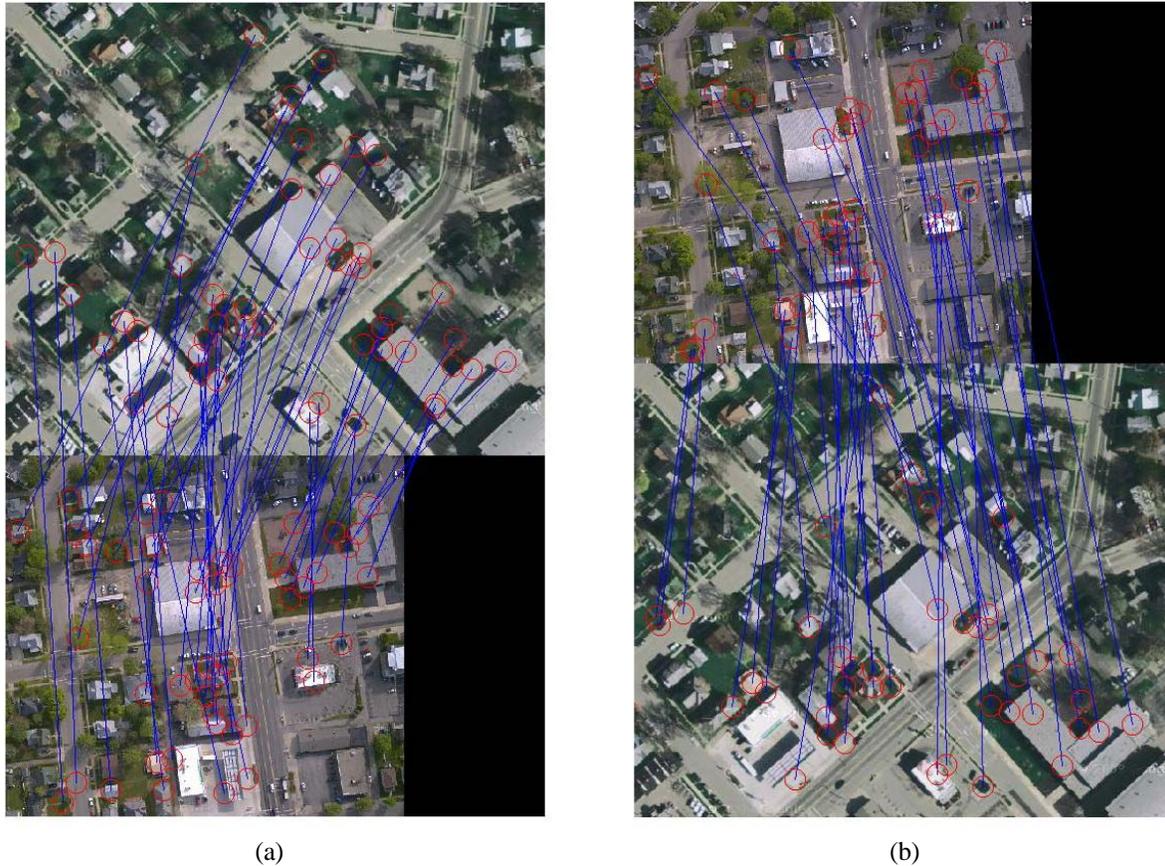
## MATCHING SATELLITE AND AERIAL IMAGERY

A QuickBird and aerial image pair of an industrial area is shown in Figures 3a-b. Compared to the aerial matching case, the number of SIFT features extracted from the satellite image are similar but the number of matched features is significantly less and, alike, the number of matched features after the blunder detection is also relatively smaller than that of the aerial matching. Yet from a co-registration perspective, the number of matched SIFT features, and their distribution, is more than enough to establish a transformation between the two images. Figures 3c-d show the distribution of the errors and its histogram, respectively; note that errors were computed based on a 2D affine transformation. The results, 3 pixel matching performance ( $1\sigma_{xy}$ ) are comparable to the aerial matching case.



**Figure 3.** Satellite and aerial camera image pair with matched SIFT features (a-b), error distribution (c) and histogram (d).

One important characteristics of SIFT matching is that the SIFT descriptor-based matching lacks the commutative property; in other words, matching image 1 to 2 may not produce the same matched SIFT features as matching 2 to 1. Figure 4a-b show a QuickBird and aerial image pair of a residential area to demonstrate the differences in matching results for satellite image to aerial image and aerial image to satellite image matching, respectively.

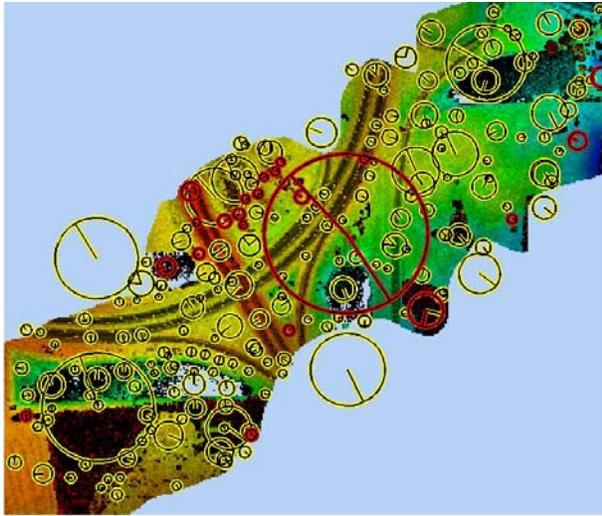


**Figure 4.** Satellite and aerial camera image pair with matched SIFT features, satellite to aerial (a) and aerial to satellite (b).

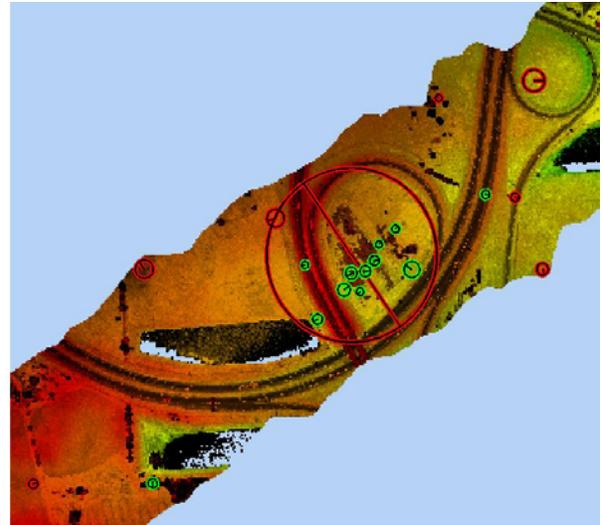
## MATCHING LIDAR IMAGERY

LiDAR data, including the point cloud (elevation) and intensity, come in irregular spacing, and thus, SIFT is not directly applicable. However, for visualization, data distribution, etc., LiDAR data are frequently converted (interpolated) to a regular grid. The impact of this conversion is ignored in our study, and only regularly spaced data were considered. Figure 5a-b show a LiDAR elevation strip pair, with overlaying the extracted and matched, and then the matched and cleaned SIFT features, respectively. The number SIFT features per area is noticeably less than that of the aerial and satellite imagery. It is interesting, as the visual appearance of the elevation image is close to the optical images used earlier. There is also a significantly lower number of matched SIFT features; however, the relative number of SIFT outliers is similar to the optical image cases. It is likely that using advanced conversions for gridding and/or filtering the elevation data could improve the SIFT performance. Figures 5c-d show the distribution of the errors and its histogram, respectively; note that errors were computed based on a 2D affine transformation.

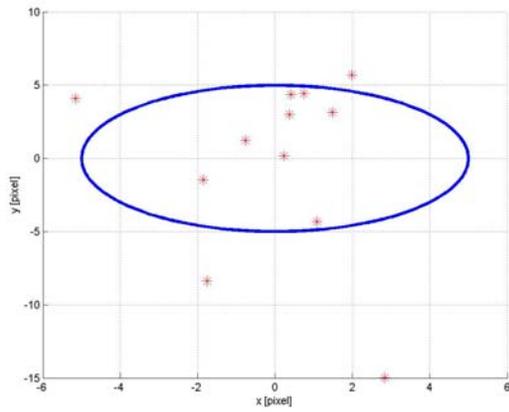
The LiDAR intensity data closely resembles panchromatic optical imagery; strictly speaking it is a narrow band image of the near-IR spectrum (at 1,064 nm). Therefore, a better SIFT matching performance is expected. Figures 6a-b show two LiDAR strips acquired over the same intersection, overlaying the extracted (yellow) and matched (red) features and then matched (red) and cleaned (green) SIFT features, respectively. Despite the noisy look of the intensity images, the number of SIFT features extracted as well as the number matched and kept features is high, indicating a good matching performance. Figures 6c-d show the distribution of the errors and its histogram, respectively; note that errors were computed based on a 2D affine transformation. The results, 4 pixel matching performance ( $1\sigma_{xy}$ ) are just slightly worse than that of the aerial to aerial matching case.



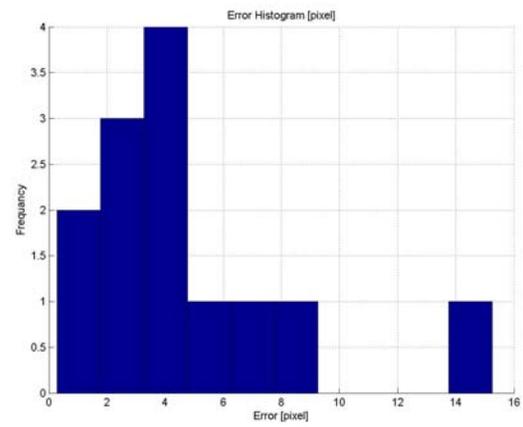
(a)



(b)

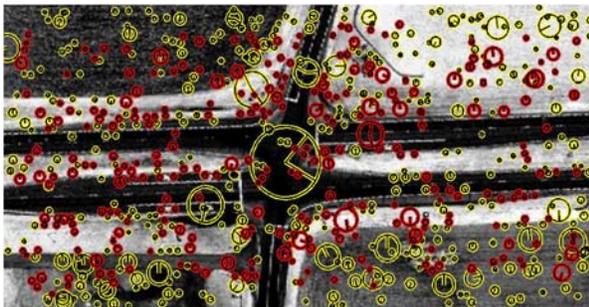


(c)

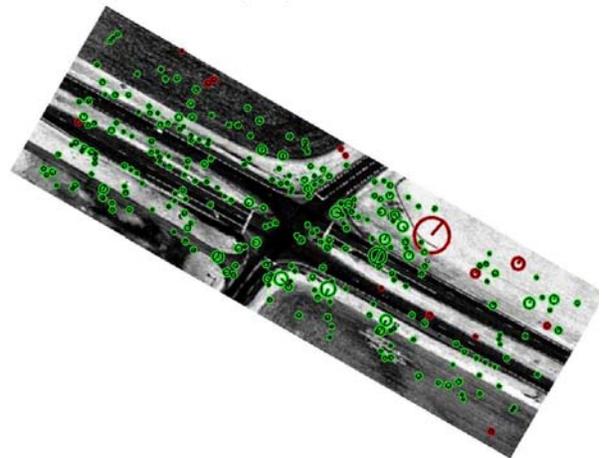


(d)

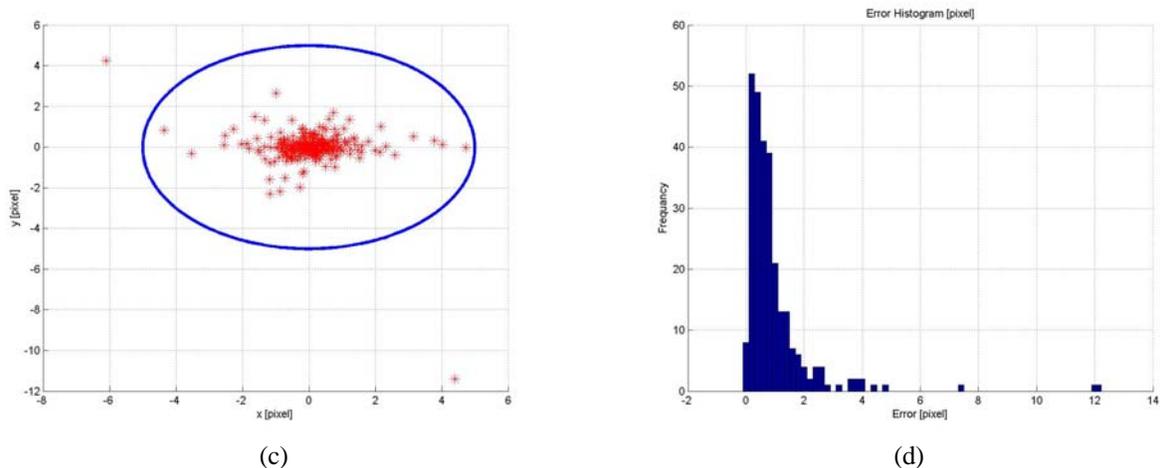
**Figure 5.** LiDAR strip pair (elevation) with matched SIFT features (a), cleaned SIFT features (b), error distribution (c) and histogram (d).



(a)



(b)



**Figure 6.** LiDAR strip pair (intensity) with matched SIFT features (a), cleaned SIFT features (b), error distribution (c) and histogram (d).

## CONCLUSIONS

The applicability of the SIFT feature-based matching for different image sensor data, acquired from airborne and satellite platforms, was studied in this paper. This preliminary investigation mainly focused on the in-domain matching performance, and only the matching between aerial and satellite imagery was discussed. The initial experiences are encouraging. The SIFT features are rather invariant to image distortions, and thus, can efficiently handle different sensor models, such as frame and push-broom camera models, or orthorectified images in any combinations. The SIFT matching for LiDAR data also showed a potential for both in-domain and between domain matching, though the performance was significantly lower for the elevation data compared to the intensity and optical imagery cases. The investigation also revealed that there is no direct correlation between SIFT feature strength and its matching value. In other words, a strong feature doesn't guarantee a successful match and the same time weak features can make good matches. This also confirms that there is no simple measure to characterize a SIFT feature in terms of matching performance; i.e., the 128-dimensional descriptor cannot be replaced by a merit number. In SIFT matching, there is a non-negligible number of incorrect matches, thus a geometry-based outlier removal is essential to clean the matches. In our testing, RANSAC showed a good performance remove mismatched features.

In summary, our present level of research indicates that the SIFT algorithm has substantial capacity to support image matching between different domains, and thus appears to be an efficient technique for co-registering image data acquired by different sensors.

This research effort continues to assess the feasibility and the performance potential of SIFT for matching images of different domain, such as LiDAR intensity with airborne and satellite imagery or LiDAR elevation data to any optical imagery, including airborne, satellite and orthoimagery. The significance of co-registering a variety of imagery is that there are several areas where this step is a prerequisite for any downstream processing. For example, the need for geographical data update is rapidly going and the update process needs an initial, preferably good, co-registration of past and current data. In addition, since better resolution satellite imagery, in terms of higher spatial resolution, shorter revisit time, and improving georeferencing accuracy, is increasingly becoming available, the satellite data could be used as inexpensive ground control to georeference airborne imagery that can be used for both mapping and terrain-based navigation worldwide; in particular, in areas, which lack a geodetic infrastructure.

## ACKNOWLEDGEMENT

The authors thank the Office of Aerial Engineering at the Ohio Department of Transportation for providing the data for this research.

## REFERENCES

- Abdel-Hakim, A.E., and A.A. Farag, 2006. CSIFT: A SIFT Descriptor with Color Invariant Characteristics, In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.1978-1983.
- Bay, H., A. Ess, T. Tuytelaars, L.V. Gool, 2008. SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, 110(3): 346-359.
- Brown, M. and Lowe, D.G. 2002. Invariant feature from interest point groups, In *British Machine Vision Conference*, Cardiff, Wales, pp. 656-665.
- Fischler, M. A. and R. C. Bolles, 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*. 24: 381–395
- Ke, Y and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors, In: *Proceedings International Conferences on Computer Vision*, Washington DC, 2004, pp.506-513.
- Li, Qiaoliang., Guoyou. Wang, Jianguo. Liu, and Shaobo. Chen, 2009. Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration, *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 6(2):287-291.
- Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: *Proceedings International Conferences on Computer Vision*, Corfu, Greece, pp.1150-1157.
- Lowe, D., 2004, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Volume 60, Number 2, pp. 91-110.
- Mikolajczyk, K., and C. Schmid, 2005. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, No. 10, pp.1615-1630, Oct. 2005.
- Yi, Z., C. Zhiguo, and X. Yang, 2008. Multi-spectral remote image registration based on SIFT, *Electron. Lett.*, Vol. 44, No.2, pp.107-108.