

CLASSIFICATION OF HYPERSPECTRAL DATA WITH ENSEMBLE OF SUBSPACE ICA AND EDGE-PRESERVING FILTERING

Junshi Xia^{1,2}, Lionel Bombrun^{1,2}, Tülay Adalı³, Yannick Berthoumieu^{1,2}, Christian Germain^{1,2}

¹ Univ. Bordeaux, IMS, UMR5218, F-33405 Talence, France

² CNRS, IMS, UMR5218, F-33405 Talence, France

³ University of Maryland Baltimore County, Baltimore, MD 21250, USA

ABSTRACT

Conventional feature extraction methods cannot fully exploit both the spectral and spatial information of hyperspectral imagery. In this paper, we propose an ensemble method of subspace independent component analysis (ICA) and edge-preserving filtering (EPF) for the classification of hyperspectral data to achieve this task. First, several subsets are randomly selected from the original feature space. Second, ICA is used to extract spectral independent components followed by a recent and effective EPF method, rolling guidance filter (RGF), to produce spatial features. The spatial features are treated as the input of a random forest (RF) classifier. Finally, the classification results from each subset are integrated together to produce the final map. Experimental results on real hyperspectral data demonstrate the effectiveness of the proposed method. A sensitivity analysis of this new classifier is also performed.

Index Terms— Classification, hyperspectral data, independent component analysis, edge preserving filter

1. INTRODUCTION

Due to its ability to provide detailed spectral information of different materials present in a given scene, hyperspectral data has increasingly become preferred for remote sensing applications [1, 2]. However, classification of hyperspectral data is a challenging problem largely due to the curse of dimensionality (Hughes phenomenon) [3]. Random Forest (RF) is one of the most widely used pixel-wise classifier for hyperspectral data, which is insensitive to the dimensionality of data and also has provided good performance in terms of classification accuracy. In addition, the RF classifier offers two main advantages, i.e. very low computational complexity and having only few parameters to tune [4, 5]. Another common strategy to overcome the Hughes phenomenon is to reduce the dimensionality via feature extraction or feature selection techniques [6]. In recent years, independent component analysis (ICA) has received attention for feature extraction of hyperspectral remote sensing images [7, 8, 9, 10]. In particular,

ICA extracts source components that generate the mixed signal measured by the sensor and the independent components refer to the different classes presented in the scene [7, 10]. Classically, when we apply ICA to hyperspectral data, principal component analysis (PCA) is performed first, and then, ICA is performed on a percentage of top, most important principal components, while the remaining components are discarded [10]. Although this method is effective for extracting features of hyperspectral data, it might lead to a loss of useful information in the discarded components. In order to alleviate this problem, we propose a random subspace ensemble approach, in which several subsets are randomly selected from the original spectral bands and then ICA is applied on each subset.

For high spatial resolution hyperspectral data, the classification performance is generally very low when only the spectral information is considered [11, 12]. In order to improve classification performance, spatial contextual information (relationship between neighboring pixels) should be included in the analysis [11, 12]. Previous approaches for including such information include mathematical morphology (MM) and Markov random fields (MRF). Li *et al.* [13] developed a generalized composite kernel (GCK) framework by combining spectral information and the most effective MM features: extend multi-attribute profiles (EMAPs). Recently, Xia *et al.* proposed two powerful classifiers, namely RoF-MRF [14] and RS-EMAPs [15]. The former combines the class posterior probabilities produced by Rotation Forest (RoF) and the spatial information represented by MRF-based multilevel logistic (MLL) prior [14]. The latter uses random subspace (RS) ensembles to classify EMAPs features [15].

Recently, edge-preserving filtering (EPF) has been successfully applied in many fields such as denoising [16]. It is used to remove noise, weak edges, and small details whereas the overall structure of the image is preserved. In this paper, the spatial contextual information is exploited by means of an EPF. Hence, it helps reduce the intraclass uncertainty, leading to high classification performance. Among the recent effective EPF solutions, we promote the rolling guidance filter (RGF) for extraction of spatial features from hyperspectral

data.

The main contribution of the paper is to propose a spectral-spatial classification method based on ensemble of subspace ICA and EPF. This framework offers many advantages.

1. ICA extracts discriminant features of each class in the image;
2. RGF provides complementary spatial information of the structures present in the scene;
3. Ensemble scheme can enhance the classification performance when compared to the use of a single technique.

The rest of the paper is organized as follows. In Section 2, ICA and EPF are introduced. The proposed method is described in Section 3 and the experimental results are shown in Section 4. The conclusion is given in Section 5.

2. BACKGROUND

Next, we briefly describe the ICA and RGF that play a critical role in the performance of the proposed method.

2.1. Independent component analysis (ICA)

ICA is an attractive solution to the blind source separation (BSS) problem, which decomposes an observed set of mixtures into a set of statistically independent components [7]. We consider the observed mixture $\mathbf{x} = [x_1, \dots, x_D]^T = \mathbf{A}\mathbf{s}$, which can be viewed as a linear combination of D random variables—or random processes when sample dependence is taken into account— $\mathbf{s} = [s_1, \dots, s_n]^T$ through a $D \times D$ non-singular mixing matrix \mathbf{A} . The statistically independent components (ICs) $\mathbf{y} = [y_1, \dots, y_D]^T$ are estimated by forming $\mathbf{y} = \mathbf{W}\mathbf{x}$, where \mathbf{W} is the estimated demixing matrix that makes use some form of diversity. The most widely used type of diversity is non-Gaussianity, i.e. higher-order-statistics (HOS). Two popular ICA approaches used in the remote sensing community, FastICA [17] uses a fixed nonlinearity to maximize non-Gaussianity and Joint Approximate Diagonalization of Eigenmatrices (JADE) [18] extracts the demixing matrix \mathbf{W} by joint a diagonalization of the fourth-order cumulant matrices.

A more attractive approach to making use of HOS is to use a dynamic nonlinearity that is matched to each of the estimated source densities, y_d , for $d = 1, \dots, D$ separately. Entropy bound minimization (EBM) utilizes an efficient entropy estimator to approximate the density of the sources by maximizing the entropy bound and using a finite number of prespecified measuring functions [19]. It provides robust performance using the four measuring functions proposed in [19] but also allows for selection of nonlinearities using prior information about the sources. Another important type of diversity, which is of particular interest here, is sample depen-

dence. The adjacent pixels in an image are highly correlated and making use of this additional statistical property when achieving ICA promises to further improve the performance of ICA [20]. Entropy rate bound minimization (ERBM) [21] effectively combines the dynamic nonlinearity selection of EBM with an invertible filter model and hence achieves better performance in terms of minimization of the entropy rate

$$\mathcal{I}_r(y_1; \dots; y_D) = \sum_{i=1}^D H_r(y_i) - \log |\det(\mathbf{W})| - H_r(\mathbf{x}) \quad (1)$$

where, $H_r(y_i)$ is the entropy rate of the process y_i and $H_r(\mathbf{x})$ is a constant with respect to \mathbf{W} .

2.2. Rolling Guidance Filter (RGF)

Rolling guidance filter, which has been proposed for non-linear image decomposition based on a modification of the bilateral filter, effectively removes noise and small details while preserving large-scale structures automatically, which the standard bilateral filter often fails to do [22].

Rolling guidance filter is composed of two steps, i.e. small structure removal and edge recovery. Small structure removal can be done by Gaussian filtering. Then, a joint bilateral filter is used to recover the edge iteratively. If we let J^1 initially set as the output of the Gaussian filtering stage and then J^{t+1} is the output of the t th iteration of joint bilateral filtering with the input image I and J^t , and is written as

$$J^{t+1}(i) = \frac{1}{Q_i} \sum_{j \in \mathcal{N}_p} \exp \left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J^t(i) - J^t(j)\|^2}{2\sigma_r^2} \right) I(j) \quad (2)$$

where, $Q_i = \sum_{j \in \mathcal{N}_p} \exp \left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J^t(i) - J^t(j)\|^2}{2\sigma_r^2} \right)$. i and j are the coordinates of pixels in the image, \mathcal{N}_p denotes the set of neighbor pixels of i , σ_s and σ_r control the spatial and range weights respectively.

In order to make RGF more effective, two main steps can be combined into one by starting rolling guidance simply from a constant-value (C) image. In this case, the initial step can be saved by starting rolling guidance from J^0 , where $\forall i, J^0(i) = C$. Algorithm 1 depicts the final rolling guidance filter implementation.

Algorithm 1 Rolling Guidance Filter

Input: I, σ_s, σ_r and n^{iter}

- 1: Initialize J^0 as a constant image
- 2: For $i = 1 : n^{iter}$
- 3: $J^t \leftarrow \text{JointBilateral}(I, J^{t-1}, \sigma_s, \sigma_r) \triangleright$ using (2)
- 4: Endfor

Output: The output image $G \leftarrow J^{n^{iter}}$

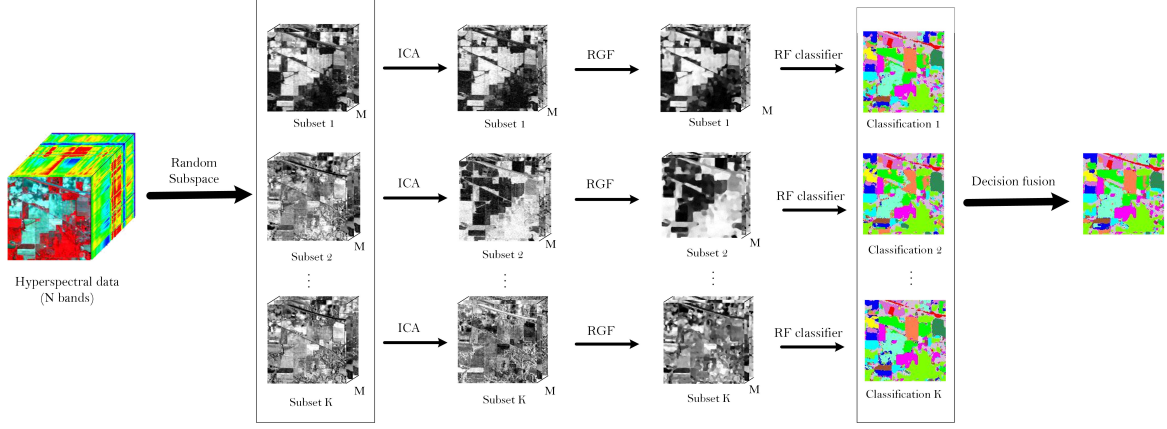


Fig. 1. Schematic of the proposed ensemble classification method

3. ENSEMBLE OF SUBSPACE ICA WITH RGF FOR CLASSIFICATION

As shown in Fig. 1, the proposed classification method consists of five steps: 1) randomly select M features from the original feature set of hyperspectral data; 2) perform ICA on the M selected features; 3) apply RGF on each extracted components; 4) perform classification on the filtered images in each subset; 5) combine the results together to generate the final classification map by a majority vote rule. Next, we describe the implementation for each step.

1) Random subspace: K subsets of spectral bands in hyperspectral data are randomly selected. Each subset contains M features. In this case, there are different features in each subset, resulting in generating different input features of RF classifier by using ICA and RGF in the next two steps, which could be beneficial for the decision fusion in the last step.

2) ICA: The aim of this step is to extract discriminant ICs for the classes in each subset, to be used for classification. Here, we propose to use ICA-EBM due to its superior separation performance and moderate computational complexity [19].

3) RGF: RGF is performed on each extracted component to obtain the m th feature in i th subset,

$$G_k^m \leftarrow \text{Algorithm 1} (IC_k^m, \sigma_s, \sigma_r, n^{iter}) \quad (3)$$

where, σ_s and σ_r are the spatial and range standard deviations of the filter. IC_k^m is the m th ($m = 1, \dots, M$) component derived from ICA in the k th ($k = 1, \dots, K$) subset. G_k^m is the resulting feature obtained by RGF.

4) Classification: RF is used for the classification of the filtered features [4]. In RF, two parameters, number of trees and number of selected features are empirically selected. In this paper, the number of trees is set to be 100 and the number of selected features is set to be \sqrt{M} .

5) Decision fusion: Different classification results are combined for the final class decision by a majority vote rule.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed approach is evaluated using real hyperspectral data, which is recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana, USA. This scene, which comprises 220 spectral bands in the wavelength range from 0.4 to 2.5 μm with spectral resolution 10 nm, is composed of 145×145 pixels, and the spatial resolution is 20 m/pixel. This dataset has 16 classes of interest. We randomly select 30 samples per class to form the training set and the rest of the pixels are used for testing (for the minor classes *Grass/pasture-mowed* and *Oats*, the training samples remain constant). K and M are set to be 10 and 16, respectively. σ_s and σ_r in RGF are set to be 7 and 0.1, respectively. The results are obtained after 10 Monte Carlo runs.

In the first experiment, the performance of the proposed method is compared with the following algorithms: 1) original spectral information; 2) random subspace (RS): proposed method without ICA and RGF; 3) random subspace with ICA (RS-ICA): proposed method without RGF and 4) random subspace with RGF (RS-RGF): proposed method without ICA. The classification results are shown in Table. 1. It can be seen that RS method does not improve the performance. The main reason is that we select a small number of features ($M = 16$) in each subset. RS-ICA and RS-RGF produced the better results than RS and original, indicating the effectiveness of ICA and RGF techniques. The proposed method yields the best performance, with 31.6, 31.6, 24.3 and 20.4 percentage points over original, RS, RS-ICA and RS-RGF, respectively. The proposed method achieves to the best classification results for all classes.

In the second experiment, we present comparisons with the proposed approach with different ICA algorithms (e.g. JADE and FastICA) (Table 2). For the ERBM, the filter length is set to be 3. Increasing this value does not significantly influence the result. For the RS-ICA, ERBM gives the best result.

Table 1. Overall, average and class-specific accuracies achieved by the classifiers.

Class	Original	RS	RS-ICA	RS-RGF	Proposed
Alfalfa	81.25	85.00	90.00	96.67	98.75
Corn-no till	40.16	41.84	66.80	68.75	90.70
Corn-min till	48.54	50.19	58.69	64.83	89.47
Bldg-Grass-Tree-Drives	67.60	70.69	81.13	95.25	99.75
Grass/pasture	83.49	82.68	82.46	78.61	93.53
Grass/trees	79.61	80.85	95.54	83.53	98.84
Grass/pasture-mowed	87.69	89.23	86.92	94.62	99.23
Corn	86.54	84.47	96.56	96.12	99.80
Oats	77.00	76.00	81.00	83.00	100.00
Soybeans-no till	64.37	65.82	72.75	66.52	91.86
Soybeans-min till	52.25	50.96	40.14	59.62	88.24
Soybeans-clean till	48.99	48.46	72.24	69.14	94.67
Wheat	94.62	94.34	99.51	94.45	99.34
Woods	83.31	83.41	92.99	82.48	98.27
Hay-windrowed	48.83	41.31	56.54	93.57	98.23
Stone-steel towers	98.00	98.00	96.00	98.92	99.08
OA	61.60	61.53	68.82	72.75	93.15
AA	71.39	71.45	79.33	82.88	96.23

Table 2. Classification accuracies obtained from the proposed methods using different ICA methods.

Classifier	RS-ICA			
	EBM	ERBM	JADE	FastICA
OA	68.82	72.61	65.36	65.29
AA	79.33	83.55	71.67	72.27
Classifier	Proposed method			
	EBM	ERBM	JADE	FastICA
OA	93.15	93.58	92.75	93.16
AA	96.23	96.63	96.14	95.82

Table 3. Classification accuracies obtained from the proposed methods in comparisons of other spatial-spectral classifiers

Classifier	RoF-MRF [14]	GCK [13]	RS-EMAPs [15]	Proposed method
OA	86.24	88.35	92.14	93.15
AA	92.76	92.62	94.71	96.23

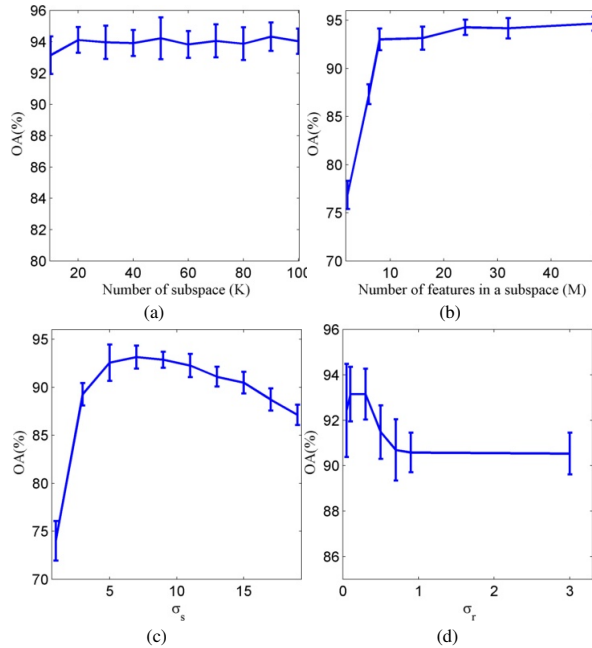


Fig. 2. Sensitivity of the proposed method. (a) influence of K ($M=16$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (b) influence of M ($K = 10$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (c) influence of σ_s ($K=10$, $M = 16$ and $\sigma_r = 0.1$) and (d) influence of σ_r ($K=10$, $M = 16$ and $\sigma_s = 7$)

This is due to the fact that ERBM not only matches a wide range of distributions but also considers sample dependence. EBM achieves the second best classification performance due to the use of an efficient entropy estimator and prior information. When we apply RGF on the features extracted from ICA techniques, the difference among the ICA techniques is not significantly, indicating that RGF plays an important role to improve the performance. Although ERBM is superior to the other ICA techniques using both spectral and spatial information as input, it requires more computation time than others. In this case, EBM is a good tradeoff between performance and computational complexity.

In the third experiment, we present comparisons of the proposed method against the aforementioned state-of-the-art spectral-spatial classifiers, such as GCK [13], RoF-MRF [14] and RS-EMAPs [15]. From Table 3, we can conclude that the proposed method outperforms those methods in this dataset.

In the fourth experiment, we investigate the parameter sensitivity of the proposed method to parameter choice. From Fig. 2, we observe that: 1) There is no pattern of dependence between K and the accuracy; 2) when M becomes larger, the proposed method tends to give better performance at the expense of increased computational complexity; 3) The proposed method achieves the best classification performance when $\sigma_s = 7$ but the performance is satisfactory over a wide range of values as well. As observed in Fig. 2(d), the optimal range of σ_r for this dataset is between 0.1 to 0.3.

Hence, selection of parameters is not very critical for the proposed methodology, which is an important added advantage. The only one that seems to provide better performance in a small range is σ_r . In practice, the users might select a small value of σ_r as in our case to better preserve the edges of hyperspectral data in order to increase the discrimination between the classes.

5. CONCLUSION

In this work, we developed a novel ensemble approach that combines random subspace, ICA, and rolling guidance filter for the classification of hyperspectral data. Experimental results confirmed improvements of the new approach over RS, RS-ICA and RS-RGF. The proposed method is superior to the recently proposed state-of-the-art spectral-spatial classifiers as well.

6. ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University to provide the AVIRIS hyperspectral data. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the ‘‘Investments for the future’’ Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02).

7. REFERENCES

- [1] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Plenum Publishing Co., 2003.
- [2] C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*, Wiley-Interscience, Hoboken, NJ, 2007.
- [3] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory.*, vol. 14, no. 1, pp. 55–63, 1968.
- [4] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [5] J. C. W. Chan and D. Paelinckx, "Evaluation of Random Forest and AdaBoost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.
- [6] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lotfi Zadeh, *Feature Extraction: Foundations and Applications*, Springer-Verlag New York, Inc., 2006.
- [7] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press. Elsevier., 2010.
- [8] J. Wang and C. I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, 2006.
- [9] M. Dalla Mura, A. Villa, J.A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sensing Lett.*, vol. 8, no. 3, pp. 542–546, 2011.
- [10] N. Falco, J.A. Benediktsson, and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2183–2199, June 2014.
- [11] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.
- [12] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [13] J. Li, P. Reddy Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, 2013.
- [14] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, 2015.
- [15] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015.
- [16] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 35, no. 6, pp. 1397–1409, June 2013.
- [17] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.
- [18] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 362–370, December 1993.
- [19] X. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5151–5164, Oct 2010.
- [20] T. Adali, M. Anderson, and G. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Proc. Mag.*, vol. 31, no. 3, pp. 18–33, 2014.
- [21] X. Li and T. Adali, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 1934–1937.
- [22] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Computer Vision ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8691 of *Lecture Notes in Computer Science*, pp. 815–830. Springer International Publishing, 2014.