PCA applied to Data Fusion for Subsurface Target Imaging of Full-polarimetric GPR

Cewen Xue1, 2, Xuan Feng1, 2, 3, 4*, Haoqiu Zhou1, 2, Xiaotian Li1, 2, Wenjing Liang1, 2, Ying Wang1, 2

1 College of Geo-Exploration Science and Technology, Jilin University, No.938 Xi MinZhu Street, Changchun 130026, China;
2 Science and Technology on Near-Surface Detection Laboratory, Wuxi 214035, China
3 Institute of National Development and Security Studies, Jilin University, No.2699 Qianjin Street, Changchun 130012, China
4 Key Laboratory of Geophysical Exploration Equipment, Ministry of Education (Jilin University), No.938 Xi MinZhu Street, Changchun 130026, China
*E-mail: fengxuan@jlu.edu.cn.

Abstract. Full-polarimetric ground penetrating radar (GPR) can obtain more comprehensive polarization data (called VV, HH, VH) for the same target than traditional commercial radar (only VV). We need to use data fusion technology to combine the polarization information of the three different polarization modes. However, the full-polarimetric GPR data fusion method has one weighted average fusion, which will mask the advantages of full polarization. Principal component analysis (PCA) is a technology of data dimensionality reduction and compression which can use VV, HH and VH as a three-dimensional data to conduct data dimensionality reduction and find the best data fusion results. In order to check the reliability, we obtained the full-polarimetric GPR data of three typical targets in the laboratory for analysis. Then we compare PCA with the weighted average fusion method by using the instantaneous amplitude and conclude that PCA can fuse full-polarimetric GPR data better than weighted average fusion.

1. Introduction

Ground-penetrating radar (GPR) is a geophysical method for detecting shallow surface with electromagnetic waves, and it has been widely used in engineering [1], archaeology, hydrology, mine detection [2] and so on. Traditional commercial GPR only can measure single-polarimetric data (VV), while full-polarimetric GPR can obtain more polarization data of the same target, called VV, HH, VH. Therefore, full-polarimetric GPR has obvious advantages in measuring complex targets. The processing methods of full-polarimetric GPR data are probably divided into two categories. One is polarization decomposition technology, which is processed by remote sensing polarization analysis combined with the data of full-polarimetric GPR. For example, Pauli decomposition technique was combined with migration processing [4], H - alpha decomposition was combined with migration processing [5], Freeman decomposition was combined with 3D migration imaging [6]. The other method is data fusion. The three images of VV, HH and VH are fused to make the fused image have all the characteristics of them, and then the fused image is processed by the traditional GPR processing methods to obtain better image of the subsurface target with higher resolution.
Image fusion is a technology for synthesizing image information of the same target from different sensors. By extracting and synthesizing multiple image information, an accurate, comprehensive and reliable image description of the same target can be obtained. Daily et al. first applied the fusion image of radar image and Landsat-MSS image to geologic interpretation in 1979 [3]. In the last 20 years, image fusion has been mainly used in intelligent robot, medicine, remote sensing and so on. For example, a multi-sensor data fusion method based on a Gaussian process model was proposed for complex surface measurements [8], a robust principal component analysis method was used to solve the fusion problem of multi-focus images [9], a perceptual image decomposition model with a local regression method divided images into regular and irregular layers and used different fusion strategies to fuse them [10]. However, there is only a simple weighted average fusion method in the field of full-polarimetric GPR data fusion, and this method will cover up the advantages of full-polarization. So we urgently need a data fusion method which can retain full-polarimetric information.

Principal component analysis (PCA) is a multidimensional orthogonal linear transformation method based on statistical features, which is commonly used for feature extraction of signals and dimensionality reduction of data [7]. PCA was first proposed by Person in 1901[11], and has been widely used in face recognition and network intrusion detection, image noise reduction. For example, Extended Two-Dimensional PCA (E2DPCA) was proposed to improve the accuracy and time of face recognition [12], PCA and Fuzzy PCA can retain the most relevant information from the network traffic data [13], PCA was used to extract features in target classification of SAR images [14]. PCA can obtain the principal component of VV, HH, VH through the maximum variance for full-polarimetric GPR, so as to achieve the effect of data fusion.

In this paper, we select three typical targets in the laboratory to obtain their full-polarimetric GPR data sets, and then principal component analysis (PCA) is used to obtain the fusion images of the three typical targets. The instantaneous amplitude is used to compare the effects of weighted average fusion and PCA fusion, and finally, a conclusion is given.

2. PCA applied to data fusion of full-polarimetric GPR

We can use the VV, HH, and VH GPR systems as different sensors to detect the same subsurface target for full-polarimetric GPR data fusion and obtain VV, VH, and HH data sets. The formula of weighted average fusion is shown as follow:

\[
S = S_{VV} \times P_{VV} + S_{HH} \times P_{HH} + S_{VH} \times P_{VH}
\]  

(1)

where \(S_{VV}, S_{HH}\) and \(S_{VH}\) are data matrices measured by full-polarimetric GPR system. \(P_{VV}, P_{HH}\) and \(P_{VH}\) are weight parameters that balance the contribution among \(S_{VV}, S_{HH}, S_{VH}\) [6].

Principal component analysis (PCA) can transform n-dimensional data into k-dimensional data, and the k-dimensional data is called the principal component. The principal component can represent the original data because it has the largest variance which represents the most features of the original data. So we can use the method of finding the principal components to remove the redundant same parts of the three polarization data and fuse different parts.

In order to use the principal component analysis for the full-polarimetric GPR data \(S = \{S_{VV}, S_{HH}, S_{VH}\}\), three matrices \(S_{VV}, S_{HH}\) and \(S_{VH}\) need to be processed into a full-polarimetric GPR matrix. The steps of principal component analysis algorithm of full-polarimetric GPR data fusion are as follows:

1) Assuming that the dimensions of the \(S_{VV}, S_{HH}\) and \(S_{VH}\) matrices are all \(m \times n\), the full-polarimetric GPR matrix is \(X = [X_{vv}, X_{hh}, X_{vh}]\), \(X_{vv}, X_{hh}, X_{vh}\) are one-dimensional vectors with the length of \(m \times n\), and all of them are obtained by the transformation of \(S_{VV}, S_{HH}, S_{VH}\), respectively.

2) Calculating the mean of data \(X\), \(\overline{X}\) is the mean of matrix \(X\).

\[
\overline{X} = \frac{1}{3} (X_{vv} + X_{hh} + X_{vh})
\]  

(2)

3) Removing average, \(\hat{X}\) is the full-polarimetric GPR matrix \(X\) minus mean matrix \(\overline{X}\).

\[
\hat{X} = X - \overline{X}
\]  

(3)
\[ \tilde{X} = [X_{VV} - \bar{X}, X_{HH} - \bar{X}, X_{VH} - \bar{X}] \]

(4) Constructing covariance matrix V of \( \tilde{X} \), where T is the transpose of the matrix.

\[ V = \frac{1}{3} \tilde{X} \tilde{X}^T \]  \( (4) \)

(5) Eigenvalue decomposition is performed on the covariance matrix V to obtain the eigenvalue \( \lambda = \{ \lambda_1, \lambda_2, \lambda_3 \} \) and the corresponding eigenvector \( W = \{ W_1, W_2, W_3 \} \).

\[ VW = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} W \]  \( (5) \)

(6) Choosing the maximum eigenvalue \( \lambda_{\text{max}} \) and the corresponding eigenvector \( W_{\text{max}} \), then the principal component C is obtained by multiplying \( W_{\text{max}} \) and covariance matrix \( \tilde{X} \).

\[ C = \tilde{X} W_{\text{max}} \]  \( (6) \)

(7) C is a one-dimensional vector with the length of \( m \times n \), and then it is retransformed into a matrix F with the dimension of \( m \times n \). F is the full-polarimetric GPR matrix after data fusion.

3. Experiment and discussion
In order to detect the effect of PCA data fusion and the imaging effect of the subsurface target including three scattering mechanisms, we selected three typical metal targets. Among them, the plate represents single scatter, the dihedral represents double scatter, the multi-branch scatterer represents volume scatter. The length and width of the metal plate are 35cm and 20cm respectively, it is single scatter and relatively simple to reflect. The dihedral is composed of two metal plates arranged at an angle of 90 degrees, which can cause double scatter, especially at the junction. The more of the single scatter toward the edge, the less of the double scatter. The length of the multi-branch scatterer is approximately 40cm, it consists of multiple scaffolds and is an irregularly shaped scatterer, so its scatter is not uniform, and volume scatter occurs. The buried depths are respectively 23cm, 32cm, 25cm. They are buried in dry sand in the laboratory and measured separately, then we obtain the full-polarimetric GPR data of three typical targets. The amount of the measuring points is 99, and the distance between every two measuring points is 1cm, the frequency band ranges from 800MHz to 4500MHz, and there are 1024 sampling points in total.

Then, we preprocess and convert the full-polarimetric GPR data into time-domain data by using inverse fast Fourier transform, and then process the time-domain data with inter-channel equalization processing. The results of the three typical targets after processing are shown in figure 1-3. For the single scattering target, as shown in figure 1, VV and HH polarization can clearly reflect the target position. Although the VH image is a bit messy, it still reflects the target position. For the double scattering target, as shown in figure 2, VV can clearly reflect the reflection profile of dihedral, HH can reflect the position of the double scatter that is the intersection of the dihedral, and VH can reflect the reflection profile of left side dihedral more. For the volume scattering target, as shown in figure 3, the target can’t be seen in VV and HH but the position of the target can be clearly seen in VH.

![Figure 1](image-url)  \( 3 \)

**Figure 1.** Full-polarimetric GPR image of metal plate. (a) VV polarization (b) HH polarization (c) VH polarization.
Finally, we use the full-polarimetric GPR data of the three typical targets to compare the conventional weighted average fusion and PCA fusion. In the weighted average fusion, an average fusion with a weight of 1 is selected for convenience. PCA fusion follows the PCA applied to data fusion of full-polarimetric GPR as mentioned above. Then we use instantaneous amplitude to judge their performance, as shown in figure 4-9. For the convenience of comparison, the color scale and vertical coordinate of instantaneous amplitude are same.

For the single scattering target, as shown in figure 4, the two kinds of fusion methods can clearly reflect target position, because VV occupies more in the principal component, PCA has similar effect with the weighted average. Both of them can see the reflection geometry of the target, and the polarization information of the entire target is single scatter. As shown in figure 5, PCA is larger than the weighted average for the amplitude of the target, so PCA can reflect the position of the target more.

For the double-scattering target, as shown in figure 6, the weighted average only has the weak reflection boundary of dihedral, while PCA can well reflect the position of double scattering which is

\[
\text{Distance(cm)}
\]

\[
\text{Time(ns)}
\]

\[
\text{Distance(cm)}
\]

\[
\text{Time(ns)}
\]

\[
\text{Distance(cm)}
\]

\[
\text{Time(ns)}
\]
the intersection of the dihedral, it also represents the polarization information of the dihedral. PCA can also see the clearer reflection geometry of the target. As shown in figure 7, the instantaneous amplitude of the weighted average has low energy at the junction and the center of the two plates, but PCA highlights the junction of the dihedral and shows the location of the double scatter.

![Figure 6. Data fusion image of metal dihedral. (a) Weighted average fusion (b) PCA fusion.](image)

![Figure 7. Instantaneous amplitude of the data fusion of metal dihedral. (a) Weighted average fusion (b) PCA fusion.](image)

![Figure 8. Data fusion image of metal multi-branch scatterer. (a) Weighted average fusion (b) PCA fusion.](image)

For the volume scattering target, as shown in figure 8, the reflected signal is almost invisible in the weighted average because VV and HH mask target signal that can only be measured by VH. However, PCA can find the real principal component, because the principal component occupied by the volume scattering target is VH, so the location of the subsurface target can be clearly seen. PCA can see obvious and complex reflection geometric feature. As shown in figure 9, the instantaneous amplitude of the weighted average has a little energy, while PCA can reflect the energy concentration of the volume scattering target and find the position of the subsurface target. PCA can see polarization information that is single scatter at the two highest energy peaks and volume scatter at the other messy low-peak.

![Figure 9. Instantaneous amplitude of the data fusion of metal multi-branch scatterer. (a) Weighted average fusion (b) PCA fusion.](image)
4. Conclusion
In this paper, we applied PCA to full-polarimetric GPR data fusion and achieved good results. The fused image has clearer geometric features and the polarization information of single scatter, double scatter or volume scatter. Moreover, we use the instantaneous amplitude to compare the results of the PCA fusion with the weighted average fusion method, finding that PCA can better reflect the positions of the three scattering mechanisms and get higher resolution GPR images.

Acknowledgement
This work was supported by the Science and Technology on Near-Surface Detection Laboratory under Grant 6142414180911, the Technology Development Program of Jilin Province: Suppression of polarization accompanying interference for full polarimetric Ground Penetrating Radar under Grant 20180101091JC.

References
[1] Luca B C, Alessandro C and Fabrizio D 2019 Railway Ballast Monitoring by GPR: A Test-Site Investigation Remote Sens. 11 2381
[2] Feng X and Sato 2004 M Pre-stack migration applied to GPR for landmine detection Inverse Prob. 20 99-115
[3] Daily M I, Farr T, Elachi C 1979 Geologic interpretation from composited radar and Landsat imagery Photogrammetric Engineering and Remote Sensing 45 1109-1116
[4] Feng X, Yu Y, Liu C and Fehler M 2015 Subsurface polarimetric migration imaging for full polarimetric ground-penetrating radar Geophys. J. Int. 202 1324-1338
[5] Feng X, Yu Y, Liu C and Fehler M 2015 Combination of H-alpha decomposition and migration for enhancing subsurface target classification of GPR IEEE Trans. Geosci. Remote Sens. 53 4852-4861
[6] Feng X, Liang W J and Liu C 2017 Application of Freeman decomposition to full polarimetric GPR for improving subsurface target classification Signal Process 132 284-292
[7] Jose C 2000 A Fast On-line Algorithm for PCA and Its Convergence Characteristics IEEE, Transactionson Neural Network 4 29-307
[8] Ding J, Liu Q, Bai M and Sun P 2020 A Multisensor Data Fusion Method Based on Gaussian Process Model for Precision Measurement of Complex Surfaces Sensors 20 278
[9] Wan T, Zhu C and Qin Z 2013 Multifocus image fusion based on robust principal component analysis Pattern Recognition Letters 34 1001-1008
[10] Deng C, Liu X, Chanussot J, Xu Y and Zhao B 2020 Towards perceptual image fusion: A novel two-layer framework Information Fusion 57 100-114
[11] Pearson K 1901 LIII. On lines and planes of closest fit to systems of points in space The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 559-572
[12] Safayani M, Manzuri Shalmani M T and Khademi M 2008 Extended Two-Dimensional PCA for efficient face representation and recognition (Cluj-Napoca: 4th International Conference on Intelligent Computer Communication and Processing) pp. 295-298
[13] Hadri A, Choudgali K and Touahni R 2016 Intrusion detection system using PCA and Fuzzy PCA techniques (Marrakesh: 2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)) pp 1-7
[14] Qiu C, Ren H, Zou H and Zhou S 2009 Performance comparison of target classification in SAR images based on PCA and 2D-PCA features (Xian, Shanxi: 2009 2nd Asian-Pacific Conference on Synthetic Aperture Radar) pp 868-871