Clutter Reduction Based on Principal Component Analysis Technique for Hidden Objects Detection

Václav KABOUREK, Petr ČERNÝ, Miloš MAZÁNEK

Dept. of Electromagnetic Field, Czech Technical University in Prague, Technická 2, 166 27 Prague, Czech Republic
kabouvac@fel.cvut.cz, petr.cerny@fel.cvut.cz, milos.mazanek@fel.cvut.cz

Abstract. This paper brings a brief overview of the statistical method called Principal Component Analysis (PCA). It is used for clutter reduction in detection of hidden objects, targets hidden behind walls, buried landmines, etc. Since the measured data, imaged in time domain, suffer from the hyperbolic character of objects’ reflections, the utilization of the Synthetic Aperture Radar (SAR) method is briefly described. Besides, the basics of PCA as well as its calculation from the Singular Value Decomposition are presented. The principles of ground and clutter subtraction from image are then demonstrated using training data set and SAR processed measured data.

Keywords
Synthetic Aperture Radar, through-wall imaging, Singular Value Decomposition, Principal Component Analysis.

1. Introduction

The non-invasive detection of hidden objects (together with through-wall imaging and buried landmines detection) remains a challenge. Unlike metallic targets, which are easily detectable by techniques grounded on magnetic induction principle, the non-metallic ones are very difficult to locate. The ability of correct and precise detection of these objects depends primarily on factors that are ‘force majeure’ (for instance physical, electric and magnetic properties of mine and its surroundings). Apart from that, there are variable factors that should be selected if reliable results are to be achieved. They embody the used transmitted signals, frequency bandwidth, scanning techniques, post-processing and imaging algorithms. Consequently, various techniques derived from different detection principles have been developed, e.g. for the purposes of landmines detection; [1].

The electromagnetic-based techniques comprise detection methods rested mostly upon variety of radar concepts. For most applications, Ground Penetrating Radar (GPR) is usually used. Depending on a data processing speed and desired quality of final results (2D or 3D images), the signals received by GPR can be measured and processed in various ways. In order to obtain images with high azimuthal (in-line) resolution, the measuring system with synthetic aperture (SA) can be employed. It coherently receives and processes the reflected signals and synthetically extends the length of measuring antenna. There are many suitable algorithms, such as migration techniques (Synthetic Aperture Radar processing, K-F Migration, Kirchhoff Migration), de-convolution and image processing techniques; [2], [3].

For the purposes of this work, the Synthetic Aperture Radar (SAR) processing is applied in order to correct the shape and position of the target. However, the contrast of the investigated target and surroundings is still poor due to their similar electromagnetic properties as well as other clutter contributions. Clutter represents small unwanted reflections caused mainly by rough surface, buried metal, rocks, interfering signals, features incurred by SAR processing, antenna coupling and wall coupling or multiple reflections in case of the through-wall imaging. It is quite common to imaging radars that most of the transmitted energy is reflected back from the first obstruction (wall, ground surface etc.). This phenomenon suppresses targets visibility in the image. In case of even surfaces, the surface reflections can be eliminated by a simple subtracting of the mean image from collected data set. In practice, this method gives sufficient results only in case of flat and even surfaces. Otherwise it is rugged and inhomogeneous, so a more efficient method has to be applied.

As far as the statistical analysis is concerned, the Principal Component Analysis method (PCA) has proven to be a beneficial technique for image compression as well as finding patterns in high-dimensional data, relationship among variables, face recognition and in other fields; [4], [5]. Thus, PCA has also been previously applied as the main image processing tool to landmines detection techniques based on IR, UV and mm-wave sensing; [6], [7]. Yet it is the most frequently applied classification method for pre-processed GPR data [8] - [10]. Furthermore, PCA plays a significant role in the through-wall imaging, owing to its feature extraction property; [11], [12]. Contrary to that, in this paper, the PCA method is combined with other GPR processing method (namely SAR) and the values of the target and clutter components are under scrutiny.
2. Synthetic Aperture Radar Technique

The SAR represents a measurement technique, where the transmitting and receiving antenna, mounted on a platform, is moving along the investigated area and, unlike the Real Aperture Radar (RAR), collects a complex of returned echo waveforms. The acquired data are then subject to processing, while the length of antennas path is mathematically formed as a huge synthetic aperture and, accordingly, the final image shows a better azimuthal resolution. The fundamental SAR processing involves a migration technique that can correct a hyperbolic character of scanned objects by means of spatial frequency conversions, mapping and interpolation.

In case of Stepped Frequency Continuous Wave (SFCW) measurement, it is possible to describe each measured data set for exact antenna positions \( x \) and \( y \) as follows:

\[
E(x, y, f) = E_{Tr}(x, y, f) \sum_{i=1}^{N} \rho_i(f) e^{-j2\pi f_i}
\]

(1)

where \( E_{Tr} \) is the transmitted signal, \( \rho_i \) denotes the amplitude of reflected wave from \( i \)-th scatterer in the distance \( r \), from the antenna and \( k \) represents the wavenumber.

For the SFCW radar, the transmitted signal in the frequency domain takes the form of \( E_{Tr}(x, y, f) = 1 \) and therefore, the equation (1) is simplified. The data subsequently stated in this form are converted to the domain of spatial wavenumbers \( k_x \) and \( k_y \) by 2D Fourier transform and final solution of the transformation is then:

\[
E(k_x, k_y, f) = 
\frac{4 \sigma_{E2} e^{-j\pi/2}}{4k^2 - k_x^2 - k_y^2} \sum_{i=1}^{N} \rho_i e^{-j\sqrt{4k^2 - k_x^2 - k_y^2} \sqrt{\sum_{j=1}^{M} x_i x_j} - j\sum_{j=1}^{M} y_i y_j}.
\]

(2)

For imaging purposes, the amplitude term stated in front of the integral can be neglected; [13]. The aforementioned simplified equation is similar to the 3D point spread function. The correction is performed via mapping function:

\[
k_z = \sqrt{4k^2 - k_x^2 - k_y^2}.
\]

(3)

The conversion in question gives rise to a non-uniform spacing of unevenly distributed data. As a result, the mapping is followed by interpolation back onto the rectangular grid. The corrected data are transformed back to the spatial domain through the 3D inverse Fourier transform:

\[
E(x, y, z) = \iiint E(k_x, k_y, k_z) e^{j(k_x x + k_y y + k_z z)} dk_x dk_y dk_z.
\]

(4)

2.1 Principal Component Analysis (PCA)

The main goal of the PCA is to express the original data set in another domain by means of any appropriate linear transformation; [14]. The latter can be defined as:

\[
Y = SX.
\]

(5)

The original data set \( X \) (where rows denote particular measurements and columns stand for their data samples) is expressed by a new transformed data matrix \( Y \) through the transformation matrix \( S \). The rows of \( S \) represent eigenvectors – principal components of \( X \). The question stipulates the desired form of transformed data \( Y \) and, consequently, stipulates the transformation matrix that ought to be selected. The measured data set can be characterized by the Signal-to-Noise Ratio and data redundancy, which stands for parts of measured signals correlated to each other, i.e. wall reflections, object reflection etc. It leads to the evaluation of correlation properties and, therefore, also construction of the covariance matrix from input data.

In this case, PCA finds a set of orthonormal vectors of principal components (matrix \( S \)) describing the distribution of original data and transforms them into the new set of uncorrelated data.

The cornerstone prerequisite of PCA technique states that the final data \( Y \) ought to be mutually uncorrelated. Hence the covariance matrix of \( Y \) has to be diagonalized, so that the appropriate transformation matrix \( S \) is found. The derivation [15] enables to obtain the principal components in the form of eigenvectors of covariance matrix of the original data \( X \).

The covariance matrix consists of the correlation characteristics (covariance and variances) existing among all pairs of the data set \( X \). The covariance between two measurements measures the degree of mutual similarity. A large absolute value denotes a high redundancy (or correlation) of respective data. Contrary to that, zero covariance points to completely uncorrelated data. The covariance of one measurement represents, at the same time, its variance (diagonal components of covariance matrix). In order to construct the covariance matrix from the data set \( X \), it is firstly necessary to subtract the mean value from each measurement:

\[
\bar{x}_{ij} = x_{ij} - \frac{1}{N} \sum_{j=1}^{N} x_{ij}
\]

(6)

where \( x_{ij} \) denotes the \( j \)-th data sample of the \( i \)-th row (i.e. measurement). Each measurement is centered and new data set takes the following form:

\[
X_0 = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N]^T.
\]

(7)

The covariance matrix can be then expressed as a dot product of the centered matrix \( X_0 \) that is divided by the number of data samples:

\[
C = \text{cov}(X) = \frac{1}{N} X_0 X_0^T.
\]

(8)

Principal components of PCA method are represented by eigenvectors of the covariance matrix \( C \). These can be obtained by means of the Singular Value Decomposition method (SVD) that allows factorizing the data matrix onto its \( A \) eigenvalues and eigenvectors \( U \) and \( V \):
\[ X_{x,n} = U_{x,n} \Lambda_{x,n} V_{n}^T. \] (9)

The left matrix \( U \) represents eigenvectors of the \( XX^T \) matrix, which is in fact identical to the covariance matrix, while \( V \) embraces eigenvectors of \( X^T X \). Accordingly, the principal components of the PCA are evaluated in equation (10) as the matrix \( U \). Yet apart from the principal component, the complete PCA method can be carried out via SVD as well by rewriting the equation (5) in the following manner:

\[ \Lambda V^T = U^T X_{b}. \] (10)

On the right side, the input data are multiplied by the matrix of principal components, so \( U^T \equiv S \). Subsequently, the left side of (10) represents a transformed input data \( X_{b} \) with a zero correlation between particular measurements (\( AV^T \equiv Y \)). Furthermore, in the PCA the eigenvalues from \( \Lambda \) are sorted in descending order. Since each eigenvalue is related to a certain eigenvector (principal component), the matrix \( U \) is reorganized as well. Every eigenvalue represents a specific amount of variance in measured data and thus reflects the importance of particular eigenvector. In fact, the eigenvector (belonging to the largest eigenvalue – first principal component) contributes decisively, among others, to the reconstruction of the original data set, because it shows the highest correlation. The second principal component is correlated with some of the original data that are uncorrelated with the first component. Since the matrix \( S \) of principal components is orthonormal (\( S^{-1} = S^T \)), the original data are obtained by means of the following transformation:

\[ S^T Y = S^T SX_0 = X_0. \] (11)

Based on the selected components, the required data set can be restored only from the selected ones. The reconstruction applying only to \( n \) components \( \{c_1, c_2, \ldots, c_n\} \) is expressed by the following equation:

\[ \bar{X} = S^T Y \bar{c} = \left[ \begin{array}{c} \vdots \\ s_{c_1} \\ \vdots \\ s_{c_2} \\ \vdots \\ \vdots \\ s_{c_n} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right] \left[ \begin{array}{c} \cdots y_{c_1} \\ \cdots y_{c_2} \\ \cdots y_{c_2} \\ \vdots \\ \vdots \end{array} \right]. \] (12)

Dimension of the reconstructed matrix \( \bar{X} \) is identical to the one of the original matrix \( X \).

3. Hardware Setup and Measurement

To obtain the testing data set, the SFCW radar measurement in monostatic configuration was performed. At each of its positions, the utilized double ridged horn (DRH) antenna transmitted a set of harmonic waves within the frequency range from 45 MHz to 26 GHz at 401 frequency points. A wooden box filled with dry homogeneous sand was selected as a surveyed medium. The data were collected from the area 0.5 meters long and 0.5 meters wide with the help of the 2D scanner and Agilent E8364A Vector Network Analyzer (VNA). The antenna was placed 150 mm above the sand surface, see Fig. 1. The complex parameter \( F_{\text{meas}}(\theta) \) was collected for imaging purposes. The parameter \( F_{\text{meas}}(\theta) \) is linearly proportional to the ratio of the \( E(x,y,f) \) to \( E_{\text{RX}}(x,y,f) \) through the proportional constant called Antenna Factor and Transmit Antenna Factor. In addition, the own antenna reflection coefficient is additionally included in the measured reflections. It should be noted that the antenna reflections are eliminated as a mean value using (6). Three different measurements were undertaken. First, a metallic round-shaped target was buried approximately 50 mm under the surface and then replaced with a plastic anti-personal (AP) mine and measured in two different depths (50 and 80 mm). The measurements were proposed so that it was possible to obtain three dimensional images (C-scans). The collected data were processed and computed in MATLAB.

4. Experimental Results

4.1 SAR Processing

The results of SAR technique allow more accurate detection and localization of hidden objects. Given the data obtained from the GPR measurement, the images reconstructed from the raw data show, unfortunately, a hyperbolic character of detected objects.

Since the data were collected in the frequency domain, the most trivial way to display them is to convert them into the time domain through the inverse Fourier transform. For easier interpretation, the C-scan is normalized to 1. Only useful reflections were subject to imaging, i.e. there was a threshold value equal to 0.029, while the data values not reaching it were neglected. The threshold value was set experimentally from the 2D cuts of data sets obtained from different measurements. The raw C-scan of metal object is depicted in Fig. 2.

Fig. 2 demonstrates a typical hyperbolic character of the measured data. Neither the shape nor the position of the object can be correctly determined.

Nonetheless, better results are attained, provided that the 3D SAR processing is utilized. As the SAR method eliminates the hyperbolic reflection effect, both the shape and size of the image are similar to a real object. The situation is demonstrated in Fig. 3. The threshold value is the same as in the first case.
The 3D SAR method represents a time-consuming process and is ineligible for real-time processing. In most cases, only the 2D imaging is practically used. Accordingly, the 2D SAR processed data were considered for clutter reduction.

### 4.2 PCA Processing

Following section 2.1, the PCA can be employed for clutter and surface reflection reduction. The crucial contribution of this method consists in its apparent suitability for both, subsurface and through-wall survey and detection. Prior to the utilization of the PCA, initial issues listed below have to be taken into consideration:

- The surface of explored area is nearly even and most of the energy is reflected straight from the surface. The given reflection creates the largest contribution to the image. It represents the first principal component of the PCA.
- Other considerable part of the image stands for the reflection of searched target and, indeed, constitutes the second principal component of the PCA method.
- In ideal case, the other principal components embody unimportant ‘noises’, such as weaker reflections from smaller objects and other multiple reflections.

These assumptions are valid only in an ideal case. In reality, the surface is non-even. In addition, in the PCA domain, not only is the surface reflection represented by the first component, but it is also determined by a finite number of the most important principal components. The reflection from the target is then characterized by another or even several components.

#### 4.2.1 Training Data Set

To validate the statements mentioned in section 4.2, a simple simulation of the metallic target behind a wall was carried out in Transient solver of CST Microwave Studio. The modeled scenario is depicted in Fig. 4. The metallic 20 mm thick target with a radius of 40 mm is located 80 mm behind a 100 mm thick concrete wall. The complex data (A-scans) are collected along the 0.5 m long wall from the distance of 40 mm. Its dispersive permittivity, approximated by the first order of Debye model, is $5.5 + 0.11j$ at the frequency 3.1 GHz. The frequency bandwidth was set from 1 to 12 GHz.

The collected data can be described from the simplified signal flow graph of the modeled scenario; see Fig. 5. The $S$ parameters stand for parameters of the antenna, $\Gamma_{\text{air}}$ represents the reflection from air, exponential terms denote the wave propagation and $\Gamma_{\text{obj}}$ includes reflections from the wall and target. The measured data are then expressed as the total reflection coefficient $\Gamma_{\text{meas}}$:

$$\Gamma_{\text{meas}} = S_{11} + \frac{S_{21}S_{12}(\Gamma_{\text{air}} + \Gamma_{\text{obj}}e^{-jkr})}{1 - S_{22}(\Gamma_{\text{air}} + \Gamma_{\text{obj}}e^{-jkr})}.$$  \hspace{1cm} (13)

Since $S_{11}$ and $S_{22}$ describe the connector and aperture reflections of the antenna, their influence can be removed from the data set through the process of time-gating.

Furthermore, $\Gamma_{\text{air}}$ can be considered as equal to zero. The denominator can be considered as equal to one due to its negligible exponential term. In the remaining relation, $S_{21}S_{12}$ describes the antenna transmission parameters and the exponential term represents the wave propagation (in both directions). To extract the $\Gamma_{\text{obj}}$ from this simplified equation, the second simulation (concerning only two
antennas) was carried out. The separation equals to the distance between the antenna and metallic target. The signal flow graph of this scenario can be seen in Fig. 6.

From this signal graph, the total transmission coefficient is computed and simplified as in the first case and, given the matched receiver ($\Gamma_L = 0$) and negligible backward transmission (exponential term), the transmission coefficient takes the final form as follows:

$$b/a \approx S_{21}S_{12}e^{-jk\rho}.$$  

Assuming constant propagation velocities in different environments (air, wall), the simulated data $\Gamma_{\text{meas}}$ can be corrected by the data described by (14).

![Fig. 7. 2D SAR image of the target behind concrete wall.](image)

![Fig. 8. Corrected 2D SAR image of the target behind concrete wall.](image)

Raw and corrected data were firstly processed in MATLAB by the SAR technique. The results are depicted in Fig. 7 and 8. The depth shift of corrected data arises from the exponential term of (14). The ripple on both images originates from the performed SAR algorithm and can be extracted via PCA. As a result, the processed corrected data are converted into the Principal Component domain using (10). Given the strength of wall reflections, it is presumed that the first component could only restore the wall reflection, so the second component corresponds to the target. The assumption can be proved exclusively by the image restoration from the first two components with the help of (12); see Fig. 9. The other smaller components are constituted by clutter, noise or reflections of insignificant importance. The shape of detected object is obviously restored better using (13-14). The vertical axis depicts time-related samples. Due to the difference between the propagation velocities in respective environments (air or wall), the axis corresponding to the dimension is disproportional to the time. The boundaries between air and wall are highlighted in Fig. 7 and 8 by the red arrow.

![Fig. 9. Reflection of the metallic target and its surface reconstructed by PCA; 1st and 2nd components.](image)

**4.2.2 Measured Data Set**

In the next step, the PCA technique was applied after 2D SAR processing in three different measurements (from $\Gamma_{\text{meas}}$) using a metallic round-shaped disc and a plastic mine. The decomposition is demonstrated only for the case of the AP mine measurement. Unlike the raw data, the ones processed with SAR are corrected and the hyperbolic character of mine reflection is suppressed; see Fig. 10. Even if the position of mine is more precise, its contrast with surrounding remains poor.

![Fig. 10. 2D SAR image of the plastic landmine under surface.](image)

As the simulated results show, an even surface can be expressed only by the first principal component. In case of the given measured data (Fig. 10), four most important principal components are to be utilized in order to restore the surface reflection; see Fig. 11. The reconstructed surface takes almost the same form as in Fig. 10.

As it was mentioned before, the second largest part of the measured data should be embodied in the landmine reflection. Accordingly, the next (fifth) component is incorporated into its reconstruction process. The rest of com-
Fig. 11. PCA reconstruction of the sand surface from 2D SAR; first four principal components are used.

Fig. 12. PCA reconstruction of the plastic landmine from 2D SAR; fifth principal component is employed.

Fig. 13. Sum of the reconstructed surface and landmine.

4.2.3 Principal Components Selection

Every principal component belongs to its given eigenvalue. The size of each eigenvalue is related to the variance or correlation of the part of the original data represented by the appropriate component. In other words, the biggest eigenvalue generally creates the most significant pattern in the imaged data. Consequently, owing to the knowledge of eigenvalues size, the target can be extracted from the whole image. Based on the given assumptions and results, the target reflection is represented by principal components, whose eigenvalues are ten times lower than the largest one. The eigenvalues exceeding the threshold stand for either the surface or wall backscatter. Clearly, the image can be reconstructed by less than 20 out of 300 components, without losing any important information.

It should be noted that the threshold amounting to 0.1 is valid for the presented set of results and similar scenarios. It is typical to express the eigenvalues as a percentage of the total, so they are normalized to the biggest one. The size of the target components varies with the target size, reflectivity and, to a certain extent, also with the ratio between the target dimensions and overall dimensions of the tested area. The results from the metallic target measurement reveal that the reflection of the disc can be restored from the fourth and smaller components. In comparison, the reflectivity of an AP mine located 8 and 5 cm under the surface is smaller, therefore the target can be determined from the fifth component (see measurements 2 and 3 in Fig. 14).

Fig. 14. Normalized eigenvalues of the measured and simulated data.

5. Conclusions

In this paper, the relation among SVD, PCA and PCA applications in the field of the hidden objects detection was presented. The measurements were performed in laboratory conditions and the complex reflection parameters were collected by means of VNA. Prior to the utilization of PCA decomposition, the SAR processing was applied in order to correct phases of the acquired datasets. As the SAR method enabled to suppress the hyperbolic effect, the target shape dimensions were enhanced and the clutter removal the PCA subtraction was proposed. Additional enhancement was performed using antenna parameters corrections. Besides, in order to prove and verify the functionality and given the hypothesis of the PCA, the simulation of the metallic target behind a concrete wall was carried out. It
was demonstrated that the correlated data could be decomposed into a set of uncorrelated components sorted by their importance. It turned out that each component carried information about the (correlated) patterns in the original signal sets (image). Unlike few last components that created only a noise and almost/nearly uncorrelated data, the most important (and also the first principal) component represented the strongest or the most distinct pattern. These considerations were confirmed by image restoration with only the first two components, where the most important one stood for wall reflections and the second one for the target reflection. It should be emphasized that this rule cannot be generalized (see measurements).

With respect to the results, the PCA was applied to the real SAR processed data. The ground was non-even and inhomogeneous and the contrast of the detected target was low. It led to the necessity to restore the ground reflections from more than one component. As a consequence, the target had to be represented by the next one.

The target extraction based on the size of data eigenvalues was shown via simulation and three measurements. The enhanced image could be restored only from the principal component representing the target reflection so that its eigenvalue amounted approximately to one tenth of the largest one. The larger components denoted the surface reflection, whereas the smaller ones represented other clutter.

On the other hand, in practice, it is highly complicated to set up the threshold value and to set correctly both the number of principal components creating unwanted reflection and components constituting the searched target.

Acknowledgements

The research is part of activities of the Department of Electromagnetic Field of the Czech Technical University in Prague. It was carried out within the research project of the Czech Science Foundation No. 102/09/P536. We would also like to express gratitude to Iveta Černá for the proof-reading of the paper.

References


About Authors...

Václav KABOUREK was born in 1985. He received his M.Sc. degree from the Czech Technical University in Prague in 2010. His research interests include UWB communication, RCS measurement, microwave detection of non-metallic object, analysis and development of detection techniques and signal transformations in MATLAB.

Petr ČERNÝ was born in 1976. He was awarded his M.Sc. and Ph.D. degrees by the Czech Technical University in Prague in 2001 and 2008, respectively. His contemporary research activities are focused on high-resolution microwave and terahertz spectroscopy, microwave circuits and antennas, ultra wideband devices and die bonding. He is a member of IEEE and Radioengineering Society.

Miloš MAZÁNEK was born in 1950. He received his M.Sc. and Ph.D. degrees from the Czech Technical University in Prague in 1974 and 1980, respectively. He has been a head of the Dept. of Electromagnetic Field since 1997. He is a senior member of IEEE, head of the Radioengineering Society and Radioengineering journal executive editor. His research interests are aimed at antennas, EMC, microwave radiometry and propagation.