2019

Fusion of Landsat and Worldview Images

Chiman Kwan

Bryan Chou

Jerry Yang

Daniel Perez
Old Dominion University

Yuzhong Shen
Old Dominion University, yshen@odu.edu

See next page for additional authors

Original Publication Citation

Kwan, C., Chou, B., Yang, J., Perez, D., Yuzhong, S., Jiang, L., & Koperski, K. (2019). *Fusion of Landsat and Worldview images.* Paper presented at the SPIE Defense + Commercial Sensing, 15-17 April 2019, Baltimore, Maryland, April 15-17, 2019.

Fusion of Landsat and Worldview Images

Chiman Kwan*[a], Bryan Chou[a], Jerry Yang[a], Daniel Perez[b], Yuzhong Shen[b], Jiang Li[b], and Krzysztof Koperski[c]

[a]Applied Research LLC, 9605 Medical Center Dr., Rockville, MD, USA 20850; [b]Old Dominion University, Norfolk, VA; [c]Digital Globe Inc., 1300 W. 120th Ave., Westminster, CO 80234,USA.

ABSTRACT

Pansharpened Landsat images have 15 m spatial resolution with 16-day revisit periods. On the other hand, Worldview images have 0.5 m resolution after pansharpening but the revisit times are uncertain. We present some preliminary results for a challenging image fusion problem that fuses Landsat and Worldview (WV) images to yield a high temporal resolution image sequence at the same spatial resolution of WV images. Since the spatial resolution between Landsat and Worldview is 30 to 1, our preliminary results are mixed in that the objective performance metrics such as peak signal-to-noise ratio (PSNR), correlation coefficient (CC), etc. sometimes showed good fusion performance, but at other times showed poor results. This indicates that more fusion research is still needed in this niche application.

Keywords: Image fusion, Landsat, Worldview, temporal resolution, spatial resolution

1. INTRODUCTION

Worldview images have a super high resolution of 0.5 m for pansharpened images[1]. However, the revisit times are rare. For a studied area[2], there are only a handful high resolution (HR) satellite images available over a period of two years (2014 to 2015). On the other hand, Landsat images have only 15 m resolution even after pansharpening, but they are available every 16 days. It will be useful to fuse these two types of images to generate a high temporal resolution time series at the resolution of the WV images. The enhanced image will be useful for anomaly detection, target detection, and change detection[3,4].

In the past few years, there have been new developments in two groups of algorithms in image fusion. One group focuses on fusing high spatial resolution low spectral resolution images with low spatial resolution high spectral resolution images. This group is known as pansharpening. Many methods have been developed. See references[5-9] and references therein. Another group focuses on integrating low spatial resolution high temporal resolution images with high spatial resolution low temporal resolution images. See papers[10-13] and references therein. Some applications include the fusion of MODIS and Landsat[10-12], and the fusion of Planet and Worldview images[13]. In the latter group, pansharpening cannot be applied because only low spatial resolution is available at the time of prediction/fusion.

In this paper, we present some preliminary results on the fusion of Landsat and Worldview images. The proposed algorithm is known as hybrid color mapping (HCM)[11], which was developed by our team and has been applied to several interesting applications in recent years[14-20]. The basic idea is to learn a mapping between a pair of Landsat images at t_1 and t_2 and then the mapping is applied to the Worldview image at t_1 to predict the high resolution image at time t_2. The HCM algorithm is simple, efficient, parallelizable, and has comparable performance as other state-of-the-art algorithms[10,12]. We demonstrated the performance of our approach by using three sets of actual Landsat and Worldview images. Six objective performance metrics were used in our evaluations. Moreover, subjective visualization was used in our evaluations. Preliminary results show that our results are mixed. That is, in some cases, we do see improvements and in some other cases, we see worse prediction results. This means that more research is needed in this area.

This paper is organized as follows. In Section 2, we will briefly summarize the proposed fusion approach. Several objective performance metrics will be described. Section 3 presents the fusion results. Objective and subjective evaluations will be presented. Finally, concluding remarks and future research directions will be given in Section 4.

* Address all correspondence to: Chiman Kwan, chiman.kwan@signalpro.net

2. IMAGE FUSION APPROACH

2.1 HCM Prediction Approach

Figure 1 illustrates the HCM approach. Based on the available Landsat images collected at t_k and t_p, we learn the pixel by pixel mapping between the two images. The learned matrix, F, is then applied in the prediction step. The prediction of the WV image at t_p can be achieved by

$$W(x, y, t_p) = F \times W(x, y, t_k) \tag{1}$$

where $W(\bullet,\bullet,\bullet)$ denotes a pixel vector (up to K with K being the number of bands) for this application and F is a pixel to pixel mapping/transformation matrix with appropriate dimensions. F can be determined by using the following relationship:

$$P(x, y, t_p) = F \times P(x, y, t_k) \tag{2}$$

where $P(\bullet,\bullet,\bullet)$ denotes a pixel vector (K bands). To account for intensity differences between two images, a variant of Equation (4) can be described as

$$P(x, y, t_p) = F_1 \times P(x, y, t_k) + F_2 \tag{3}$$

where F_2 is a vector of constants. Procedures to obtain F can be found in reference[11].

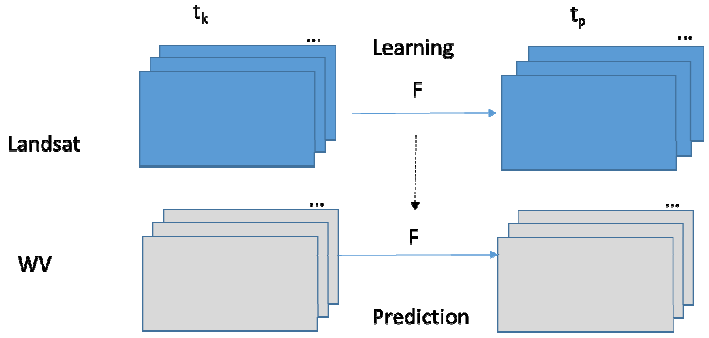


Figure 1. The Hybrid Color Mapping (HCM) approach for image fusion.

Based on our observations, in some cases, prediction results will be more accurate if we divide the images into patches. Each patch will have its own mapping matrix. Figure 2 illustrates the local prediction approach. The patches can be overlapped or non-overlapped. Moreover, for each local patch, which can be a single band or a multi-band image, we use the same estimation algorithm[11] to determine the local mapping matrix, F_i.

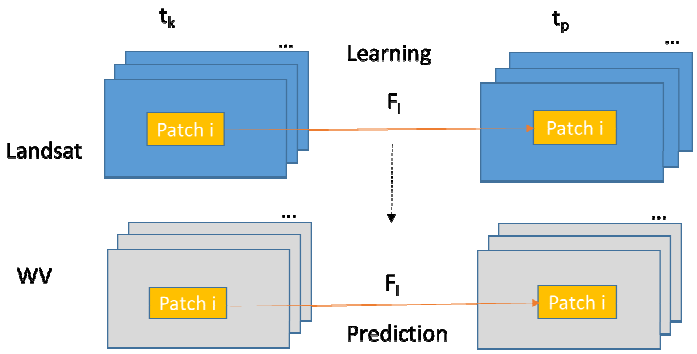


Figure 2. Proposed prediction approach based on local mapping.

2.2 Performance Metrics

Although there are many performance metrics in the literature, we selected the following ones: absolution difference (AD) [13], root mean squared error (RMSE)[13], peak signal-to-noise ratio (PSNR)[13], cross correlation (CC)[13], Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) [13], and structural similarity (SSIM) [13].

- Absolute Difference (AD). The AD of two vectorized images $\mathbf{S}$ (ground truth) and $\hat{\mathbf{S}}$ (prediction) is defined as

$$\mathrm{AD}(\mathbf{S},\hat{\mathbf{S}}) = \frac{1}{Z}\sum_{j=1}^{Z}\left|s_j - \hat{s}_j\right| \tag{4}$$

  where Z is the number of pixels in each image. The ideal value of AD is 0 if the prediction is perfect.
- RMSE (Root Mean Squared Error). The RMSE of two vectorized images $\mathbf{S}$ (ground truth) and $\hat{\mathbf{S}}$ (prediction) is defined as

$$\mathrm{RMSE}(\mathbf{S},\hat{\mathbf{S}}) = \sqrt{\frac{1}{Z}\sum_{j=1}^{Z}\left(s_j - \hat{s}_j\right)^2} \tag{5}$$

  where Z is the number of pixels in each image. The ideal value of RMSE is 0 if the prediction is perfect.
- PSNR (Peak Signal to Noise Ratio). PSNR is related to RMSE defined in (5). If the image pixels are expressed in doubles with values between 0 and 1, then

$$\mathrm{PSNR}=20\log(1/\mathrm{RMSE}(\mathbf{S},\hat{\mathbf{S}})) \tag{6}$$

- CC (Cross-Correlation). We used the codes from Open Remote Sensing website (https://openremotesensing.net/). The ideal value of CC is 1 if the prediction is perfect.
- ERGAS (Erreur Relative. Globale Adimensionnelle de Synthese). The ERGAS is defined as

$$\mathrm{EGARS}(\mathbf{S},\hat{\mathbf{S}}) = 100d\,\frac{\mathrm{RMSE}}{\mu} \tag{7}$$

  for some constant d depending on the resolution and μ is the mean of the ground truth images. The ideal value of ERGAS is 0 if a prediction algorithm is perfect.
- SSIM (Structural Similarity). This is a metric to reflect the similarity between two images. An equation for SSIM can be found in paper[10]. The ideal value of SSIM is 1 for perfect prediction. We also use the SSIM map to display the similarity values at each pixel location. Bright pixels have high similarity.

3. EXPERIMENTS

3.1 Data

In order to make our paper self-contained, we include the following specifications of Landsat and Worldview images in Table 1.

Table 1. Band specifications of Landsat and Worldview images.

Pansharpened Landsat Image (15 m Resolution) with 4 bands		Pansharpened Worldview Image (0.5 m Resolution) with 4 bands	
Blue	455 – 515 nm	Blue	450 -510 nm
Green	500 – 590 nm	Green	510 – 580 nm
Red	590 – 670 nm	Red	630 – 690 nm
NIR	780 – 860 nm	Near-IR1	770 – 895 nm

Our area of interest is an airport[2]. From data archives of both Landsat and Digital Globe, the following images were used in our study:

Landsat images: 4/25/2014, 10/27/2014, 5/23/2015, and 8/27/2015;
WV images: 4/24/2014, 10/30/2014, 5/28/2015, and 7/30/2015

It should be noted that it is difficult to retrieve Landsat and WV images for the same dates because the two satellites seldom visit the same location at the same day. However, this also justifies our research, as our goal is to generate high spatial resolution images when high resolution WV images are not available. It is also emphasized that the registration of the two different types of satellite images is non-trivial, as WV images are not taken at nadir. As a result, automated registration algorithms using corner points from buildings may lead to large registration errors at ground level pixels. In this research, we manually selected ground feature points such as road intersections for image alignment.

3.2 Results

From the above collected images, we focused on three different scenarios.

Scenario 1: 4/25/14 – 8/27/15

In this case, we used two Landsat images collected on 4/25/2014 and 8/27/2015, and two WV images collected on 4/24/2014 and 7/30/2015. The prediction scenario is summarized in Figure 3. We only show four bands because the pansharpened WV images only have four bands. We used patch sizes of 5, 10, 10, 1 for R, G, B, and NIR bands, respectively; there is no overlap between patches. Each band is predicted separately. A comparison between the ground truth, predicted image, and the Landsat image of the RGB bands is shown in Figure 4. At first glance, it appears that the predicted image seems to be better in terms of clarity and resolution. However, objective metrics in Table 2 show that the results are mixed. In some cases, we see that the predicted image performed better and in other cases, the predicted image is worse.

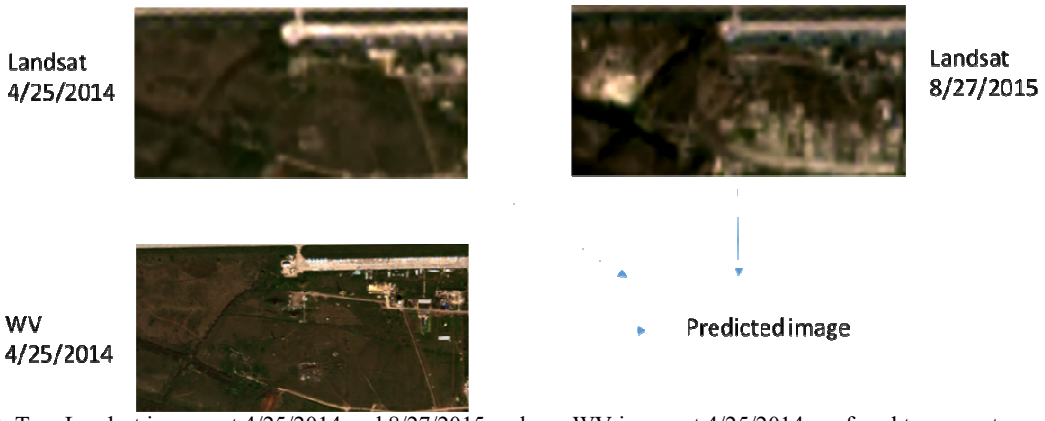
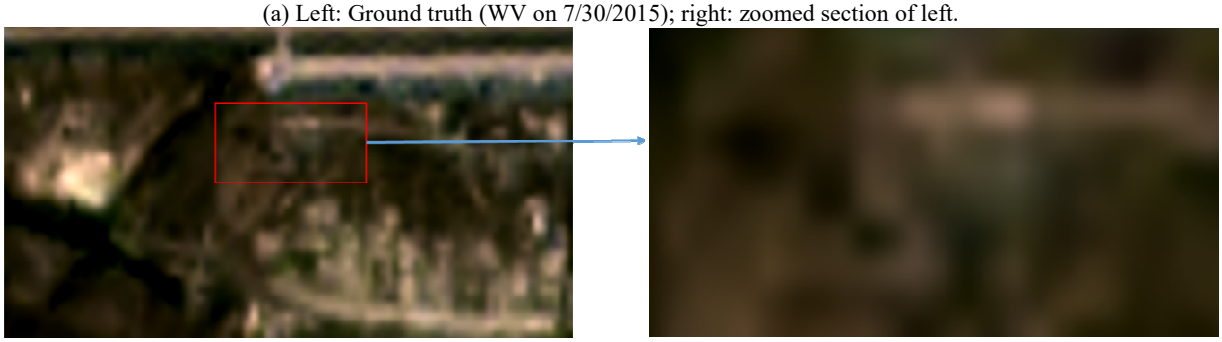
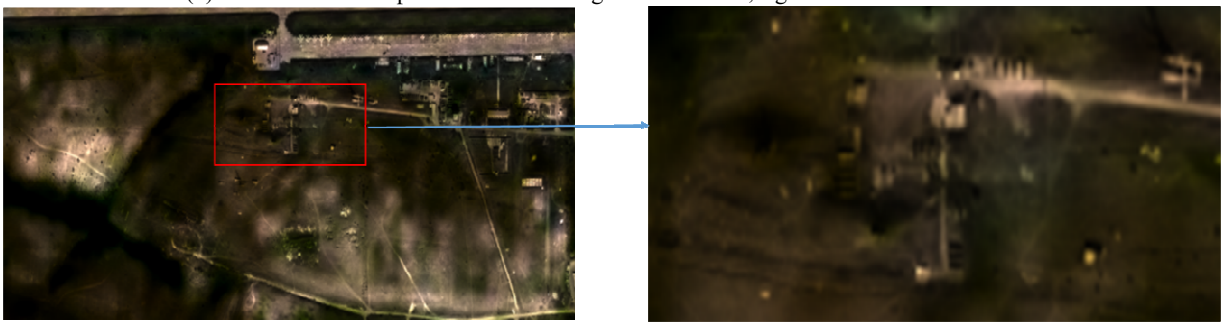


Figure 3. Two Landsat images at 4/25/2014 and 8/27/2015 and one WV image at 4/25/2014 are fused to generate a prediction. The prediction is then compared to a WV image collected on 7/30/2015.

(a) Left: Ground truth (WV on 7/30/2015); right: zoomed section of left.



(b) Left: Bicubic interpolated Landsat image on 8/27/2015; right: zoomed section of left.



(c) Left: HCM predicted image on 7/30/2015; right: zoomed section of left.

Figure 4. Comparison of different fused images with the ground truth (WV) and the low resolution Landsat image at the prediction time (7/30/2015).

Table 2: Fusion performance metrics for image pair (4/24/2014-7/30/2015).

	Fusion Red	Landsat Red	Fusion Green	Landsat Green	Fusion Blue	Landsat Blue	Fusion NIR	Landsat NIR	Fusion Avg	Landsat Avg.
AD	0.0032	0.0032	0.0029	**0.0028**	0.0029	**0.0028**	0.0073	0.0073	0.004075	**0.004025**
CC	0.7963	**0.821**	0.7555	**0.7949**	0.7327	**0.7752**	0.4627	**0.4987**	0.6868	**0.72245**
ERGAS	48.5367	**45.3185**	**45.5343**	45.8658	**73.685**	77.5776	31.3093	**29.0069**	49.76633	**49.4422**
PSNR	46.1483	**46.7442**	**47.8845**	47.8215	**48.227**	47.7799	39.7237	**40.3871**	45.49588	**45.68318**
RMSE	0.0049	**0.0046**	**0.004**	0.0041	**0.0039**	0.0041	0.0103	**0.0096**	0.005775	**0.0056**
SSIM	0.8672	**0.9557**	0.878	**0.9587**	0.8922	**0.9526**	1	1	0.90935	**0.96675**

Scenario 2: 10/27/14 – 5/23/15

In this case, we used two Landsat images collected on 10/27/2014 and 5/23/2015, and two WV images collected on 10/30/2014 and 5/28/2015. The prediction scenario is summarized in Figure 5. The patch sizes are 2, 8, 10, 1 for R, G, B, and NIR, respectively. No overlapping is used and every band is separately predicted. Similar to Scenario 1, the predicted image shown in Figure 6 appears to be good. However, some of the details are incorrect because they are transferred from the earlier WV image. The objective metrics also corroborate with the visual observations. In some cases, we see better results for predicted images; but in other cases, the predicted images are actually worse.

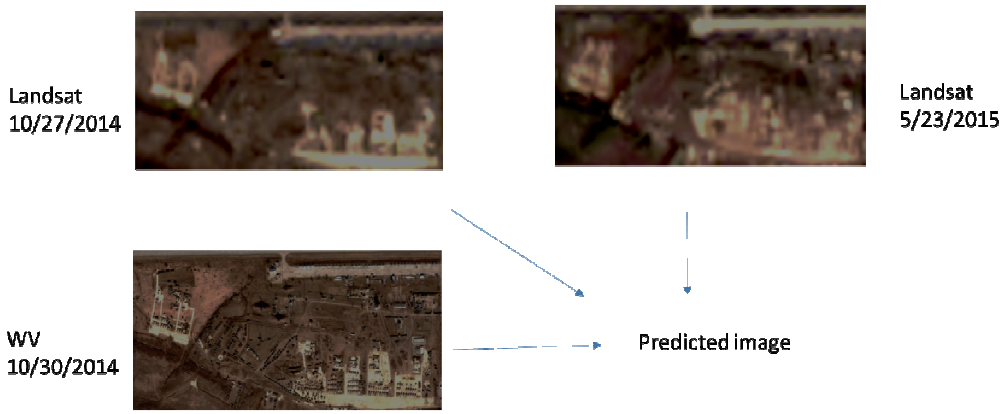
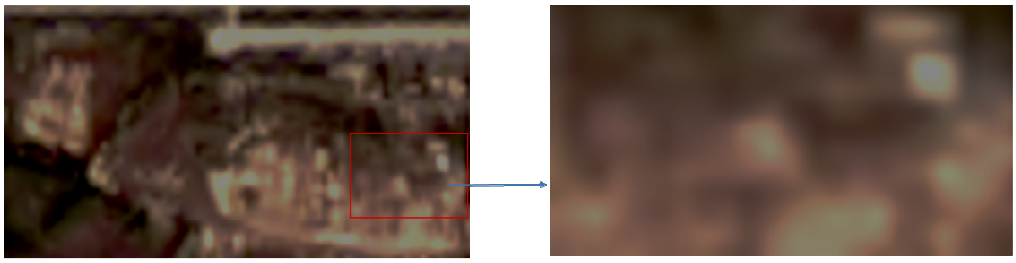


Figure 5. Two Landsat images at 10/27/2014 and 5/23/2015 and one WV image at 10/30/2014 are fused to generate a prediction. The prediction is then compared to a WV image collected on 5/28/2015.

(a) Left: Ground truth (WV on 5/28/2015); right: zoomed section of left.



(b) Left: Bicubic interpolated Landsat image on 5/23/2015 right: zoomed section of left.

(c) Left: HCM predicted image on 5/28/2015; right: zoomed section of left.

Figure 6. Comparison of different fused images with the ground truth (WV) and the low resolution Landsat image at the prediction time (5/28/2015).

Table 3. Performance metrics for Pair 2 (10/30/2014-5/28/2015).

	Fusion Red	Landsat Red	Fusion Green	Landsat Green	Fusion Blue	Landsat Blue	Fusion NIR	Landsat NIR	Fusion Avg	Landsat Avg.
AD	0.0075	**0.0062**	0.0057	**0.0053**	0.0042	**0.0041**	**0.0143**	0.0154	0.007925	**0.00775**
CC	0.6827	**0.8174**	0.6906	**0.7883**	0.7097	**0.7887**	**0.3456**	0.2799	0.60715	**0.668575**
ERGAS	56.468	**42.8349**	68.5909	**56.7317**	75.547	**64.4525**	**67.0105**	70.2925	66.9041	**58.5779**
PSNR	38.741	**41.1411**	40.7255	**42.3743**	43.0107	**44.3902**	**34.5855**	34.1702	39.26568	**40.51895**
RMSE	0.0116	**0.0088**	0.0092	**0.0076**	0.0071	**0.005**	**0.0187**	0.0196	0.01165	**0.01025**
SSIM	**0.8749**	0.8615	**0.9172**	0.8674	**0.94**	0.8761	**1**	0.93	**0.933025**	0.868333

Scenario 3: 4/25 2014 – 10/30/2014.

In this case, we used two Landsat images collected on 4/25/2014 and 10/27/2014, and two WV images collected on 4/24/2014 and 10/30/2014. The prediction scenario is summarized in Figure 7. The patch sizes are 9, 7, 9, 1 for R, G, B, and NIR, respectively. No overlapping is used and every band is separately predicted. A comparison of ground truth WV image, Landsat image, and the predicted image is summarized in Figure. 8. Similar to the first two scenarios, the predicted image appears to be better. However, some unwanted details are also introduced.

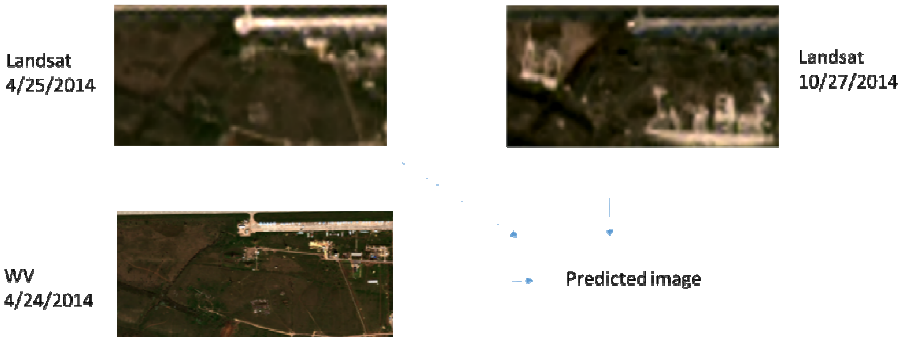
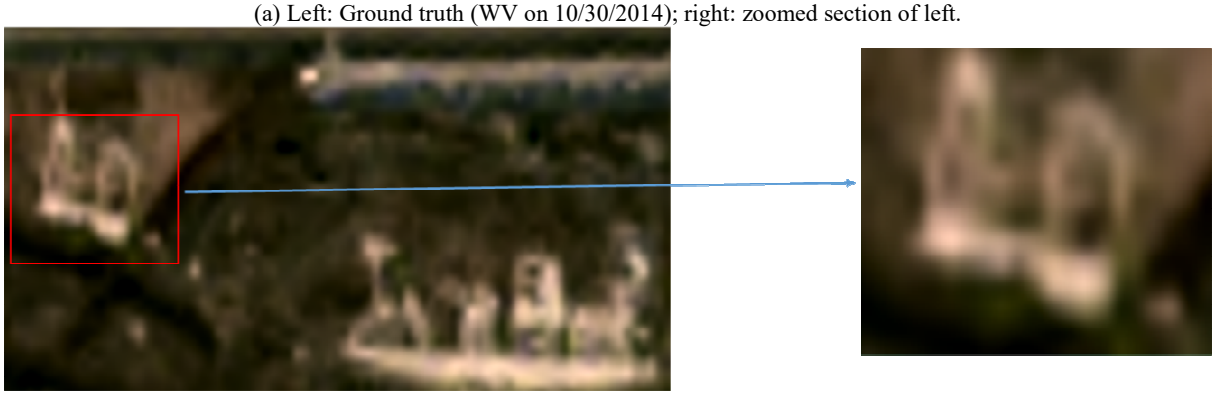
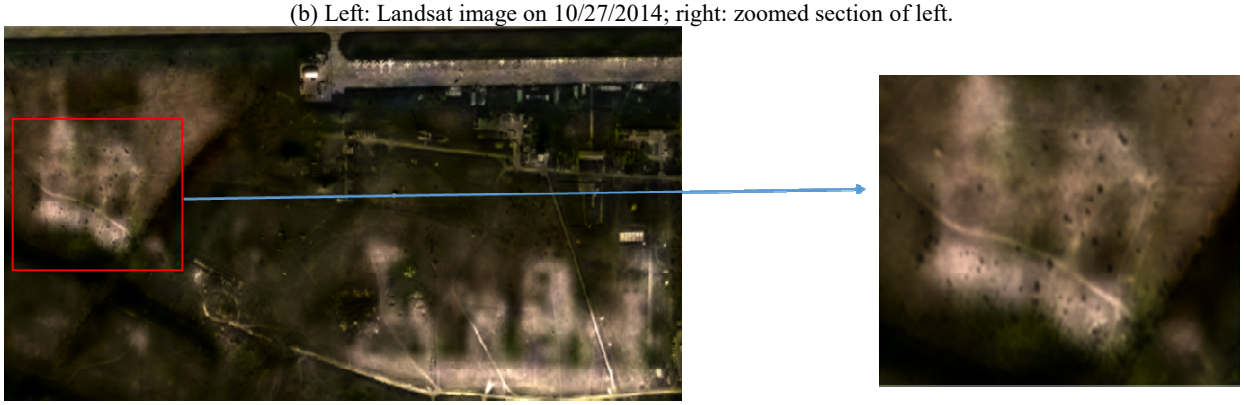


Figure 7. Two Landsat images at 4/25/2014 and 10/27/2014 and one WV image at 4/24/2014 are fused to generate a prediction. The prediction is then compared to a WV image collected on 10/30/2014.

(a) Left: Ground truth (WV on 10/30/2014); right: zoomed section of left.



(b) Left: Landsat image on 10/27/2014; right: zoomed section of left.



(c) Left: HCM predicted image on 10/30/2014; right: zoomed section of left.

Figure. 8. Comparison of different fused images with the ground truth (WV) and the low resolution Landsat image at the prediction time (10/30/2014).

Table 4. Performance metrics for Pair 3 (4/24/2014-10/30/2014).

	Fusion Red	Landsat Red	Fusion Green	Landsat Green	Fusion Blue	Landsat Blue	Fusion NIR	Landsat NIR	Fusion Avg	Landsat Avg.
AD	0.0037	0.0037	**0.0027**	0.0029	**0.0025**	0.0028	0.0065	**0.006**	0.00385	0.00385
CC	0.7208	**0.7542**	0.7492	**0.7648**	0.7452	**0.7498**	0.5707	**0.6383**	0.696475	**0.726775**
ERGAS	**46.4938**	53.0961	50.711	**49.1159**	82.5919	**81.8509**	26.8436	**24.6402**	**51.66008**	52.17578
PSNR	44.8143	**45.3684**	46.9492	**47.2268**	47.2358	**47.3141**	41.0603	**41.8042**	45.0149	**45.42838**
RMSE	0.0057	**0.0054**	0.0045	**0.0044**	0.0043	0.0043	0.0089	**0.0081**	0.00585	**0.00555**
SSIM	0.7841	**0.9521**	0.9058	**0.9599**	0.8948	**0.9503**	1	1	0.896175	**0.965575**

3.3 Discussions

From the results in Section 3.2, we have a few observations:

- For no-change areas, HCM performed better in terms of subjective evaluations.
- For changed areas, HCM tends to add some new textures that are not present in the ground truth images. This is probably because the resolution difference is 30 and the learned mapping between two Landsat images are not fine enough to help the prediction process.

4. CONCLUSIONS

In this paper, we present an image fusion algorithm that attempts to solve a challenging problem, which is to generate a high temporal resolution and high spatial resolution image sequence by fusing LR Landsat with HR WV images. Because the spatial resolution difference is 30 to 1, the mapping learned from the Landsat appears to be not good enough.

One potential idea is to apply some deep learning based deblurring algorithms to the Landsat images and then HCM is applied. Another idea is to apply Generative-Adversarial Network (GAN) to enhance the Landsat images directly.

ACKNOWLEDGEMENT

This research was supported by DARPA under contract #140D6318C0043. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] Kwan, C., Budavari, B., Bovik, A. C. and Marchisio, G., "Blind Quality Assessment of Fused WorldView-3 Images by Using the Combinations of Pansharpening and Hypersharpening Paradigms," IEEE Geoscience and Remote Sensing Letters, 14(10), 1835 – 1839, (2017)

[2] Kwan, C., Hagen, L., Chou, B., Perez, D., Li, J., Shen, Y. and Koperski, K., "Simple and Efficient Cloud and Shadow Detection Algorithms for Landsat and Worldview images," Signal, Image, and Video Processing, submitted, (2019).

[3] Kwan, C., "Image Resolution Enhancement for Remote Sensing Applications," 2nd International Conference on Vision, Image and Signal Processing, (2018)

[4] Kwan, C., "Remote Sensing Performance Enhancement in Hyperspectral Images," Sensors, 18(11), 3598, (2018)

[5] Loncan, L., de Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A. and Simoes, M., "Hyperspectral pansharpening: A review," IEEE Geosci. Remote Sens. Mag., 3(3), 27–46, (2015)

[6] Zhou, J., Kwan, C., Budavari, B., "Hyperspectral image super-resolution: A hybrid color mapping approach," Appl. Remote Sens., 10, 035024, (2016)

[7] Qu, Y., Qi, H. and Kwan, C., "Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution," Conference on Computer Vision and Pattern Recognition, (2018)

[8] Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R. and Wald, L., "A critical comparison among pansharpening algorithms," IEEE Trans. Geosci. Remote Sens., 2565–2586, (2015)

[9] Kwan, C., Dao, M., Chou, B., Kwan, L.M. and Ayhan, B., "Mastcam image enhancement using estimated point spread functions," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, (2017)

[10] Zhu, X., Helmer, E., Gao, F., Liu, D., Chen, J. and Lefsky, M., "A flexible spatiotemporal method for fusing satellite images with different resolutions," Remote Sens. Environ, 172, 165–177, (2016)

[11] Kwan, C., Budavari, B., Gao, F. and Zhu, X., "A hybrid color mapping approach to fusing MODIS and Landsat images for forward prediction," Remote Sens., 10, 520, (2018)

[12] Gao, F., Masek, J., Schwaller, M. and Hall, F., "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," IEEE Trans. Geosci. Remote Sens., 44, 2207–2218, (2006)

[13] Kwan, C., Zhu, X., Gao, F., Chou, B., Perez, D., Li, J., Shen, Y., Koperski, K. and Marchisio, G., "Assessment of Spatiotemporal Fusion Algorithms for Worldview and Planet Images," Sensors, 18(4), 1051, (2018)
[14] Dao, M., Kwan, C., Ayhan, B. and Bell, J. F., "Enhancing Mastcam Images for Mars Rover Mission," 14th International Symposium on Neural Networks, 197-206, (2017)
[15] Kwan, C., Budavari, B., Dao, M., Ayhan, B. and Bell, J. F., "Pansharpening of Mastcam images," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 5117-5120, (2017)
[16] Kwan, C., Ayhan, B. and Budavari, B., "Fusion of THEMIS and TES for Accurate Mars Surface Characterization," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 3381-3384, (2017)
[17] Kwan, C., Haberle, C., Ayhan, B., Chou, B., Echavarren, A., Castaneda, G., Budavari, B. and Dickenshied, S., "On the Generation of High-Spatial and High-Spectral Resolution Images Using THEMIS and TES for Mars Exploration," Proc. SPIE 10644, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV, 106440B, (8 May 2018)
[18] Kwan, C., Haberle, C., Echavarren, A., Ayhan, B., Chou, B., Budavari, B. and Dickenshied, S., "Mars Surface Mineral Abundance Estimation Using THEMIS and TES Images," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, (2018)
[19] Kwan, C., Choi, J. H., Chan, S., Zhou, J. and Budavari, B., "A Super-Resolution and Fusion Approach to Enhancing Hyperspectral Images," Remote Sensing, 10(9), 1416, (2018)
[20] Kwan, C., Choi, J. H., Chan, S., Zhou, J. and Budavari, B., "Resolution Enhancement for Hyperspectral Images: A Super-Resolution and Fusion Approach," IEEE International Conference on Acoustics, Speech, and Signal Processing, 6180 – 6184, (2017)