Support Vector Machine for Target Detection in Hyperspectral Images

Davood AKBARI, Abdorreza SAFARI, Iran

Key words: Hyperspectral image, Target Detection, SVM, ROC curve

SUMMARY

Hyperspectral Images are worthwhile data for many processing algorithms (e.g. Dimensionality Reduction, Target Detection, Change Detection, Classification and Unmixing). Target detection is a key issue in processing hyperspectral images. Spectral-identification-based algorithms are sensitive to spectral variability and noise in acquisition. In most cases, both the target spatial distributions and the spectral signatures are unknown, so each pixel is separately tested and appears as a target when it significantly differs from the background. On the other hand, there are many (e.g. Modified Spectral Angle Similarity (MSAS) as a Deterministic and Covariance-based Matched Filter Measure (CMFM) as sub-pixel approach) algorithms for target detection. As a new algorithm, Support Vector Machine (SVM) is a useful technique for Target Detection.

In this paper, first we propose a theoretical discussion aimed at understanding and assessing the potentialities of MSAS, CMFM and SVM algorithms in hyper-dimensional feature spaces. Then, we assess the effectiveness of SVM with respect to conventional. To sustain such an analysis, the performance of SVM is compared with those of two other Target Detection algorithms, one-against-all, the one-against-one. Finally, Different performance indicators have been used to support our experimental studies in a detailed and accurate way (i.e., Target Detection accuracy, the computational time, the stability to parameter setting).

The results obtained on a real Visible/Infrared Imaging Spectroradiometer hyperspectral dataset (CASI) allow concluding that, SVM is a valid and effective alternative to conventional Target Detection algorithms of hyperspectral remote sensing data.

چكيده

فنآوري سنجش از دور فراطيفي، در دو دهه گذشته پيشرفت چشمگيري داشته است. اين پيشرفت در طراحي و ساخت سنجنده‌ها و همچنين در توسعه و پياده‌سازي روش‌هاي پردازش داده‌ها، بسيار مشهود بوده است. در مقايسه با تصاوير چندطيفي، تصاوير فراطيفي داراي حدتفكيک طيفي بالاتري بوده از اين رو منحني طيفي مربوط به هر پيكسل نسبت به منحني طيفي پيكسل‌هاي تصاويرچندطيفي پيوسته‌تر مي‌باشد، بنابراين از تصاويرفراطيفي مي‌توان به عنوان منبع مناسبي جهت نقشه‌برداري مواد استفاده نمود.
ويژگي‌ها و پيچيدگي‌هاي داده‌هاي حاصل از سنجنده‌هاي فراطيفي باعث شده است كه روشهاي نوين و پيشرفته‌تر آناليز تصاوير سنجش از دور به منظور استخراج اطلاعات دقيقتر و كاملتر از داده‌هاي فراطيفي مورد توجه قرار گيرند، يكي از

آناليزهايي كه بر روي تصاوير فراطيفي انجام مي‌شود، آشكارسازي هدف است. روشهاي آشكارسازي هدف در تصاوير فراطيفي، معمولاً بر اساس ويژگي‌ها و اطلاعات طيفي صورت مي‌گيرد.الگوریتم‌های مختلفی برای این منظور وجود دارد که از جمله آنها می توان به دسته الگوریتمهای کلاسیک و غیر کلاسیک اشاره کرد.
در این مقاله ابتدا راجع به سه الگوریتم اندازه‌گیری زاویه طیفی، اندازه‌گیری انطباقی مبتنی بر کواریانس و ماشین بردار پشتیبان مطالبی بیان شده سپس الگوریتم ماشین بردار پشتیبان با سایر الگوریتم‌ها مقایسه گشته و بررسی های مربوطه بر روی تصویر فراطیفی CASI پیاده سازی گردید.
بررسي‌هاي انجام شده بهبود دقت قابل ملاحظه‌اي ماشین بردار پشتيبان را نشان مي‌دهد.

Support Vector Machine for Target Detection in Hyperspectral Images

Davood AKBARI, Abdorreza SAFARI, Iran

1. INTRODUCTION

Recent advances in hyperspectral sensors with high spectral and spatial resolution have led to an increased interest in exploiting spectral imagery for target detection. Given the availability of spectral libraries for a wide range of materials, detection algorithms that exploit a known target signature have been widely investigated. It has been shown (Scharf and Friedlander, 1994) that such algorithms are dependent on the degree of signal mismatch between the spectral libraries and the spectra observed in an image.

Automatic target recognition (ATR) has experienced significant strides with the advent of hyperspectral imaging (HSI) sensors. ATR systems should be able to detect, classify, recognize, and/or identify targets in an environment where the background is cluttered and targets are at long distances and may be partially occluded, degraded by weather, or camouflaged (Yamany, Farag and Hsu, 1999). HSI sensors provide plenty of spectral information to uniquely identify materials by their reflectance spectra. A material’s reflectance spectrum contains the reflectance values of the material as a function of wavelength. Although it is theoretically possible for two completely different materials to exhibit the same spectral signature, targets in ATR applications are typically man-made objects with spectra that differ considerably from the spectra of natural background materials (Manolakis and Shaw, 2002).

In HSI target detection applications, the targets are sparse and typically occupy less than 1% of the total pixels in a hyperspectral (HS) scene, rendering traditional spatial processing techniques impractical. Consequently, most HSI detection algorithms exploit the spectral information of the scene, an approach known as nonliteral exploitation in the HSI literature (Manolakis, Marden and Shaw, 2003). One of the main challenges in HSI processing is spectral variability, which refers to the phenomenon that the spectra measured from samples of the same material will never be identical. In other words, spectra of the same material are not fixed due to the inherent variations present in the material. Further spectral variability is introduced by external factors such as atmospheric conditions, sensor noise, and illumination variations (Shaw and Burke, 2003).

Although many detection algorithms have been developed over the years, spectral variability poses challenges for these algorithms. While the stochastic detectors are mathematically tractable and can work well in some situations, they are only optimal under the assumption of the multivariate normality of the data. The quadratic Neyman–Pearson detector requires the covariance matrix of the target class, which is not available if one is given a single spectral signature obtained from a library (Vapnik, 1998). In real-life scenarios, the multivariate normality assumption is often violated because an HS image may contain multiple types of terrain, thus causing detection performance to suffer (Henz and Wagner, 1997).

Kernel methods have become increasingly popular in a variety of pattern recognition (PR) applications. The recently-developed support vector machine (SVM) has its roots in statistical learning theory and is an emerging nonparametric approach for describing a set of data (Tax and Duin, 2004). It has been successfully applied in the areas of facial expression analysis, gene expression data clustering, image retrieval and remote sensing image classification.

In this letter, we will use the SVM to perform target detection in HS imagery. Experiments on urban HSI scenery confirm that the proposed SVM-based method can provide substantially lower false positive rates (FPRs) while maintaining higher true positive rates (TPRs) when compared to other detectors. Section II provides formulation of methods. Section III provides Accuracy Evaluation Method. Section IV provides the experiments and results, and conclusions and future work are discussed in Section V.

2. SUGGESTED METHODS

2.1 Modified Spectral Angle Similarity (MSAS)

Given two vectors as the target and pixel spectra, a spectral angle between this pair of vectors can be defined. In the case of a hyperspectral image, the "hyper-angle" is calculated with:

$$\alpha = \cos^{-1}(s_i . s_j \ / \ \| s_i \| \| \| \ s_j \|) = \cos^{-1}(\sum_{l=1}^{L} s_{il} s_{jl} \ / [\sum_{l=1}^{L} s_{il}^2]^{1/2} [\sum_{l=1}^{L} s_{jl}^2]^{1/2}) \qquad \text{Eq.1}$$

The smaller angle means more similarity between the pixel and target spectra. Here, we prefer to use a modified spectral angle. In above equation α is between 0 and π/2, so we can easily obtain, MSAS= $\frac{2\alpha}{\pi}$, by this rescaling the values of measure convert to [0, 1].

2.2 Covariance-based Matched Filter Measure (CMFM)

Covariance-based Matched Filter Measure, CMFM, is one of the anomaly detection methods. The aim of anomaly detection is to search and find unknown targets with low probability of existence in the image. The anomaly detection works based on properties of covariance or correlation matrix of the target.

In Eq. 2. CMFM measures similarity of targets si and sj (s includes spectral properties) after reducing their mean μ. More similarity value means both targets are members of same class in higher probability.

$$CMFM = (s_i - \mu)^T K_{L\times L}^{-1}(s_j - \mu) \qquad \text{Eq.2}$$

In which, $K_{L\times L}^{-1}$ is inverse of image covariance matrix and L is number of image bands. For pixel with CMFM value closer to unit, it means the pixel is more similar to the target.

2.3 Support Vector Machine (SVM)

We will first define the hard margin SVM, applicable to a linearly separable dataset, and then modify it to handle non-separable data. The maximum margin classier is the discriminate function that maximizes the geometric margin $\frac{1}{\|w\|}$ which is equivalent to minimizing $\|w\|^2$. This leads to the following constrained optimization problem: $\min imize_{w,b}$ $\frac{1}{2}\|w\|^2$ Subject to: $y_i(w^T x_i + b) \geq 1 \quad i = 1,...,n$. The constraints in this formulation ensure that the maximum margin classier classifies each example correctly, which is possible since we assumed that the data is linearly separable. In practice, data is often not linearly separable; and even if it is, a greater margin can be achieved by allowing the classier to misclassify some points. To allow errors we replace the inequality constraints with $y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1,...,n$ Where $\xi_i \geq 0$ are slack variables that allow an example to be in the margin ($0 \leq \xi_i \leq 1$, also called a margin error) or to be misclassified ($\xi_i > 1$). Since an example is misclassified if the value of its slack variable is greater than 1, $\sum_i \xi_i$ is a bound on the number of misclassified examples. Our objective of maximizing the margin, i.e. minimizing $\frac{1}{2}\|w\|^2$ will be augmented with a term $C\sum_i \xi_i$ to penalize misclassification and margin errors. The optimization problem now becomes: $\min imize_{w,b}$ $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$ Subject to: $y_i(w^T x_i + b) \geq 1 - \xi_i \quad \xi_i \geq 0$ The constant C > 0 sets the relative importance of maximizing the margin and minimizing the amount of slack. This formulation is called the soft-margin SVM, and was introduced by Cortes and Vapnik (Sch olkopf and Smola, 2002). Using the method of Lagrange multipliers, we can obtain the dual formulation which is expressed in terms of variables α_i (Cristianini and Shawe-Taylor, 2000):

$$\text{maximize}_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j x_i^T x_j \quad \text{Subject to: } \sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C. \quad \text{Eq.3}$$

The dual formulation leads to an expansion of the weight vector in terms of the input examples: $w = \sum_{i=1}^{n} y_i \alpha_i x_i$ The examples x_i for which $\alpha_i > 0$ are those points that are on the margin or within the margin when a soft-margin SVM is used. These are the so-called support vectors. The expansion in terms of the support vectors is often sparse, and the level of sparsity (fraction of the data serving as support vectors) is an upper bound on the error rate of the classier (Cortes and Vapnik, 1995).

The dual formulation of the SVM optimization problem depends on the data only through dot products. The dot product can therefore be replaced with a non-linear kernel function, thereby performing large margin separation in the feature-space of the kernel. The SVM optimization problem was traditionally solved in the dual formulation, and only recently it was shown that the primal formulation, can lead to efficient kernel-based learning (DeCoste and Weston 2007).

3. ACCURACY EVALUATION METHOD

For decision making to separate target from non target pixels, a threshold is necessary. One of most reliable way to find a threshold is using Receiver Operating Characteristic (ROC) Curves. It has been used with the Neyman-Pearson method in signal detection theory. It can be used to visualize a classifier performance in order to select the proper decision threshold. The ROC Curves compare a series of similarity image classification results for different threshold values with ground truth information. Aprobability of detection (Pd) versus a probability of false alarm (Pfa) curve and a Pd versus a threshold curve are reported for each selected class (rule band).

For calculating of ROC curves, Confusion Matrix is needed. A confusion matrix is a form of contingency table showing the differences between the ground true data and classified images and it is computed by cross tabulation technique. In case of a single class classification or target detection we obtain a confusion matrix such as given on Table 1.

Table 1: A Confusion Matrix for Target Detection Case

Confusion Matrix		Classified Classes		
		0	1	sum
True Classes	0	Tn	Fp	Cn
	1	Fn	Tp	Cp
	sum	Rn	Rp	N

The elements of this matrix are defined as:

Cn=Tn+Fp; Cp=Fn+Tn; Rn=Tn+Fn; Rp=Fp+Tp; Cn+Cp=Rn+Rp=N

Tn (true negative) is the number of non target pixels which are correctly classified as non target. P (Tn) is its probability or rate as calculated using: P (Tn)=Tn/Cn.

Tp (true positive) is the number of target pixels which are correctly classified as target and P(Tp) is its rate as obtained using: P(Tp)=Tp/Cp. It is also called probability of detection: Pd.

Fp (false positive) is the number of non target pixels which are incorrectly classified as target and P(Fp) is its probability as calculated by: P(Fp)=Fp/Cn. It is also called probability of false alarm: Pfa.

Fn (false negative) is the number of target pixels which are incorrectly classified as non target and P(Fn) is its probability as calculatedby: P(Fn)=Fn/Cp.

This matrix and its elements must be calculated for a set of thresholds. In practice we fix a number of thresholds between the minimum and maximum values of rule data. Then, for each threshold, a Pd and Pfa could be calculated. With each triple of (thr, Pd, Pfa) we can plot two curves: A ROC that contains the Pd against the Pfa and another curve that contains the Pd against the threshold. An example of ROC curves are presented on Figure 1.

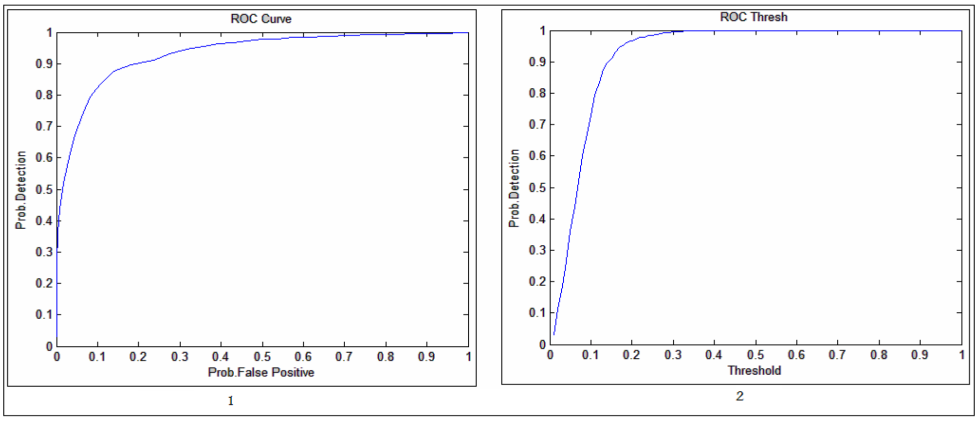

Figure1: Curves (1) probability of detection versus probability of false alarm and (2) probability of detection versus threshold

4. EXPERIMENTS

4.1 HSI Data

The above techniques are applied to CASI (Compact Airborne Spectrographic Imager) hyperspectral images. CASI has a flexible spectral resolution capability. It means that the image data may have different numbers of bands, maximum 288. These numbers of bans cover a rang from 0.4 to 1.0 µm of electromagnetic spectrum, so the wide of each bands is about 10 µm. Spatial resolution of CASI is a function of its IFOV and altitude of airborne platform. It can vary from 1 to 10 meters. Dynamic rang of sensor is another parameter which produce the image data with 12 bits or 4096 gray levels. CASI also is equipped by GPS and INS for In/Off fly rectification and geo-referencing of images. The data in this test is one set of CASI image. The spatial resolution of image is 2m and the number of bands for this image is fixed to 32 channels.

Figure 2: (a) the false color CASI image of study area (R=0.914, G=0.620, B=0.451), (b) Ground truth data for accuracy evaluation, (c) The extracted spectra of building material

4.2 Experimental Results

For applying the techniques, we have selected a test image on the same area containing man-made objects like’s roads, buildings and green spaces. It includes a 128X128 pixels image with 32 bands and a spatial resolution of 2 meters, (Figure 2-(a)). A target spectrum of building materials has been extracted by collecting and averaging the spectra of manually selected pixels for sample data (Figure 2-(c)).

The result maps for each method have been obtained (Figure 3). To compare and evaluate the results, we extracted a true data map by visual interpretation of the building materials of the scene (Figure 2-(b)). For a quantitative evaluation of the results, we retain two elements of the confusion matrix: the overall accuracy (OA), and the kappa coefficient (K). The overall accuracy is calculated by summing the number of both target and non target pixels correctly classified and dividing by the total number of pixels. Because the OA is not a very complete and reliable criterion, the Kappa coefficient is computed with other elements of the confusion matrix (Alimohammadi, 1998).

1 2 3

Figure 3: the resultant images of algorithms actions: (1) MSAS, (2) CMFM, (3) SVM

It is visible that the SVM provides the less noisy results. From Table 2, we can observe that the two quality criteria for this method are better than four other approaches. In all result maps, there are some identified pixels which are relatively similar to target but not completely. For MSAS method, the target maps have a lot of mismatching. The CMFM results are more precise for this purpose.

Table 2: computed quantities in assessing the experiments accuracy

Algorithm	Overall Accuracy	**Kappa Coefficient**
MSAS	0.91	**0.78**
MCMF	0.95	**0.83**
SVM	0.98	**0.91**

5. CONCLUSION

This work shows that it is possible to extract from hyperspectral data very important information, useful for the environmental characterization of urban areas. Even if in its preliminary stages, this research has shown many potentialities for urban remote sensing. We have applied the SVM for target detection in HSI. Experiments on urban HSI scenery illustrate that the SVM-based detector can provide higher TPRs and substantially lower FPRs than other methods in varying scenarios of target spectral variability.

Future work, we intend to investigate a more efficient selection of the kernel parameter s rather than performing a linear search over all the candidate s values. The selection of features that maximize separability is crucial in PR systems. Because of their success in a variety of PR applications, we will investigate the potential of the discrete wavelet transform coefficients as features in the context of SVM-based HSI target detection.

REFERENCES

Alimohammadi, A., 1998. Application of Remote sensing in Urban Area. Ms Courses, in RS&GIS departement of TMU Tehran.

Cristianini, N., and Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines. Cambridge UP, Cambridge, UK.

DeCoste, D., and Weston, J., 2007. Large Scale Kernel Machines. MIT Press, Cambridge.

Henz, N., and Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivar. Anal., vol. 62, no. 1, pp. 1–23.

Kraut, S., Scharf, L. L., and McWhorter, L. T., 2001. Adaptive subspace detectors," IEEE Trans. Signal Process., vol. 49, no. 1, pp. 1–16.

Manolakis, D., Marden, D., and Shaw, G., 2003. Hyperspectral image processing for automatic target detection applications. Lincoln Lab. J., vol. 14, no. 1, pp. 79–114.

Manolakis, D., and Shaw, G., 2002. Detection algorithms for hyperspectral imaging application. IEEE Signal Process. Mag., vol. 19, no. 1, pp. 29–43.

Scharf, L. L. and Friedlander, B., 1994. Matched subspace detectors, IEEE Trans. Signal Process., vol. 42, no. 8, pp. 2146–2157.

Scharf, L. L., and McWhorter, L. T., 1996. Adaptive matched subspace detectors and adaptive coherence. in Proc. 30th AsilomarConf. Signals, Syst.,Comput., Pacific Grove, CA.

Shaw, G., and Burke, H., 2003. Spectral imaging for remote sensing. Lincoln Lab. J., vol. 14, pp. 3–28.

Scholkopf, B., and Smola, A., 2002. Learning with Kernels. MIT Press, Cambridge, MA.

Tax, D. M. J., and Duin, R. P. W., 2004. Support vector data description. Mach. Learn., vol. 54, no. 1, pp. 45–66.

Vaz, C., and Thakor, N., 1989. Adaptive fourier estimation of time-varying evoked potentials. ZEEE hns. Bwmed Eng., vol. 36, pp. 448455.

Yamany, S. M., Farag A. A. and Hsu S.-Y. 1999. A fuzzy hyperspectral classifier for automatic target recognition (ATR) systems. Pattern Recognit. Lett., vol. 20, no. 11–13, pp. 1431–1438.

BIOGRAPHICAL NOTES

Davood Akbari has got his B.Sc. in Surveying Engineering from University of Imam Hossein, Tehran, Iran. He has got his M.Sc. in Surveying Engineering with the specialization of Remote Sensing from department of Surveying and Geomatics Engineering, University of Tehran, Iran. He is PhD student of Remote Sensing in University of Tehran and Assistant Professor in University of Zabol now.

CONTACTS

Mr. Davood Akbari (Assistant Professor),
Dept. of Surveying and Geomatics Engineering,Faculty of Engineering,
University of Zabol, Zabol, IRAN.
Email: davoodakbari62@gmail.com
Tel: +98-915-561-9799