Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **3**, May **2017**

Improved Endmember Extraction Algorithms Based on Endmember Independence and Thinning Technology

Ying Cui[1,2,*], Jiaqi Wang[1], Guojiao Song[1], Shubin Liu[2],

[1]College of Information and Communication Engineering
Harbin Engineering University
145 Nantong Avenue, Nangang District, Harbin, Heilongjiang

[2]Remote Sensing Technology Center
Heilongjiang Academy of agricultural science
368 Xuefu Avenue, Nangang District, Harbin, Heilongjiang
[*]Corresponding author: cuiying@hrbeu.edu.cn

Received April, 2016; revised February, 2017

Abstract. *Two improved endmember extraction algorithms are proposed based on endmember independence and thinning technology. The improved algorithms introduced endmember independence into traditional endmember extraction algorithms. One improved algorithm combines simplex volume calculation criteria and endmember independence as thinning criteria. The other improved algorithm uses high dimensional simplex volume calculation criteria and endmember independence as thinning criteria. The two improved algorithms are proposed to obtain more accurate endmembers. Compared with traditional N-FINDR algorithm and endmember extraction algorithm based on simplex volume calculation criteria, the proposed algorithms can obtain endmembers with higher classification accuracy and lower RMSE accuracy by the experiment results.*
Keywords: Image Processing; Endmember independence; Endmember Independence; Thinning Technology

1. **Introduction.** In recent years, the development of hyperspectral remote sensing technology to quantitative direction, the mixed pixels existing is an inevitable obstacle. The mixed pixels greatly affect the accuracy of ground-object recognition for the hyperspectral image analysis, as well as restrict the further development of hyperspectral remote sensing. The endmember, i.e., pure pixel, contains the same type of target. The data unmixing premise is endmember extraction. Currently, there are many representative algorithms to extract endmember, including N-FINDR [1, 2], Simplex Growing Algorithm [3](SGA), Convex Cone Analysis [4](CCA) and Iterative Error Analysis [5](IEA) etc. Among them, the N-FINDR endmembers selection is better, but it lacks of criteria to determine the number of initial endmembers and dimensionality reduction can cause deviation. Currently, some improved algorithms based on N-FINDR have been proposed. High-dimensional simplex volume formula is applied to endmember extraction. The formula is not related to the data dimension, deviation of dimension reduction is avoided in [6]. In [7], the candidate endmember set is obtained by using quantile of the chi-square distribution. Compared with the N-FINDR, this method is faster. PPI and VCA are used to extract the initial endmembers in [8], all pixels are selected with the same probability by the algorithm. In order to improve the search speed, the intelligent optimization

algorithms are used to extract endmember in [9, 10]. Furthermore, some improved algorithms have been proposed. In [11], the volume formula uncorrelated to the dimension is introduced, thus, the complexity of the algorithm is reduced. The mean square error as the objective function in [12] is used, crossover and mutation of the differential evolution is proposed to extract endmember. The methods precision is higher, the convergence and the global search ability is better. In [13], the pre-processing module of spatial and spectral information combines with the existing algorithms. It ensures the accuracy and reduces the amount of computation. Now, the nature of endmembers is not explored in the improved N-FINDR. Since, among the extracted endmembers, there may be similar. It is particularly important to conduct endmember thinning by the endmembers properties. In this paper, it is a research starting point. Firstly, based on simplex volume calculation criterion and endmember independence, the endmember thinning criterion is proposed. Secondly, based on high-dimensional simplex volume calculation criterion and the endmember independence, the N-FINDR is improved. Compared with two original algorithms, the proposed algorithms can obtain endmembers with higher classification accuracy and lower RMSE accuracy.

2. **N-FINDR algorithm.** For N-FINDR algorithm, simplex volume composed by pure pixels, i.e., endmembers, more than simplex volume composed by any other pixels in spectral space. Thus, in N-FINDR, a set of random pixels is the initial solution, through the replacement of the pixels and the endmembers, simplex volume will expand until no seeking greater simplex volume. In [1], the implementation is as follows:

Step 1 Assuming that the number of endmembers is p, the original hyperspectral data down to p-1 dimensions;

Step 2 After the image dimension reduction, randomly select p pixels as the initial candidate endmembers and calculate the simplex volume V_1. Simplex volume is calculated as follows:

$$V(E) = \frac{1}{(p-1)!} abs(|E|) \tag{1}$$

$$E = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ A_1 & A_2 & \cdots & A_p \end{bmatrix} \tag{2}$$

Where A_i is the i–th endmember corresponding p-1 dimensional column vector;

Step 3 Select a pixel P_1 from other pixels, recalculat simplex volume V_2. If $V_2 > V_1$, replace a candidate endmember as a new candidate endmember;

Step 4 Replace other candidate endmembers using pixel P, implement Step 3;

Step 5 Cyclically implement Step 3–4 for all other pixels, the largest simplex volume is obtained, and its corresponding vertex are the required endmembers;

3. **Endmember extraction algorithms of high-dimensional simplex criterion.** Since N-FINDR is limited by the data dimension, high-dimensional simplex volume calculation criterion in [6] is proposed. $p_1, p_2, \cdots, p_n$ pixels compose n-1 dimensional high dimensional simplex, its volume:

$$V_n = V(p_n - p_1, p_n - p_2, \cdots, p_n - p_{n-1}) = \frac{1}{(n-1)!} \sqrt{|A_{n-1}^t A_{n-1}|} \tag{3}$$

Where $A_{n-1} = (p_n - p_1, p_n - p_2, \cdots, p_n - p_{n-1})$. Since $A_i^T A_i$ is square, it is applied to extract the initial endmembers, the original data dimension is not reduced. The specific implementation of the algorithm refers to [6], the method is called TN algorithm.

4. **Improved endmember extraction algorithms.**

4.1. **Endmember relevance.** In order to further refine the endmembers, the endmembers characteristics are analyzed. For a hyperspectral image (including multiple objects), the obtained endmembers should be independent, i.e., endmembers have higher independence. Since the spectrum of spatial and temporal has variability, it is possible to extract different endmembers for the same object in practical operation. Therefore, this paper uses higher relevance endmembers instead of lower independence endmembers, and selects endmembers belonging to the same objects to do thinning in order to improve the accuracy of the terminal endmembers. Endmember relevance is calculated as follows in [14]:

$$r = \frac{\sum_{k=1}^{c}(R_{ik} - \bar{R}_i)(R_{jk} - \bar{R}_j)}{\sqrt{[\sum_{k=1}^{c}(R_{ik} - \bar{R}_i)^2][\sum_{k=1}^{c}(R_{jk} - \bar{R}_j)^2]}} \tag{4}$$

Where, R_{ik} and R_{jk} are the grey levels of the i-th endmember and j-th endmember on k-band, and represent respectively the average gray levels of the i-th endmember and j-th endmember, c represents the number of bands.

4.2. **Thinning Technology.** Thinning technique is common method, it obtains regional framework in the field of image processing. Its purpose is to reduce the redundant information to highlight the main shape of the targets. Thinning technique can be divided into two categories of non-iterative algorithms and iterative algorithms. Non-iterative algorithms produce regional framework by one image scanning. Iterative algorithms iteratively delete pixels that meet certain conditions until the regional framework is obtained. In iterative algorithms, parallel algorithms firstly mark deleted points, and all points are unified deleted until all the points in the image are processed. Serial algorithms find a deleted point in the process of scanning images, and the point is immediately deleted.

According to the above introduction, the serial algorithms are similar to the endmember search process. So in this paper, the algorithm is applied to endmember extraction to compose endmember extraction thinning criterion. Specific methods are as follows:

(1) In the initial extracted endmembers, using endmember independence selects endmembers and composes the scanning area.

(2) In the iterative process, improved volume calculation criteria as a condition is met, endmembers are replaced in the scanning area and volume is calculated.

(3) If the volume becomes larger, the corresponding endmember is replaced, until the largest simplex volume is obtained.

4.3. **Improved algorithms description.** In this paper, two improved endmember extraction algorithms are proposed based on the above. In the improved N-FINDR algorithm, endmember independence and high-dimensional simplex volume calculation criterion are used as the thinning criterion. The algorithm is called NFINDRCTN algorithm. The flowchart is shown as Fig. 1, the specific implementation of the algorithm is as follows:

Step 1 Extract the initial endmembers by N-FINDR algorithm, the number is p;

Step 2 Calculate high-dimensional simplex volume V_1 of the initial endmembers by Equation (3);

Step 3 Calculate the initial endmembers relevance by Equation (4), and select endmembers whose relevance value are greater than relevance threshold r to construct the scanning area;

Step 4 Select a replaced endmember D in scanning area;

Step 5 Select a pixel Q from other pixels to replace the endmember D, recalculate high-dimensional simplex volume V_2. If $V_2 > V_1$, the pixel Q replaces the endmember D, if not, reserve the endmember D;

Step 6 Determine whether traverse all pixels, if so, continue to step 7; if not, return to Step 5;

Step 7 Determine whether traverse all replaced endmembers, if so, continue to step 8; if not, return to Step 4;

Step 8 Output the final endmember set.

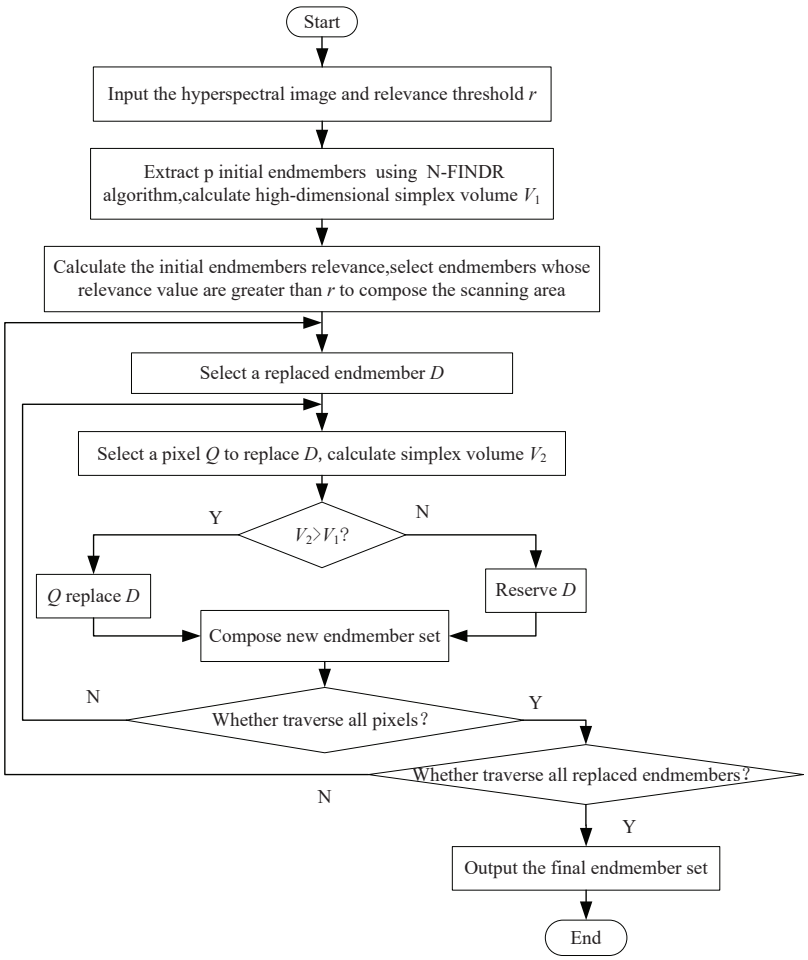

FIGURE 1. Flowchart of NFINDRCTN

In the improved TN algorithm, endmember independence and simplex volume calculation criterion are used as the thinning criterion. The algorithm is called TNCNFINDR algorithm. The flowchart is shown as Fig. 2, the specific implementation of the algorithm is as follows:

Step 1 Extract the initial endmembers by TN algorithm, the number is p;

Step 2 Calculate the initial endmembers relevance by Equation (4), and select endmembers whose relevance value are greater than relevance threshold r to construct the scanning area;

Step 3 The original hyperspectral data down to p-1dimensions, calculate simplex volume V_1 of the initial endmembers by Equation (1);

Step 4 Select a replaced endmember D in scanning area;

Step 5 Select a pixel Q from other pixels to replace a endmember D, recalculate simplex volume V_2. If $V_2 > V_1$, the pixel Q replaces the endmember D, if not, reserve the endmember D;

Step 6 Determine whether traverse all pixels, if so, continue to step 7; if not, return to step 5;

Step 7 Determine whether traverse all replaced endmembers, if so, continue to step 8; if not, return to step 4;

Step 8 Output the final endmember set.

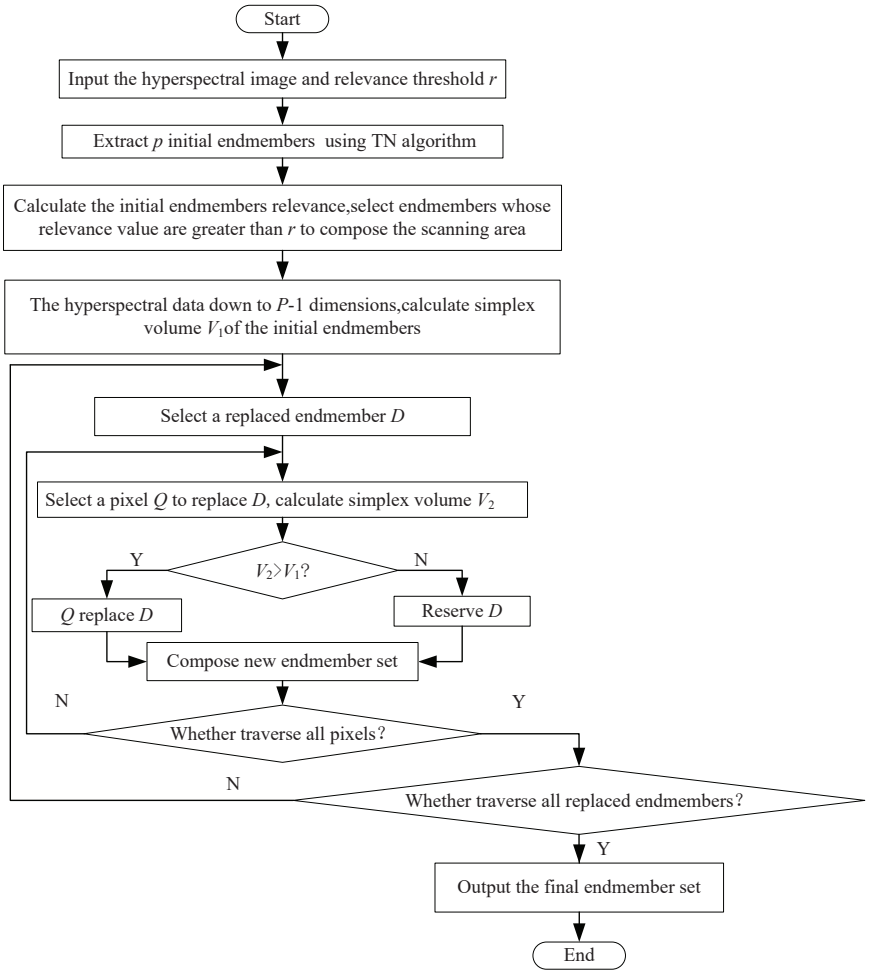

FIGURE 2. Flowchart of TNCNFINDR

5. **Experiment and analysis.**

5.1. **Experiment purpose.** In this paper, the method in [6] is called TN, the method in [15] is called SN-FINDR. Both methods proposed in this paper are called NFINDRCTN and TNCNFINDR. This paper mainly uses synthetic data and real hyperspectral

images to compare and analyze results of N-FINDR, TN, SN-FINDR, NFINDRCTN and TNCNFINDR.

5.2. **Calculation of unmixing error.** The unmixing effect is measured by unmixing error. This paper employs fully constrained least squares method as unmixing algorithm. For real hyperspectral image unmixing, endmember spectrum and its corresponding abundance information generally are unknown. At this time, using RMSE is used to evaluate and analyze unmixing effect. Assuming that X is the original spectral image, in the original image, residual of the i-th pixel X_i is:

$$\varepsilon_i = X_i - AS_i \tag{5}$$

RMSE of the i-th pixel X_i is:

$$RMSE_i = \sqrt{\frac{1}{l}\sum_{j=1}^{l}\varepsilon_{i,j}^2} \tag{6}$$

RMSE of whole image is:

$$RMSE_i = \sqrt{\frac{1}{N\cdot l}\sum_{i=1}^{N}\sum_{j=1}^{l}\varepsilon_{i,j}^2} \tag{7}$$

5.3. **Synthetic data simulation experiments.** Synthetic data are divided into two groups. The first group is to compare simplex volume of N-FINDR with NFINDRCTN, TN with TNCNFINDR. In the compared experiments, (-15, 0), (0, 20) and (15, 0) are as the theoretical endmembers. Based on three theories synthesize, 1000 points which add to gaussian noise of mean is 0 and variance is 2 are randomly generated by normalized coefficients and linear mixed model. Synthetic data and experimental results of N-FINDR and NFINDRCTN, TN and TNCNFINDR are shown as Fig. 3 and Fig. 4. The composed volume are shown in Table 1 and Table 2.

Table 1. Comparison of N-FINDR and NFINDRCTN

Endmember extraction algorithm	NFINDR	NFINDRCTN
Simplex volume	11.1073	11.1273

Table 2. Comparison of TN and TNCNFINDR

Endmember extraction algorithm	TN	TNCNFINDR
Simplex volume	0.6165	0.6481

From Table1 and Table2, the volume of NFINDRCTN and TNCNFIND are greater than N-FINDR and TN respectively. The triangle by the extracted endmembers of TN, N-FINDR, NFINDRCTN and TNCNFIND are shown in Fig. 3 and Fig. 4. Compared with N-FINDR and TN, the extracted endmembers using NFINDRCTN and TNCNFIND are closer to the theoretical endmembers, the searched volume are larger.

The second group is to compare unmixing effect of N-FINDR, TN, SN-FINDR, TNCN-FINDR and NFINDRCTN. Data synthetic rules are as follows: randomly generate vector $\boldsymbol{a}$ and $\boldsymbol{b}$. The dimension is 1*12. Vector $\boldsymbol{a}$ and $\boldsymbol{b}$ which add to random noise with mean

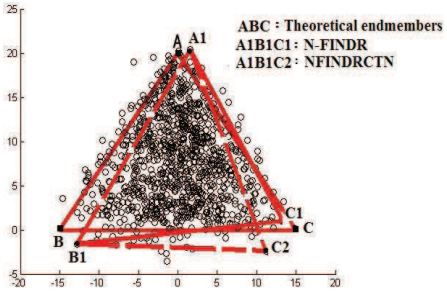

FIGURE 3. experimental results of NFINDRCTN

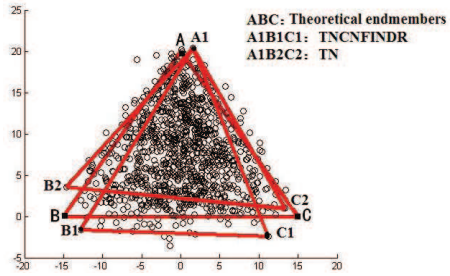

FIGURE 4. experimental results of TNCNFINDR

TABLE 3. Comparison of unmixing error with four methods with synthetic data

Endmember extraction algorithm	N-FINDR	TN	TNCNFINDR	NFINDRCTN
Unmixing error	0.1573	0.1067	0.1395	0.1244

TABLE 4. Experimental objects and the number of pixels

Name of object	Corn 1	Corn 2	Ranch	Bush	Hay	Bean 1	Bean 2	Bean 3	Arbor
Number of pixels	1428	830	483	730	478	972	2455	593	1265

TABLE 5. Comparison of experimental results with different number of pixels of 6 objects

Species of actual objects	Number of pixels	Method	Unmixing error	Species of separated objects
6 objects (the number of endmembers is 7)	360 pixels (60 pixels per type)	N-FINDR	0.013059	4
		TN	0.011774	5
		SN-FINDR	0.013291	5
		TNCN-FINDR	0.011723	5
		NFIND-RCTN	0.011175	5
	480 pixels (80 pixels per type)	N-FINDR	0.011723	4
		TN	0.012336	3
		SN-FINDR	0.017719	3
		TNCN-FINDR	0.011296	4
		NFIND-RCTN	0.010979	4

is 0 and variance is 0.5 are used to generate matrix $\boldsymbol{A}$ and $\boldsymbol{B}$. The dimension is 1200*12. Matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ with different proportions synthesize matrix $\boldsymbol{C}$. The data of $\boldsymbol{C}$ is as follows: 1 to 200 is 100% of A from 1 to 200, 201 to 400 is 80% of $\boldsymbol{A}$ from 201 to 400 and 20% of $\boldsymbol{B}$ from 201 to 400, and so on, the synthesized ratio of $\boldsymbol{A}$ decreases from 100% to 0%, the synthesized rate of $\boldsymbol{B}$ increased from 0% to 100%, the rate is 20%. In the extraction process, the number of the endmembers is two. The unmixing error of four methods is given in Table 3. From Table 3, it shows that the proposed algorithms have improvement, the unmixing error of NFINDRCTN is better than NFINDR. Because the endmembers number is small, the unmixing error of TNCNFINDR is bigger than TN. If the endmembers number is increasing, the unmixing error is decreasing. The experiment results are shown in the real hyperspectral image simulation experiments.

Table 6. Comparison of experimental results with different number of pixels of 9 objects

Species of actual objects	Number of pixels	Method	Unmixing error	Species of separated objects
9 objects (the number of endmembers is 10)	540 pixels (60 pixels per type)	N-FINDR	0.012591	7
		TN	0.01305	7
		SN-FINDR	0.015734	6
		TNCN-FINDR	0.012388	8
		NFIND-RCTN	0.012052	7
	720 pixels (80 pixels per type)	N-FINDR	0.013224	5
		TN	0.014185	6
		SN-FINDR	0.014806	6
		TNCN-FINDR	0.012164	7
		NFIND-RCTN	0.012475	5

5.4. **Real hyperspectral image simulation experiments.** The original AVIRIS image is used in this paper. It was a part of remote sensing research area in Northwest Indian, Indiana, USA. It was shot in June, 1992. In this paper, 200 bands are used removing noise and water vapor absorption. Due to spatial resolution of AVIRIS image is only 20m×20m, pixel mixing probability is higher. Some objects contains only dozens of pixels, those pixels can not fully represent the spectral properties of the object. So this paper only selects 9 objects with the larger pixels number as candidate endmembers. The names of 9 objects and the pixels number are shown in Table 4.

In [15], it only gives performance improved methods with 9 objects, and the experimental results with different objects number and different pixels number are not discussed. Therefore, in this paper, the selected types of objects respectively is 6 and 9, and the pixels numbers is 60, 80. The six objects are Corn 1, Corn 2, Ranch, Bean 1, Bean 2 and Bean 3. In order to ensure the reasonableness of comparative experiments, the number of the selected endmembers is $Z + 1$ (Z is the number of types of objects).

For unknown hyperspectral images, the endmembers number is obtained automaticly by using OSP endmember extraction algorithm combined with Endmember independence [14]. First, extracte i endmembers by OSP algorithm, and calculate endmember relevance to compose endmembers correlation matrix M_i. The number of non-diagonal elements which is higher than endmember relevance threshold on upper triangle M_i, is called strong relevance index of endmember i and denoted $ESCI_i$. The endmember number i increases from 1. When i meet $ESCI_i \geq 1$ for the first time, i is actual the endmembers number. In order to more comprehensively analyze the advantages and disadvantages of the proposed algorithms, in addition to the unmixing error, the classification accuracy of algorithm is also introduced in this paper.

Before the final experimental comparison, first, the selection of relevance threshold (r) with improved algorithms is analyzed. By analyzing the unmixing error curves, for 6 objects, the relevance thresholds with TNCNFINDR respectively are: 0.97, 0.93, the relevance thresholds with NFINDRCTN respectively are: 0.95, 0.92. For 9 objects, the relevance thresholds with TNCNFINDR respectively are: 0.93, 0.94, the relevance thresholds with NFINDRCTN respectively are: 0.92, 0.95.

For different number of pixels, the experimental results of 6 objects are shown in Table 5. From Table 5, it can conclude that the unmixing error and classification accuracy of TNCNFINDR are improved than TN algorithm. Similarly, the unmixing error and classification accuracy of NFINDRCTN algorithm are improved than N-FINDR algorithm. In addition, the results obtained by the two methods are better than SN-FINDR algorithm, and NFINDRCTN algorithm is better than TNCNFINDR algorithms.

In different number of pixels, the experimental results of 9 objects are shown in Table 6. As can be seen from Table 6, TNCNFINDR algorithm and NFINDCTN algorithm can also improve the TN algorithm and N-FINDR algorithm in different number of pixels. But TNCNFINDR algorithm is better than NFINDRCTN algorithm.

6. **Conclusions.** The improved endmember extraction algorithms based on endmember independence and thinning technology are proposed. The algorithms combine endmember independence and endmember thinning criteria to complete extraction and thinning of endmember. Based on different objects and different number of pixels, the real hyperspectral data experiments show that the algorithms can improve the original algorithms. But the algorithms have different advantages, TNCNFINDR has great advantages for more objects type, NFINDRTN has great advantages for less objects type. In practical application, the better method is selected. In addition, TN algorithm takes a long time to extract endmembers based on more pixels, and make computational time rise. How to reduce the computational time, and improve the algorithm usefulness is one of the next works to be solved.

Acknowledgment. This work was supported by the National Natural Science Fund of China (No. 61675051), the doctoral Fund of Ministry of Education of China (No.201323041 10007), Heilongjiang postdoctoral special fund (LBH-TZ0420). The authors would like to thank the associate editor and reviewers for their comments that help in improving this paper.

REFERENCES

[1] E. M. Winter, N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data, *SPIE Conference on Imaging Spectrometry V*, Denver, Colorado, pp.266–275, 1999.

[2] A. Plaza, C. I Chang, An improved N-FINDR algorithm in implementation, *Algorithms and Techonologies for Multispectral, Hyperspectral and Ultraspectral Imagery XI*, Bellingham, WA, pp.298–306, 2005.
[3] C. I. Chang, C. C. Wu, W. M. Liu and Y. C. Ouyang, A new growing method for simplex-based endmember extraction algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, vol.44, no. 10, pp.2804–2819, 2006.
[4] A. Ifarraguerri, C. I. Chang, Multispectral and hyperspectral image analysis with convex cones, *IEEE Transactions on Geoscience and Remote Sensing*, vol.37, no. 2, pp.756–770, 1999.
[5] R. A. Neville, K. Staenz, T. Szeredi J. Lefebvre and P. Hauff, Automatic endmember extraction from hyperspectral data for mineral exploration, *Proceedings of Fourth International Airbone Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing*, Ottawa, Canada, pp.21–24, 1999.
[6] X. R. Geng, Y. C. Zhao and G. H. Zhou, An automatic endmember extraction algorithm using simplex volume, *Progress in Natural Science*, vol.16, no. 9, pp.1196–1200, 2006.
[7] H. Y. Ding, W. Z. Shi, Fast N-FINDR algorithm for endmember extraction based on chi-square distribution, *Journal of Remote Sensing*, vol.17, no. 1, pp.122–137, 2013.
[8] X. J. Zhong, *Hyperspectral data mixed pixel decomposition and spectral matching verification algorithm*, Master. Thesis, Nanjing University of Science and Technology, Nanjing, China, 2013.
[9] W. Chen, X. C. Yu, P. Q. Zhang and H. Wang, Particle Swarm Optimization Genetic Algorithm for Endmember Extraction, *Computer Engineering*, vol.37, no. 16, pp.188–190, 2011.
[10] K. M. Yang, T. Zhang, S. W. Liu, L. W. Wang and H. Q. Hou, Endmember Extraction Algorithm on Hyperspectral Remote Sensing Images Based on Chaotic Discrete Particle Swarm Optimization, *Journal of Geomatics Science and Technology*, vol.31, no. 2, pp.148–152, 2014.
[11] K. Sun, X. R. Geng, P. S. Wang and Y. C. Zhao, A Fast Endmember Extraction Algorithm Based on Gram Determinant, *IEEE Geoscience and Remote Sensing Letters*, vol.11, no. 6, pp.1124–1128, 2014.
[12] Y. F. Zhong, L. Zhao, and L. P. Zhang, An Adaptive Differential Evolution Endmember Extraction Algorithm for Hyperspectral Remote Sensing Imagery, *IEEE Geoscience and Remote Sensing Letters*, vol.11, no. 6, pp.1061–1065, 2014.
[13] S. Lopez, J. F. Moure, A. Plaza, G. M. Callico, J. F. Lopze and R. Sarmiento, A New Preprocessing Technique for Fast Hyperspectral Endmember Extraction, *IEEE Geoscience and Remote Sensing Letters*, vol.10, no. 5, pp.1070–1074, 2013.
[14] J. C. Qi, S. L. Zhu, B. S. Zhu and W. Cao, An automatical method for obtaining the number of endmembers based on endmember independence, *Science of Surveying and Mapping*, vol.34, no. 6, pp.206–208, 2009.
[15] C. H. Zhao, B. Qi, and Y. L. Wang, An Improved N-FINDR Hyperspectral Endmember Extraction Algorithm, *Journal of Electronics and Information Technology*, vol.34, no. 2, pp.499–503, 2012.