Hyperspectral Anomaly Detection Method Based on Auto-encoder

Emrecan Batı*, Akın Çalışkan*, Alper Koz* and A. Aydm Alatan*†

*Center for Image Analysis, Middle East Technical University, Ankara, Turkey;
†Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey

ABSTRACT

A major drawback of most of the existing hyperspectral anomaly detection methods is the lack of an efficient background representation, which can successfully adapt to the varying complexity of hyperspectral images. In this paper, we propose a novel anomaly detection method which represents the hyperspectral scenes of different complexity with the state-of-the-art representation learning method, namely auto-encoder. The proposed method first encodes the spectral image into a sparse code, then decodes the coded image, and finally, assesses the coding error at each pixel as a measure of anomaly. Predictive Sparse Decomposition Auto-encoder is utilized in the proposed anomaly method due to its efficient joint learning for the encoding and decoding functions. The performance of the proposed anomaly detection method is both tested on visible-near infrared (VNIR) and long wave infrared (LWIR) hyperspectral images and compared with the conventional anomaly detection method, namely Reed-XiaoLi (RX) detector. The experiments have verified the superiority of the proposed anomaly detection method in terms of receiver operating characteristics (ROC) performance.

Keywords: Hyperspectral image, Anomaly detection, Representation learning, Auto-encoder, RX anomaly detector

1. INTRODUCTION

Hyperspectral target detection studies are carried on evolving in two distinct categories. In anomaly detection methods, which can be considered as the first category, the detection is assessed by the spectral difference between the hyperspectral pixel under test and its surrounding. On the other hand, spectral signature based target detection methods forming the second category do not compare the pixel with its surrounding but instead with a previously available spectral signature. Compared to the signature based methods, the anomaly detection methods have become more preferable in many applications where prior information is not available.

Anomaly can be defined as the data or observation which does not fit into the general characteristics of a given data set. Among anomaly detection methods, Reed-XiaoLi (RX) method has taken a significant attention in the literature with its pioneer aspect and simplicity. The method utilizes the Mahalanobis distance metric to describe the discrepancy between the target pixel and the background distribution to differentiate the abnormal objects. The algorithm stands as a benchmark anomaly detector, on which several modifications have been implemented so far. Kernel-RX and Local-RX are two examples of such modifications. Kernel-RX makes use of kernel trick to implicitly transform data into a higher dimensional space and calculate Mahalanobis distance in that space, in single step without computing the transformed data. Local-RX models the background distribution from pixels spatially close to the pixel under test instead of using the whole hyperspectral image.

A major challenge in these anomaly detectors is an efficient representation of background. When varying complexity of hyperspectral scenes is considered, using a single Gaussian distribution to model the background as in RX method would not be sufficient. Although the complexity of the background can be alleviated by Local-RX methods, a drawback of such methods is to decrease the number of samples necessary to learn the background. Furthermore, the selection of the suitable kernel for Kernel-RX methods is always not a simple task for the scenes composed of many different materials.

Further author information: (Send correspondence to E.B.)
E.B.: E-mail: ebati@metu.edu.tr
In this paper, we first propose to represent the hyperspectral scenes of different complexity with the state-of-the-art representation learning method, namely auto-encoder rather than the Gaussian distribution based approaches. Auto-encoders are until now used in representation learning as a building block of the deep learning architectures with applications for natural language processing, text retrieval, speech recognition and many others. Auto-encoders learn to efficiently encode a given input in a nonlinear manner to minimize reconstruction error. Compared to its linear counterparts such as principle component analysis (PCA), independent component analysis (ICA) and minimum noise fraction (MNF), auto-encoders have been verified to give better performances with its sparse-overcomplete representation. In such a representation, the anomalies are expected to yield greater reconstruction error due to their rareness. Therefore, the reconstruction error can be regarded as an anomaly metric. Based on this idea, we secondly propose in this paper a novel anomaly detection method depending on the proposed auto-encoder based representation of hyperspectral image.

The rest of the paper is organized as follows. The proposed method is explained in detail in Section 2. We then present the experimental setup and discuss the experiment results in Section 3. Finally, the conclusions are drawn in Section 4.

2. PROPOSED ANOMALY DETECTION METHOD USING AUTOENCODER

Proposed method consists of two stages (Figure 1a). In the first stage, described as learning phase, each hyperspectral pixel, $Y_i$, of the input test image is encoded into a code, $Z_i$, and then decoded to reconstruct the original pixel.

The parameters of the encoding and decoding functions are tuned to minimize the cumulative of the energies defined for each spectral pixel. We adopt the auto-encoder using the unified energy based framework for the encoding and decoding functions. Specifically, predictive sparse decomposition auto-encoder is adapted for the multidimensional hyperspectral data in the developed method. In the second phase, the reconstruction error at each pixel with the learned parameters is assigned to that pixel as a measure of being anomaly.

Before going into the description of the energy function and the learning algorithm, we first present the utilized notation in the paper in parallel with Figure 1b.
• $Y_i$: $i^{th}$ pixel of the hyperspectral image cube of dimension $h \times w \times p$. Note that this corresponds to a one dimensional vector with size $p$.

• $Z_i$: The output of the encoder function with a size $n_z$ for the input pixel $Y_i$.

• $Z_i^*$: The optimal code sought for input pixel $Y_i$.

• $\hat{Y}_i$: Reconstructed pixel from the code $Z_i$.

• $g_e(Y_i, W_e, D)$: Encoder function which transforms the input $Y_i$ to the code $Z_i$ with the parameters $W_e$ and $D$. $W_e$ corresponds to a matrix of size $n_z \times p$ and $D$ is a diagonal matrix of size $n_z \times n_z$. For predictive sparse decomposition auto-encoder,$^{12}$ $g_e(Y_i, W_e, D)$ is selected as

$$g_e(Y_i, W_e, D) = D \tanh(W_e Y_i).$$  

(1)

• $g_d(Z_i, W_d)$: Decoder function which reconstructs the pixel $\hat{Y}_i$ from the code $Z_i$ with the parameter $W_d$. $W_d$ is a matrix of size $p \times n_z$. For predictive sparse decomposition auto-encoder,$^{12}$ $g_d(Z_i, W_d)$ is selected as:

$$g_d(Z_i, W_d) = W_d Y_i.$$  

(2)

The energy function that would be minimized for hyperspectral pixels during the learning phase consists of three terms (Figure 1b). The first term contributing to the energy function is the encoder energy which describes the distance between the obtained code and the optimal code. As the encoder energy gets smaller the derived code would be more closer to the optimal code. The second term is defined as the decoder energy which indicates the success of reconstruction of the hyperspectral pixel from the code $Z_i^*$ as in Figure 1b. As the decoder energy gets smaller, the reconstructed pixel would be more closer to the original pixels. The final term code energy is $L_1$ norm of the generated code which enforces the sparsity of the code. The overall energy of the system is given as$^{12}$

$$E(Y_i, Z_i, W_e, D, W_d) = \alpha \|Z_i - g_e(Y_i, W_e, D)\|_2^2 + \|Y_i - g_d(Z_i, W_d)\|_2^2 + \beta \|Z_i\|_1. \tag{3}$$

The proposed algorithm to find the optimal code, $Z_i^*$, which minimizes the energy function in (3) is as follows:

1. Initialize randomly the parameters $W_e$ and $D$ of the encoder function and $W_d$ of the decoder function.

2. For a predefined number of iterations,

   • Randomly select a group of $k$ pixels from all the pixels of the hyperspectral image.
   • Calculate the code $Z_i$ for each hyperspectral pixel, $Y_i$, in the selected group.
   • Search $Z_i^*$ by minimizing the energy in (3) with respect to $Z_i$ while $W_e$, $D$ and $W_d$ are fixed. In the current implementation the search method is selected as gradient descent for the simplicity.
   • After the optimal code $Z_i^*$ is found, fix $Z_i$ to $Z_i^*$ and update $W_e$, $D$ and $W_d$ to minimize the average of the energy in (3) calculated over all the pixels in the selected group. The update is performed by using the gradient descent method. $W_d$ is normalized to a unit norm after the update so as to avoid the situation where $W_d$ and $Z$ are multiplied and divided by the same constant to decrease the energy in (3).

3. After the iteration stage, which yields the learned parameters $W_e$, $D$ and $W_d$, calculate code $Z_i$ by using encoding function $g_e$.

4. Assign reconstruction error, $\|Y_i - W_d Z_i\|_2^2$, to each pixel as a measure of being anomaly for that pixel.

In the proposed algorithm encoding function $g_e$ is selected as $D \tanh(W_e Y)$ and decoding function $g_d$ is selected as $W_d Y$ as suggested by Kavukcuoglu et al.$^{12}$ The number of iterations are adjusted so as to guarantee an early-stopping before the average energy enters to the saturation. The weighting factor $\alpha$ is adjusted to 1 as suggested by Kavukcuoglu et al.$^{12}$ We choose the $\beta$ value experimentally as 0.2. The code length $n_z$ is selected as a design parameter to judge the performance of the algorithm.
3. EXPERIMENTAL RESULTS AND COMPARISONS

The experiments are conducted with two images, one acquired by visible-infrared (VNIR) camera and the other by long-wave-infrared (LWIR) camera (Figure 2). The VNIR camera has a spectral range of 400-1000 nm and LWIR camera has a spectral range of 7800-11500 nm. The captured hyperspectral VNIR and LWIR images are of the size $885 \times 1500 \times 182$ and $180 \times 266 \times 121$, respectively. The scene for the VNIR image contains different fabrics and construction supplies which are placed on a rural area. The scene is captured from a height of 500 m. The scene for the LWIR image contains a piece of carpet, parquet, polyethylene foam, raincoat, gray aluminum and black aluminum captured from a distance of 10 m on the ground. The images have ground truth data for testing their performances.

The performance of the proposed algorithm is first evaluated with respect to the size of the code, $Z$. Based on this evaluation a suitable value of $Z$ is selected to compare the proposed algorithm with the base line algorithms in the literature. The methods for comparisons are selected as the variants of Reed-Xioli (RX) algorithm, namely Global-RX algorithm\(^1\) and dual window Local-RX algorithm.\(^4\) The comparisons are performed over Receiver Operating Characteristic (ROC) curves which indicates true positive rate (TPR) with respect to the false positive rate (FPR). TPR and FPR are defined as

\[
TPR = \frac{TP}{N_{anom}}, \tag{4}
\]

\[
FPR = \frac{FP}{N_{bg}}. \tag{5}
\]

where $TP$ represents the anomaly pixels detected given a certain threshold; $N_{anom}$ represents the total number of the anomaly pixels in the image; $FP$ represents the number of pixels detected as anomaly from the background; and $N_{bg}$ is the number of background pixels.

Figure 3 illustrates ROC curves for the proposed algorithm with different code sizes, $n_z$. As the code size increases, the performance of the algorithm increases as well. This can be reasoned by the more expressive capacity of the auto-encoder with the larger code size. The outputs of the proposed algorithm for the code size of 8 for VNIR hyperspectral image and for code size of 64 for LWIR image is given in Figure 4 and LWIR images.

Figure 5a illustrates generated code vectors, $Z_i$, belonging to anomaly pixels and code vectors corresponding to background pixels in a two dimensional graph for the code size, $n_z = 2$, in order to explore the distribution of the code vectors in their represented space. The same illustration is also given for the code size, $n_z = 3$, in a three dimensional graph given in Figure 5b. Compatible to the concept of anomaly, the pixels belonging to the anomaly are concentrated on the far edges of the graph, which indicates the success of the proposed auto-encoder in filtering the anomalies out.

Figure 6 presents the comparison of the proposed algorithm with Global-RX algorithm\(^1\) and dual window Local-RX algorithm.\(^4\) While the proposed algorithm gives the best ROC performance, it is followed by the
dual window Local-RX. Global-RX yields the worst performance among the three techniques. As the proposed algorithm can cope with the scene variability with its adaptable characteristics and tunable code size, it catches the dominant and major spectral signatures in the hyperspectral data quite well, while leaving the remaining pixels as anomalies.

4. CONCLUSIONS

Considering that an efficient background representation is one of the major challenges in the existing anomaly detection methods, we propose an auto-encoder based anomaly detection method which can represent the scenes of different complexity with its inherited learning capabilities. The experimental results first verify that the selection of the code size has a direct influence on the performance of the method. Secondly, the superiority of the proposed anomaly detection method over the baseline RX variants is revealed in terms of receiver operating characteristics (ROC) performance.

A basic trade-off in the proposed learning based anomaly detection method is the level of the learning versus the capability of the reconstruction error to represent the anomalies. In other words, as the auto-encoder begins to memorize, instead of learning the scene, it might be expected that the anomalies would also begin to dissipate. Therefore, the code size, the encoder and decoder functions and other parameters related to the learning capacity of the auto-encoder have emerged as critical selections in the proposed method. The future work will focus on...
Figure 5: Illustration of generated code vectors (a) for code size of 2 and (b) for code size of 3. Red and green colors correspond to background and anomaly pixels, respectively.

Figure 6: ROC curves for the proposed algorithm with Global-RX algorithm and dual window Local-RX algorithm (a) for VNIR image and (b) for LWIR image.

a detailed analysis of these factors as well as the comparisons with the other state-of-the-art anomaly detection methods.

ACKNOWLEDGMENTS

This research was partially supported by HAVELSAN Inc. and by Scientific Research Project Center of METU under the project id BAP-08-11-2015-005.

REFERENCES


