Estimation of Water Depths and Turbidity From Hyperspectral Imagery Using Support Vector Regression

Zhigang Pan, Craig Glennie, *Member, IEEE*, Carl Legleiter, and Brandon Overstreet

Abstract—We propose and evaluate an empirical method for water depth determination from hyperspectral imagery when the benthic layer is visible using support vector regression (SVR). The implementation of the empirical method is presented, and its ability to estimate water depths is compared with a more commonly used band ratio method for two distinct fluvial environments. Our analysis shows that SVR outperforms the band ratio method by providing better root-mean-square error (RMSE) agreement and higher R^2 for both clear and turbid water. We also demonstrate an extension of the nonparametric properties of SVR to provide estimates of water turbidity from hyperspectral imagery and show that the approach is able to estimate turbidity with an RMSE of approximately 1.2 NTU when compared with independent turbidity measurements.

Index Terms—Bathymetry, hyperspectral, support vector regression (SVR), turbidity.

I. INTRODUCTION

BATHYMETRY estimates provide the research community with quantitative spatially distributed models that allow the development of hydraulic and mesohabitat models [1]. Acquiring bathymetry through remote sensing techniques is well established and has been applied successfully in both coastal and inland fluvial environments [2]. Remote sensing can yield accurate quantitative bathymetry using typical methods such as echosounders, airborne bathymetric LiDAR, and passive optical imagery [3]. Due to their relatively high spatial resolution and accuracy, improved spectral resolution, and easily implemented image processing techniques, multispectral/hyperspectral imagery has emerged as a popular method for bathymetry retrieval [4].

Bathymetry using passive multispectral imagery can provide relatively high accuracy, and source imagery can be easily accessed through a number of free or low-cost portals. Once acquired, the algorithms to process the imagery are available in many conventional image processing software packages. The physical principle behind passive optical image bathymetry is that water depth, substrate type, water column properties, and water surface roughness all affect the observed reflectance [5]. Water attenuates the propagated light due to scattering and absorption in water body, with a rate of attenuation depending on wavelength. Through investigation of water optical properties, many analytical and empirical methods have been proposed to retrieve water depth directly from passive multispectral/hyperspectral imagery. Lyzenga [6] and Philpot [7] proposed and validated the efficacy of passive remote sensing for water bathymetry retrieval using a linear solution. Legleiter *et al.* [5] proposed a band ratio method to retrieve bathymetry from passive optical imagery data using a statistical test to select an optimal band pair for bathymetric regression. Bachmann *et al.* [8] used a manifold coordinate representation to retrieve bathymetry from hyperspectral imagery with a reduced representation. Sandidge and Holyer [9] used the Airborne Visible/Infrared Imaging Spectrometer data to train a neural network system to retrieve bathymetry with established quantitative empirical relationship.

The band ratio method implemented by Legleiter *et al.* [5] and Ma *et al.* [10] has been shown in the literature to be easy to implement, provide accurate bathymetric results from hyperspectral imagery, and have a minimum number of tuning parameters. The main drawback of the band ratio method is that only a portion of the spectral information is utilized; however, all bands, theoretically, are attenuated by the water column with specific, but wavelength-dependent, attenuation coefficients. Therefore, all spectral bands should potentially contain water bathymetry information. A more generalized model, which takes advantage of a number of spectral channels, could potentially improve bathymetric estimates. However, a specific model would be difficult to construct due to the complex and unknown relationship between the spectral bands and the water column response. A nonparametric method, which utilizes all spectral information, can be formulated and could potentially improve bathymetry retrieval. It would be also preferable if, like the OBRA method, the algorithm could be easily implemented without a requirement for extensive radiative transfer modeling.

Support vector machine (SVM) is a supervised machine learning algorithm used to analyze patterns within data, and support vector regression (SVR) is a realization of SVM for prediction [11]. SVR is a nonparametric regression method and a data-dependent learning scheme, and thus, no explicit model is required to fuse the observations with the physical measurement quantities. The regression model is generalized and can

be used to estimate any spectrally related parameters where no explicit physical model exists. A review of SVMs and regression can be found in [12]. Bruzzone and Melgani [13] have estimated biophysical parameters from remotely sensed data using a multiple-estimator system, which incorporated SVM in combination with a multilayer perceptron neural network. Camps-Valls *et al.* [14] utilized the SVR method to estimate ocean chlorophyll concentration with satellite remote sensing data to provide a more accurate, less biased, and noise-resistant model.

In this letter, we propose and test using SVR as an alternative method for bathymetric retrieval in order to more accurately account for the nonlinearity existing in the observed hyperspectral data. A conventional band ratio bathymetry retrieval method is used as a baseline for comparison with the proposed SVR method, and a comparison with *in situ* water depth observations is presented. To demonstrate the generality of the SVR method, we also evaluate its ability to predict other water column parameters empirically by using it to estimate water turbidity from hyperspectral imagery. We again compare the results with *in situ* measurements of turbidity.

II. SVR

SVR is a nonparametric regression technique, and therefore, no assumptions regarding the underlying data model are required. SVR can transform a nonlinear regression problem into linear regression through the implementation of a kernel function, which projects the original feature space into higher dimensional space. A hyperplane is then used to fit the projected space, and the estimated parameters can be used for subsequent prediction [15].

SVR is a supervised machine learning algorithm, and therefore, calibration samples are essential. We define (x_1, z_1), $(x_2, z_2), \ldots, (x_n, z_n)$, where $x_i \in R^N$ is the feature vector, and $z_i \in R^1$ is the target output. Here, N is the dimension of feature space, and n denotes the number of samples. With $\varepsilon - $ SV regression, the goal is to find a hyperplane $f(x)$ for calibration data set. The linear function $f(x)$ can be described as

$$y = f(x) = \langle \mathbf{w} \cdot \phi(\mathbf{x}) \rangle + b = \sum_{i=1}^{n} w_i \phi_i(x) + b \qquad (1)$$

where y is the predicted value, $\mathbf{w}$ is the weight vector, ϕ is the nonlinear mapping function for reprojection, and b is the bias term; more details for SVR can be found in [16].

Slack variables ξ_i and ξ_i^* are introduced to accommodate a soft-margin SVR [16] (see Fig. 1). Under given parameters $C > 0$ and $\varepsilon > 0$, the standard form of SVR is

$$\text{Min:} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \qquad (2)$$

$$\text{Subject to} \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i^* \\ f(x_i) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, \ldots, n. \end{cases} \qquad (3)$$

Here, C is the penalty parameter, which tunes the tradeoff between the generalization of functional relationship and the accuracy of the fitted hyperplane. ε is the maximum allowed deviation from the fitted hyperplane.
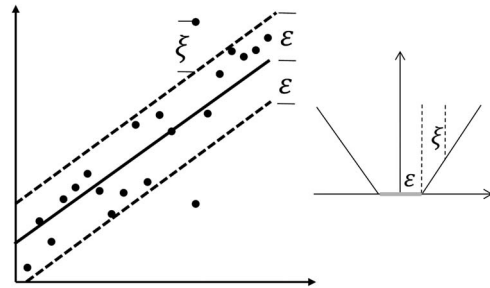

Fig. 1. Slack variable ξ for soft-margin SVR; ε is the maximum allowed deviation.

Kernel functions are also introduced into SVR to accommodate a nonlinear relationship in the linear formulation earlier. The kernel functions project the original feature space into a higher dimensional space that allows SVR to fit a hyperplane in a transformed feature space. There are many commonly used kernel functions, including linear, polynomial, radial basis function (RBF), and hyperbolic tangent [17]. RBF is widely used and implemented here because of its good performance and smaller number of input parameters. The RBF kernel can be described as

$$K(x_i, x_j) = \exp \left(-\lambda \|x_i - x_j\|^2 \right). \qquad (4)$$

Here, K denotes the kernel on two samples vectors x_i and x_j; λ is related to the kernel width, which requires tuning to achieve the best performance. The performance of SVR with RBF kernel is highly correlated to the three input parameters: C, ε, and λ. To optimize the selection of three parameters, a general k-fold cross-validation method and a grid searching scheme are used [18].

To evaluate the performance of the proposed method, root-mean-square error (RMSE) and R square value R^2 are utilized in this study.

III. EXPERIMENT DATA SETS

A. Airborne Hyperspectral Imagery

Both the Snake River in Wyoming Grand Teton National Park (see Fig. 2.1) and the confluence of the Blue and Colorado Rivers in north-central Colorado (see Fig. 2.2) are investigated in this study. The Snake River is predominantly clear water, and the study focused on an area referred to as Rusty Bend. The Colorado River originates in Rocky Mountain Park, and the Blue River enters the Colorado River from the south near Kremmling, Colorado. This site has variable water conditions because the Colorado River is also joined near the confluence by Muddy Creek, which, as the name implies, was turbid.

Hyperspectral imagery was collected with an ITRES Compact Airborne Spectrographic Imager (CASI)-1500 sensor. CASI-1500 is a pushbroom camera with 1500 across-track pixels spanning a $40°$ field of view and has a programmable spectral range that extends from 380 to 1050 nm with a maximum of 288 bands. The CASI image data were directly georeferenced using trajectory information from the Global Positioning System and the Inertial Navigation System onboard the aircraft.

The hyperspectral imagery was calibrated with manufacturer-provided software and calibration constants to convert the raw measurement into spectral radiance. ATCOR-4 software was used for atmospheric correction to convert the
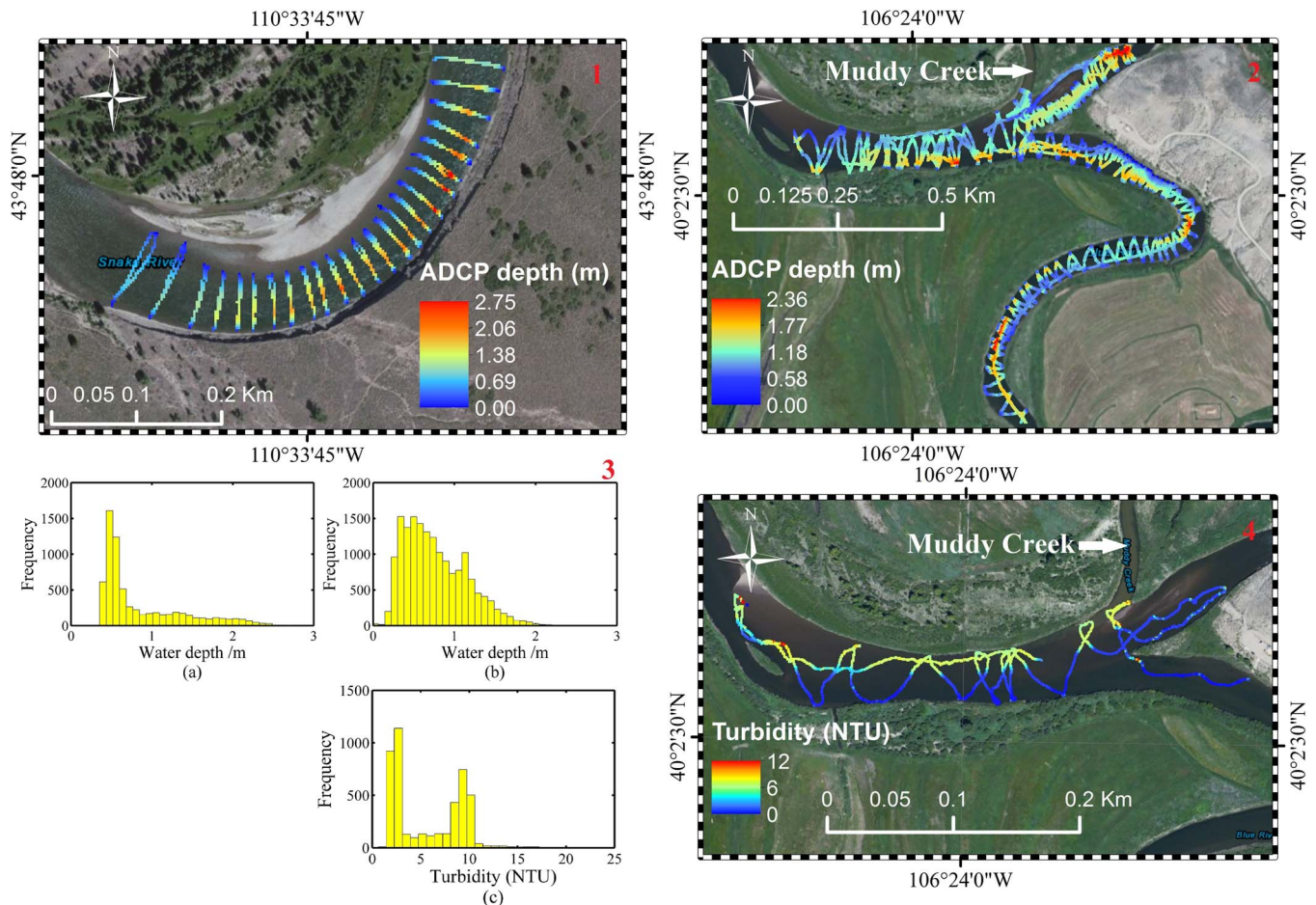
Fig. 2. (1) Snake River and (2) Blue/Colorado River study areas overlayed with ADCP depth observations. (3) Histograms of the *in situ* ADCP data for both rivers and turbidity measurements for the Blue/Colorado confluence. (4) Blue/Colorado River study area overlayed with *in situ* turbidity observations.

sensor-based reflectance to surface reflectance [19]. This program uses a digital elevation model and a database derived from Modtran-5 radiative transfer simulation to account for topographic and atmospheric effects on surface reflectance.

B. ADCP Data

Both the band ratio method (used as a baseline for comparison) and SVR are empirical methods that require a calibration data set to relate the physical measurements to observed spectral data. Acoustic Doppler Current Profiler (ADCP) data are used in this study to calibrate both the band ratio and SVR methods. The ADCP reference data were collected with a SonTek River Surveyor S5 ADCP deployed from a kayak. SonTek reports a depth resolution of 0.001 m and an accuracy of 1% over the range of 0.2–15 m. ADCP data are our primary ground reference data. We estimate the accuracy of the ADCP depths to be better than 3 cm for these two projects as all water was shallower than 3 m (see Fig. 2.3(a) for Snake River and Fig. 2.3(b) for Blue/Colorado River depth histograms).

C. Water Turbidity Measurements

Water turbidity observations are used in this study to validate the efficacy of SVR for deriving spectrally influenced parameters without an explicit model. Attenuation of light in water is wavelength dependent, and water turbidity is a primary environmental parameter that depicts the water column reflectance, and hence, the reflected energy received by a hyperspectral camera should be influenced by water turbidity. In this letter, a WET Labs EcoTriplet was deployed from a kayak on the Blue/Colorado River to measure the portion of the total backscattering associated with particulates in the water column, and turbidity was then derived from the measured backscatter (see Fig. 2.4). The distribution of the turbidity measurements is shown in Fig. 2.3(c), which displays a bimodal distribution due to the introduction of higher turbidity water from Muddy Creek (see Fig. 2.4).

IV. EXPERIMENT SETUP AND RESULTS

We studied the accuracy of SVR for bathymetry retrieval using hyperspectral imagery and compared its performance with optimal band ratio analysis (OBRA) [8]. To investigate the influence of calibration sample size, we increased the number of training samples from 100 to 500 in increments of 100, with a validation sample size fixed at 1000 to maintain the same comparison baseline. The calibration samples and validation samples were randomly selected from the *in situ* measurements. The calibration samples were first used to estimate the OBRA parameters for water depth retrieval; the retrieved parameters were applied to estimate water depths at the validation sample
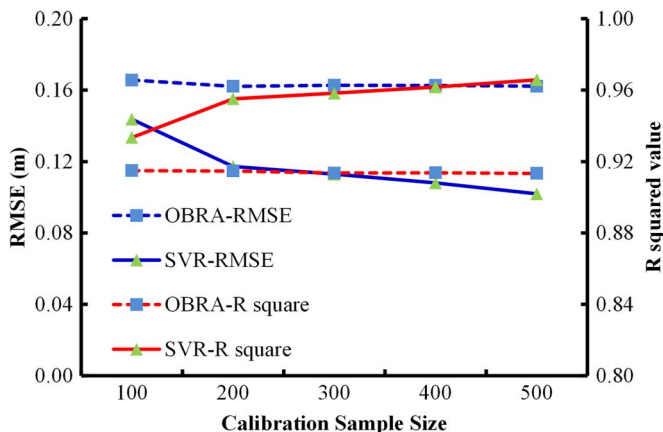
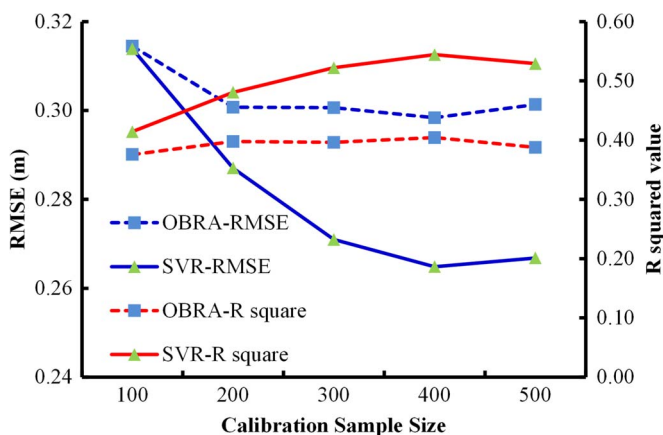Fig. 3. RMSE (m) and R^2 of OBRA and SVR for Snake River water depths.

Fig. 5. RMSE (m) with varying water depth of OBRA and SVR for Snake River water depths.

Fig. 4. RMSE (m) and R^2 of OBRA and SVR for Blue/Colorado River water depths.

Fig. 6. RMSE (m) with varying water depth of OBRA and SVR for Blue/Colorado River water depths.

locations. The same calibration samples were then fed into SVR for estimation of water depths; the retrieved water depths from both OBRA and SVR estimation were compared with the validation samples to calculate RMSE and R^2.

In order to ensure consistent and stable results from both algorithms, the experiment was repeated 20 times. In addition, to maintain a good model fit across all depths, the field measured water depths were first categorized into different depth bins with an interval of 0.1 m [see Fig. 2.3(a) and 2.3(b)]. The random calibration samples were chosen to have the same relative percentage as the water depths distribution. The RMSE and R^2 between retrieved water depth and validation samples were calculated for each iteration, and the average value was used for result.

The RMSE for SVR decreases as the calibration sample size increases for the Snake River, whereas R^2 also increases (see Fig. 3). OBRA shows a fairly consistent performance with all training sample sizes. SVR outperforms OBRA with both lower RMSE and higher R^2 regardless of the training sample size. Both SVR and OBRA depth estimates degraded in the more turbid Blue/Colorado River as the overall RMSE is higher and R^2 for both algorithms are lower than Snake River (see Fig. 4). However, SVR still significantly outperforms OBRA with a lower RMSE and higher R^2 value.

To further investigate the influence of water depth on the accuracy of depth determination for both OBRA and SVR,
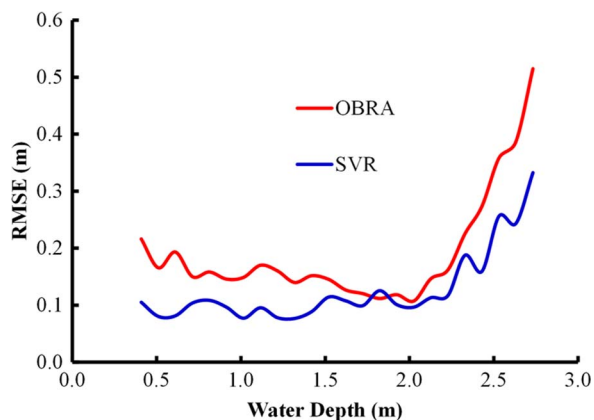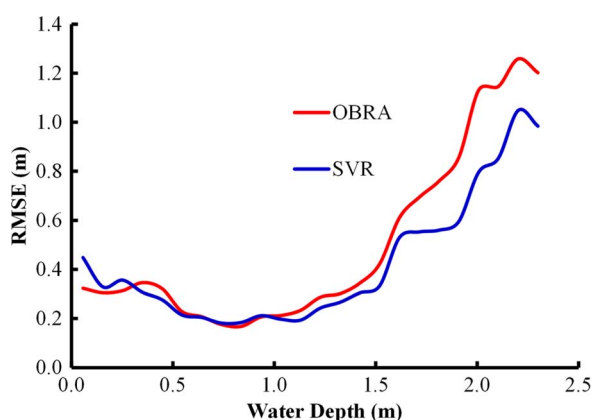
Fig. 5 (Snake River) and Fig. 6 (Blue/Colorado River) show the RMSE associated with varying water depths. The SVR results show significantly better RMSE for both shallower (< 1.5 m) and deeper water (> 2.5 m) for the Snake River; however, SVR results only show notable improvement in RMSE for deeper water (> 1.5 m) in the Blue/Colorado River. RMSE increases at deeper water (> 2 m for Snake River, > 1.5 m for Blue/Colorado River) due to the saturation of the water column radiance over bottom radiance signal in deeper water.

Finally, we briefly demonstrate the application of SVR for estimation of water turbidity from hyperspectral imagery. As the relationship between optical turbidity and observed spectra remains unknown, a physical model to predict turbidity is difficult to establish. Therefore, herein, we use the same k-fold cross-validation and calibration–validation procedures to train an SVR model for extraction of water turbidity. From our *in situ* turbidity observations, we again varied the number of training samples from 100 to 500 in 100 sample increments and kept 1000 observed turbidity measurements as validation samples. Fig. 7 shows that the RMSE between retrieved water turbidity decreases as the calibration sample size increases and is matched by a corresponding increase in R^2.

V. DISCUSSION AND CONCLUSION

In this letter, bathymetry extraction using SVR has been proposed and applied to two hyperspectral imagery data sets when
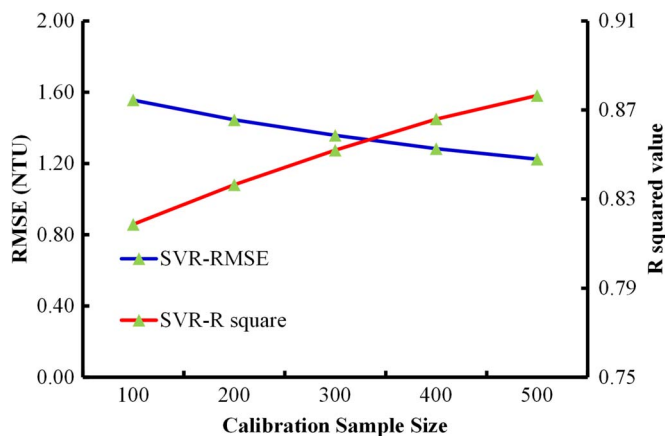
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

PAN *et al.*: ESTIMATION OF WATER DEPTHS AND TURBIDITY FROM HYPERSPECTRAL IMAGERY

5

Fig. 7. RMSE (NTU) and R^2 of OBRA and SVR for Blue/Colorado River water turbidity.

the benthic layer is visible in two distinct fluvial environments. The retrieved water bathymetry was compared with a band ratio method. SVR incorporates all available spectral bands instead of only examining the ratio of a pair of bands. This is important, as all spectral bands were influenced by both benthic reflection and water attenuation. The band ratio method, i.e., OBRA, establishes a physical model and then uses an empirical method to find the optimal pair of spectral bands and neglects all other spectral channels. The physics-based water depth retrieval characteristic of OBRA makes it more generalized. The results also show that OBRA provides consistent results with relatively few training samples; however, its performance, unfortunately, does not improve with an increase in the training sample size. As an empirical method, a band ratio method may not be optimal for the accurate estimation of bathymetry. In contrast, SVR does not require a physical model to bridge the observed data with the desired product. The bathymetry retrieved from SVR outperforms OBRA with a lower RMSE and better R^2 for both the Snake and Blue/Colorado Rivers. The increase of training sample size was shown to improve SVR water depth estimation performance as well. Because no explicit relationship is required between the observations and the estimated physical parameters, SVR can be potentially extended to estimate other spectral-based physical parameters. We also demonstrated how SVR can be used to estimate water turbidity, and the results show that hyperspectral imagery can be used to estimate turbidity with a low RMSE and high R^2.

For depth determination, the RMSEs for both SVR and OBRA degrade in deeper water (> 2.0 m for Snake River, > 1.5 m for Blue/Colorado River); this is caused by the non-linear influence of the water column radiance in deeper water. However, for both rivers, SVR still outperforms OBRA, likely because of its use of all spectral radiance measurements. Both SVR and OBRA are less effective in the more turbid Blue/Colorado River (SVR has a R^2 of 0.5); however, the estimated water turbidity shows an obvious coherence with the observed radiance (SVR has a R^2 of 0.87 when estimating turbidity). This is due to the assumption that, for spectral bathymetric retrieval, the observed benthic radiance is the more dominant term compared with the water column radiance. However, the substantial turbidity of the Blue/Colorado River has resulted in

significant water column radiance, which masks radiance from the benthic layer.

In summary, the proposed SVR method is effective at extracting water depths from hyperspectral imagery and outperforms the band ratio method. Due to its nonparametric formulation, SVR has been also successfully used to retrieve water turbidity from the observed hyperspectral imagery with high coherence with the physical measured water turbidity.

ACKNOWLEDGMENT

REFERENCES

[1] P. Carbonneau, M. A. Fonstad, W. A. Marcus, and S. J. Dugdale, "Making riverscapes real," *Geomorphology*, vol. 137, no. 1, pp. 74–86, Jan. 2012.
[2] D. Feurer, J. Bailly, C. Puech, Y. Le Coarer, and A. Viau, "Very-high-resolution mapping of river-immersed topography by remote sensing," *Progr. Phys. Geogr.*, vol. 32, no. 4, pp. 403–419, Aug. 2008.
[3] J. Gao, "Bathymetric mapping by means of remote sensing: Methods, accuracy and limitations," *Progr. Phys. Geogr.*, vol. 33, no. 1, pp. 103–116, Feb. 2009.
[4] W. A. Marcus and M. A. Fonstad, "Optical remote mapping of rivers at sub-meter resolutions and watershed extents," *Earth Surf. Process. Landforms*, vol. 33, no. 1, pp. 4–24, Jan. 2008.
[5] C. J. Legleiter, D. A. Roberts, and R. L. Lawrence, "Spectrally based remote sensing of river bathymetry," *Earth Surf. Process. Landforms*, vol. 34, no. 8, pp. 1039–1059, Jun. 2009.
[6] D. R. Lyzenga, "Passive remote sensing techniques for mapping water depth and bottom features," *Appl. Opt.*, vol. 17, no. 3, pp. 379–383, Feb. 1978.
[7] W. D. Philpot, "Bathymetric mapping with passive multispectral imagery," *Appl. Opt.*, vol. 28, no. 8, pp. 1569–1578, Apr. 1989.
[8] C. M. Bachmann *et al.*, "Bathymetric retrieval from hyperspectral imagery using manifold coordinate representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 884–897, Mar. 2009.
[9] J. C. Sandidge and R. J. Holyer, "Coastal bathymetry from hyperspectral observations of water radiance," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 341–352, Sep. 1998.
[10] S. Ma *et al.*, "Bathymetry retrieval from hyperspectral remote sensing data in optical-shallow water," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1205–1212, Feb. 2014.
[11] H. Drucker *et al.*, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process.*, 1997, vol. 9, pp. 155–161.
[12] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011.
[13] L. Bruzzone and F. Melgani, "Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 159–173, Jan. 2005.
[14] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, "Robust support vector regression for biophysical variable estimation from remotely sensed images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 339–343, Jul. 2006.
[15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, Apr. 2011.
[16] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[17] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.
[18] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.
[19] ReSe. Atcor4: Atmospheric & Topographic Correction for Wide FOV Airborne Optical Scanner Data. Accessed 9/5/2014. [Online]. Available: http://www.rese.ch/products/atcor/atcor4/