A Novel Adaptive Scheme for Evaluating Spectral Similarity in High-resolution Urban Scenes

Bin Chen, Student Member, IEEE, Anthony Vodacek, Member, IEEE,
and Nathan D. Cahill, Senior Member, IEEE

Abstract

The analysis of high spatial resolution urban images for object recognition must deal with variable illumination conditions and many spectrally similar materials in the built environment. Spectral similarity measures have the potential to contribute to the effective analysis of urban scenes, however, without readily available surface reflectance conversion, the characteristics of existing spectral measures may lead to unacceptable performance. To better account for these spectral imaging scenarios for an urban environment, a simplified in-scene radiometric calibration approach is presented to preserve data collinearity and a novel spectral similarity measure based on the geometric characteristics of the Mahalanobis distance is developed to incorporate both spectral direction and spectral magnitude. With a minimum of human input to define representative pixels, the experimental results demonstrate through the analysis of ROC curves the potential advantages of the novel distance measure when applied to the identification of materials in urban images.

Index Terms

spectral similarity, high resolution, SAM, SCM, SID, Euclidean distance, Mahalanobis distance, radiometric conversion

I. INTRODUCTION

In context of remote sensing, imaging spectroscopy reveals the unique property of the imaged target, which can be exploited to identify all objects with similar spectral characteristics. Spectral imaging from remote platforms can provide periodic, large-scale, sustainable, and efficient surveillance of the earth surface and is crucial to target detection and land mapping efforts. The mapping process is often based upon the quantitative evaluation of the spectral similarity or dissimilarity of a target and a reference object in terms of

B. Chen and A. Vodacek are with the Digital Imaging and Remote Sensing Laboratory, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, 14623 USA e-mail: vodacek@cis.rit.edu.

N.D. Cahill is with the Center for Applied and Computational Mathematics, School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY, 14623 USA.
a single pre-defined criterion, e.g., spectral angle mapper (SAM) [1], spectral information divergence (SID) [2], [3], spectral correlation mapper (SCM) [4], [5], squared Euclidean distance (ED), and squared Mahalanobis distance (MD) [6]. The reference spectral data can either be collected from field measurements, obtained from a spectral library, or extracted directly from the image. These measures are important tools in remote sensing imagery analysis and generic algorithm development. There are several prior publications describing applications and comparisons of spectral similarity measures [7], [8]. These works have demonstrated that spectral similarity measure can be very useful, but is often scene dependent.

Recent advances in sensing technology have led to improved spatial and spectral capabilities for remote sensing systems and enabled new detailed applications in urban areas. Small objects, e.g., architectural details, vehicles, signage, and even pedestrians are observable in urban settings. Yet the advanced resolving capability gives rise to the problem that excessive details, i.e., many unique man-made materials and local spectral variation caused by the topographical illumination effect [1], [9] (shadows and shading due to varied surface normals), are apparent in an urban scene. The non-uniform illumination issue can be solved using brightness normalization, but this only works on reflectance data. For example, to most effectively apply SAM, SID, and possibly SCM, the data should be reduced to “apparent reflectance”, with all dark current and path radiance biases removed. [1] However, surface reflectance is not easily obtainable for high-resolution commercial satellite sensors. The data may only be available as digital number (DN) or uncalibrated spectral radiance data. The conversion of DN to reflectance requires ground truth data collected at the moment when the sensor platform flies overhead, or alternatively, some algorithms, e.g., FLAASH [10] and QUAC [11] can compensate for atmospheric effects without in situ data, but the fidelity cannot be guaranteed and the added complexity hinders their application.

Another issue in the analysis of high-resolution urban scenes is that many man-made materials have very similar spectral signatures. For example, gray asphalt roads and parking lots are often confused with bright roofing materials by spectral similarity measures that operate only on spectral direction, whereas magnitude (brightness) can be used to separate the materials. A measure such as MD is more appropriate in this case but the effort to find sufficient representative training data for each material in complex scenes becomes overwhelming.

One way to overcome the inherent limitations of single similarity measures is to combine them into hybrid measures. However, hybrid measures may inherit the weakness of each combined measure, determining how to combine measurements with different physical meanings (e.g., degree vs. distance) is a challenge, and training of optimal relative weight for each measure component may be burdensome. Thus, we will not consider hybrid measures in this paper and direct the reader to the literature [9], [12], [13], [14], [15], [16],
Without reverting to a hybrid measure, one promising way to include both spectral direction and magnitude is to use the squared Mahalanobis distance, however, as previously mentioned, the training requirements to build the necessary covariance matrix are excessive for applications in urban environments. In this paper we describe a novel spectral similarity measure that is built similar to MD but rather than collecting training data used to develop the covariance matrix, a user finds a few representative target pixels and then tunes an adjustable shape parameter for the target hyper-ellipsoid in spectral space. We also describe a simple user directed dark object subtraction method for calibration to apparent radiance to be applied to the data prior to the application of a spectral similarity measure. We then test our novel measure against a set of other single spectral similarity measures using receiver operator characteristic (ROC) curves to assess the results. The remainder of the paper is organized as follows: in Section II, the overview of the common spectral similarity measures is given, followed by presenting a new radiometric calibration method and a new spectral similarity measure in Section III; in Section IV experimental results are provided for comparison; and finally the paper concludes in Section V.

II. REVIEW OF SINGLE SPECTRAL SIMILARITY MEASURES

In this section we review the mathematical definitions and geometrical descriptions of the common single similarity measures presented in the previous section. This overview provides the context and mathematical basis that leads to the development of our new similarity measure. The pixel spectrum is mathematically represented as a vector in N-dimensional spectral space, where N is the number of spectral bands.

A. Single Spectral Similarity Measures

Since it is impossible to visualize data with more than three dimensions, the isosurfaces of the similarity measures are depicted in a 3-D spectral space in Fig. II.1. Note that the isovalue contours are not completely depicted in Figs.II.1b and II.1c due to the undetermined locus of SCM and SID. Note that the surfaces of the isovalue contours must be bounded in the first quadrant due to the positivity of realistic physical quantities. \( \mathbf{x} \) and \( \mathbf{y} \) are defined as two vectors of random variables such that \( \mathbf{x} = (x_1, x_2, \cdots, x_N)^T \), \( \mathbf{y} = (y_1, y_2, \cdots, y_N)^T \), with the same dimensionality of \( N \), where T denotes the operation of transposition.

1) Spectral Angle Mapper: \( \text{SAM} \) is an extensively used metric to evaluate the spectral similarity of a pair of two pixel spectra and is defined as

\[
\text{SAM}(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left( \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \right).
\]
Figure II.1: Graphical isovalue surfaces of five spectral similarity measures. The black dots indicate the positions of arbitrarily chosen reference pixel points. The solid lines or the surfaces encompassed by solid lines represent the sets of points most similar to the reference in terms of similarity measures, while the outer surfaces encompassed by the dotted lines denote the sets of isovalue regions. Here “most similar” means having extreme spectral similarity scores. (a) SAM. (b) SCM. (c) SID. (d) ED. (e) MD with three different isosurfaces, which is equivalent to the proposed measure.

The score given by SAM is normally measured in radians. The loca of pixel points with zero SAM value are depicted in Fig. II.1a as a solid line running from the origin. The reference point also lies exactly on this straight line. An isovalue surface is overlaid as the dotted lines/contours and forms an upside-down cone with the origin as the vertex, but expanding to infinity. These characteristics show that SAM measures only the angular difference of spectral direction rather than the magnitude or brightness, because it is invariant to the scaling of spectral magnitude.

2) Spectral Correlation Mapper: SCM is defined as Pearson’s correlation coefficient based on population statistics:

$$\text{SCM}(x, y) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are mean values of the components of two comparative vectors. SCM takes into account brightness differences and shape differences between spectra [7]. The correlation can either be positive or negative, but its maximum cannot exceed 1, which means fully (linearly) correlated. SCM is only sensitive to a linear relationship between two random variables. Pixel points with SCM values of 1 lie on a subset of a plane determined by the reference point and a space line $B1 = B2 = B3$ as highlighted by a series of parallel segments in Fig. II.1b. SCM is invariant both to the spectral direction and translation of spectral magnitude.

3) Spectral Information Divergence: SID is defined as symmetrized discrete-form Kullback-Leibler (KL) divergence by adding together two conjugate KL terms:
\[
\text{SID}(x, y) = \sum_{i=1}^{N} P(x_i) \log \frac{P(x_i)}{P(y_i)} + \sum_{i=1}^{N} P(y_i) \log \frac{P(y_i)}{P(x_i)},
\]

where the probability mass functions are defined as the normalized pixel spectra such that

\[
P(x_i) = \frac{x_i}{\sum_{j=1}^{N} x_j}, \quad P(y_i) = \frac{y_i}{\sum_{j=1}^{N} y_j}, \quad i = 1, 2, \ldots, N.
\] (II.1)

If two random variables \(x\) and \(y\) have distinct probability distributions, \(\text{SID}\) will tend to be large. \(\text{SID}\) is applied to measure the spectral similarity between two pixel spectra \(x\) and \(y\) in a stochastic manner and is essentially the divergence based on the discrete probability of each band per pixel. Owning to the normalization, the pixel points with the zero \(\text{SID}\) value form a ray starting from the origin and passing through the reference point (see Fig. II.1c), which is analogous to \(\text{SAM}\). Given the difficulty of depicting an isosurface of \(\text{SID}\), only one easily derived iso-value curve (dotted line in Fig. II.1c) is drawn instead for illustration. \(\text{SID}\) is also invariant with the scaling of spectral magnitude.

4) \textit{Squared Euclidean Distance:} \(\text{ED}\) is computed as the \(L_2\)-norm difference of magnitudes of two spectra:

\[
\text{ED}(x, y) = (x - y)^T (x - y).
\]

\(\text{ED}\) is an isotropic metric and does not favor any specific spectral direction. As shown in Fig. II.1d, its isosurface is spherical with the centroid as the reference point and the radius as the square root of its score.

5) \textit{Squared Mahalanobis Distance:} \(\text{MD}\) is a class separability measure under the equal covariance and multivariate Gaussian assumption and is defined as

\[
\text{MD}(x, y) = (x - y)^T \Sigma^{-1} (x - y),
\] (II.2)

where the covariance matrix \(\Sigma\) is typically learned from labeled data [20], [21], [22].

There is only a single point, which is the reference, with the similarity score of zero as illustrated in Fig. II.1e. The isosurface of \(\text{MD}\) is an elongated (or equivalently squashed) ellipsoid along the direction of the reference vector. The length of principal axis is larger than that of any other minor axis with equal length.

\textbf{B. Remarks}

\textit{SAM} and \(\text{SID}\) should only be used for evaluation of spectral similarity in terms of reflectance and they are insensitive to the lengths of spectral vectors. To the contrary, \(\text{ED}\) makes no distinction on spectral direction
and merely records the relative $L_2$ distance depending only on the spectral magnitude. $MD$, however, considers both spectral direction and spectral magnitude and emphasize according to covariance information. But $MD$ is a class separability measure and cannot be used on spectral comparison of two single vectors. Although the statistical assumption of $SCM$ is sound, the interpretation of its physical basis is not easy. It is difficult to explain why the pixel points most spectrally similar (fully correlated) to the reference point are located on an asymmetric subspace with respect to the reference vector (2-D plane in 3-D space). Further, its invariance with translation of magnitude is not physically reasonable. $SID$ has a similar problem that its isosurface shows no symmetry, while $SAM$, $ED$, and $MD$ are at least symmetric with respect to the line passing through both origin and reference point.

$SID$, $ED$, and $MD$ are actually instances of Bregman divergence [23] with different convex functions. Among them, $MD$ is simply the $ED$ between two points transformed by a matrix dependent on $\Sigma$ and $MD$ is also a special case of $SID$ under continuous Gaussian assumption of identical covariances and means as single pixel vectors. [24] Given that $SID$ only sets probability measures as “normalized” spectral vectors, it seems that $MD$ gives a richer representation of the underlying statistical distribution of the reference datum point.

Mathematical limitations of single spectral similarity measures lead to practical constraints: $ED$ cannot handle spatially varying illumination issue; $SAM$, $SID$, and possibly $SCM$, can fix this but also require reflectance data. They also are weak in identifying objects with similar spectral profiles; $MD$ cannot be used without a large training set. These limitations restrain the application of existing similarity measures to high-resolution urban scenes.

III. PROPOSED SCHEME

Since reflectance data from high-resolution commercial satellite sensors are not routinely available, in Section III.A, we present a simple radiometric calibration approach that can effectively preserve data collinearity to enable correction for the illumination effect. Then in Section III.B, we describe a novel spectral similarity measure with feature similar to $MD$ where both spectral direction and spectral magnitude are captured in a single measure.

A. Radiometric Conversion Method

In reflectance space, image data of the same type of material should distribute along a ray passing through the origin, and thus is the ideal situation for the application of similarity measures that leverage spectral direction. High-resolution image data, however, are normally recorded as DN or radiance and include an atmospheric component. Further, high-fidelity radiometric calibration is not readily available for these
images. To tackle the problem, a radiometric conversion approach, analogous to the well-known empirical line method (ELM), is proposed to perform atmospheric correction and facilitate spectral analysis.

If a target lies on a slanted plane or its sky dome is obstructed by adjacent objects, the downwelled radiance onto the target will be reduced. The fraction of the sky hemisphere above the target, also known as shape factor, is defined in accordance with geometry as \( \mathcal{F} = 1 - \frac{1}{2} \sin \sigma \mathcal{F} \), where \( \sigma \mathcal{F} \) is the target slope angle. Scene geometry relation is established with regard to \( \sigma \mathcal{F} \): \( \sigma' (\mathcal{F}) = \sigma'_s - \sigma \mathcal{F} \), where \( \sigma'_s \) is the solar zenith angle and \( \sigma' \) is the angle from the normal of the target to the sun.

The effective radiance at the sensor aperture, also known as top-of-atmosphere (ToA) radiance, is effectively approximated as \( L_{ToA} (\lambda) = E_s' (\lambda) \cos \sigma' (\mathcal{F}) \tau_1 (\lambda) + \mathcal{F} E_d (\lambda) \frac{r (\lambda)}{\pi} \tau_2 (\lambda) + L_u (\lambda) \), \( \text{(III.1)} \) where \( E'_s (\lambda) \) is the exoatmospheric spectral irradiance onto a surface perpendicular to the incident beam, \( \tau_1 (\lambda) \) is the atmospheric transmission along the sun-target path, \( E_d (\lambda) \) is the total downwelled spectral irradiance, \( r (\lambda) \) is the bidirectional reflectance factor, \( \tau_2 (\lambda) \) is the atmospheric transmission along the target-sensor path, \( L_u (\lambda) \) is the radiance from the sun scattered upward into the sensor’s line of site along the sensor-target path. Since the physical parameters are all dependent on the wavelength, the explicit dependency of \( \lambda \) will be removed for brevity.

ELM \cite{26}, \cite{27} is a linear regression of observed radiance values against known reflectance values (ground truth) according to

\[
L_{ToA} = [E'_s \cos \sigma' (\mathcal{F}) \tau_1 + \mathcal{F} E_d] \tau_2 \pi^{-1} \cdot r + L_u, \tag{III.2}
\]

where the leading factor before \( r \) is the slope of the regression and \( L_u \) is the intercept. The image data can then be calibrated to surface reflectance based on this linear relationship. ELM assumes that ToA radiance is proportional to the varied reflectance. For our application, constant reflectance of a single object is desired, but the true reflectance value may remain unknown. We would like to relate radiance data to the shape factor \( \mathcal{F} \), rather than the reflectance \( r \), so as to achieve consistent spectral similarity measurements with various scene geometries, which result in non-uniform illumination. \( \tau_1, \tau_2, \text{and} L_u \) are assumed to be constant over the scene as long as the sensor does not have a very wide filed of view and a very large area of coverage. In addition, we assume that the samples are approximately Lambertian so that the errors introduced by sensor viewing angle effects could be minimized. If this is not the case, solving for the reflectance value \( r \) requires estimating the shape of the bidirectional reflectance distribution function (BRDF).

If the relative amount of skylight to the total direct solar irradiance incident on the target is known, i.e.,
$E_d = \alpha E_s' T_1$, where $\alpha$ is a scaling factor for each band. (III.2) can be subsequently reduced:

$$L_{ToA} (\mathcal{F}) = [\cos \sigma' (\mathcal{F}) + \mathcal{F} \alpha] E_s' T_1 T_2 \pi^{-1} \cdot r + L_u$$

$$\equiv A \cdot \beta (\mathcal{F}) + B,$$

(III.3)

where $\beta (\mathcal{F}) = \cos \sigma' (\mathcal{F}) + \mathcal{F} \alpha$, $A = E_s' T_1 T_2 \pi^{-1} r$, and $B = L_u$. Derived from (III.3), the ground-leaving radiance of the target is then expressed as the difference between the ToA radiance and the upwelled radiance,

$$L_{gd} \equiv L_{ToA} (\mathcal{F}) - B = A \cdot \beta (\mathcal{F}).$$

(III.4)

Suppose that two targets, $T_1$ and $T_2$, consisting of the same material lie on a surface with varied surface normals, which leads to different shape factors $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. From (III.4), The ratio of ground radiance of two samples is then given:

$$\frac{L_{gd,T_1}}{L_{gd,T_2}} = \frac{L_{ToA} (\mathcal{F}_1) - B}{L_{ToA} (\mathcal{F}_2) - B} = \frac{\beta (\mathcal{F}_1)}{\beta (\mathcal{F}_2)} = \frac{\cos \sigma' (\mathcal{F}_1) + \mathcal{F}_1 \alpha}{\cos \sigma' (\mathcal{F}_2) + \mathcal{F}_2 \alpha}.$$  

(III.5)

Note that $\alpha$ in (III.5) is implicitly dependent on $\lambda$. If we relax the constraint on band-dependency and the scaling factor no longer depends on wavelength $\lambda$, i.e., $\alpha (\lambda) = \alpha$, (III.5) will yield a constant value. The ratio is constant across all spectral bands for any two arbitrarily given target pixels representing exactly the same material. Now ground-leaving radiance has properties similar to reflectance, and spectral similarity measures can be applied to image data with upwelled radiance removed, because the data collinearity is preserved.

There are several techniques for estimating the upwelled radiance. Since we assume that there are no ground truth data available, we implemented - dark object subtraction (DOS) - to correct for the atmospheric effect. DOS assumes the existence of a dark object with zero or small surface reflectance in the scene and a horizontally homogeneous atmosphere. [28] The minimum ToA radiance value in the histogram from the entire scene is thus attributed to the atmospheric effect and is subtracted from all the pixels [29]. We simply seek to find one of the darkest pixels in the scene which is defined as the one without direct solar illumination and also receiving negligible sky light illumination. As a result, deeply shadowed pixels surrounded by high-rise buildings and/or tall trees are potentially feasible candidates for an urban dark object and can be easily identified by the user. Hence, ToA radiance of the dark pixel is contributed solely by the upwelled radiance and (III.1) reduces to a much simpler form:
The cumulatively upwelled radiance is assumed to be uniform across the whole image scene. It can be seen that ToA radiance data of two targets $T_1$ and $T_2$ are not collinear with a line passing through the origin, due to the offset caused by the upwelled radiance. Only after $L_u$ is subtracted from the ToA radiance can the data collinearity be found. This type of data could also be possibly used to match with a spectral library that is mainly documented as calibrated data.

**B. Proposed Spectral Similarity Measure**

We previously discussed how MD has the advantage of capturing more spectral features, but the disadvantage of requiring the collection of a large training data set to construct a nonsingular covariance matrix. It is challenging to find adequate representative spectral training pixels in a complex urban scene, where there are too many variants within a single class. To exploit spectral features as much as possible and lessen the dependence on training, a new spectral similarity measure is introduced based upon MD but with a user adjustable isovalue hyper-surface. It incorporates spectral direction and spectral magnitude naturally, because it does not require any kind of combination. In addition, the proposed approach offers more freedom in designing a distance isosurface in spectral space, which facilitates the manipulation of the comparative tendency towards either spectral direction or spectral magnitude.

Referring to (II.2), covariance matrix $\Sigma$ is always symmetric and thus can be diagonalized as $\Sigma = \Phi \Lambda \Phi^T$, where $N \times N$ orthonormal matrix $\Phi$ has as its columns the corresponding eigenvectors of $\Sigma$, and $\Lambda$ is a $N \times N$ diagonal matrix whose diagonal elements are variances of the corresponding bands. Its first element is intentionally made $w^2$ times of the rest such that

$$\Lambda = \begin{bmatrix}
w^2 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \\
\end{bmatrix},$$

where $w \geq 1$ and $w \in \mathbb{R}$, thus it is assured that the first element $w^2$, corresponding to $x$, is reasonably larger than the other elements. It is implicitly assumed no preference for spectral direction other than the direction of the reference vector. By altering $w$, we can achieve different shapes and orientations of the isovalue hyper-curves in spectral space. The preference for the extent of spectral “purity” is literally controlled by the selection of $w$. Larger $w$ means spectral direction dominates more over magnitude, which visually makes the
hyper-ellipsoid “flatter” and the measure more sensitive to the change of spectral direction, and vice versa. The customized covariance matrix $\Sigma$ can then be easily constructed because only the subspace spanned by the axial vectors normal to the reference matters, see

$$\Sigma = I + \left( w^2 - 1 \right) xx^T / \|x\|^2,$$  \hspace{1cm} (III.6)

where $I$ is a $N \times N$ identity matrix. The construction only depends on the reference vector $x$ and variable multiple $w$. Given $\Lambda$ is of full rank, $\Sigma$ is always invertible.

According to (III.6), the proposed measure is named anisotropy-tunable distance (ATD), because it uses a user-defined covariance matrix $\Sigma$, rather than one generated from training data in a traditional way. Since only linear operations are involved, the generation of a covariance matrix is not computationally expensive. The hyper-ellipsoidal isosurface of ATD is elongated along the principal axis, i.e., the direction of $x$, which indicates its sensitivity to the change of spectral direction. The shape of the isosurface can be tuned by adjusting the first eigenvalue of $\Sigma$, upon which the extent of sensitivity to varying spectral angles is strengthened or weakened. This configuration results in a series of hyper-ellipsoidal isosurfaces with respect to varied $w$ (see Fig. II.1e).

It is also straightforward to see that ATD is equivalent to ED when $w = 1$. Unlike the existing measures, choosing $x$ or $y$ as the reference will result in different measure scores for ATD. In other words, ATD is not a metric because it is not symmetric. But this is not a concern, since for most applications we compare pixel vectors with only one reference vector, which can be derived from the image manually or automatically. For example, one application of spectral similarity measures is active contour based image segmentation, which uses mean intensity within a region as the spectral reference to be compared with pixel intensities within this region. The extraction of reference is image-based and fully automatic. In the following, we will use $\text{ATD}(\omega)$ to denote the similarity score of ATD when parameter is $\omega$.

C. Summary

The proposed scheme involves a novel distance measure for evaluating spectral similarity in high-resolution urban scenes and an integrated radiometric calibration approach. The new measure is generated under the Gaussian assumption of data distribution and uses a single reference vector as the mean. The shape of isosurface is adjusted by changing the first eigenvalue of the covariance matrix. By doing this we can tune its inclination towards spectral direction or magnitude and make it adaptive to complex urban scenes. In addition, the radiance calibration is greatly simplified because no reflectance data are needed. Though no reflectance data are used, the collinearity of radiance data has been effectively preserved, which benefits any
IV. Results & Discussion

Images used in the experiments are from two commercial high-resolution multispectral satellite sensors, GeoEye-1 and WorldView-2. They are pansharpened to 0.5-meter resolution, covering the spectral range from visible to near IR bands. In this section, we show the results of different spectral similarity measures applied on the imagery from both sensors. Based on the linear relationship, original DN data has been converted to ToA radiance for both sensors according to [30].

A. Data Collinearity Verification

The proposed scheme is first applied to a pan-sharpened WorldView-2 image. In order to test the effectiveness of DOS in preserving data collinearity in spectral space, an image portion (484×484) containing a typical hip roof with four sloped orientations is cropped from the original and is shown as natural color in Fig. IV.1a. Due to the slanted angles, sloped faces of the red roof receive different solar illuminations, resulting in distinct shading effects. Here it is assumed that this whole rooftop consists of a single material, ignoring small structures (e.g., two dark unresolvable objects on east face). Thus it is expected that data points of the rooftop should be collinear with respect to the origin in a multidimensional space. Four ground truth regions of interest (ROIs) are selected such that each region represents a unique illuminating condition, which are illustrated in Fig. IV.1b. The image data in RGB color space within these ROIs before and after similarity measure which leverages spectral direction.

This approach may fail if the upwelled radiance data are spatially varying, the scene objects are far from Lambertian, or the scaling factor \( \alpha \) changes dramatically from band to band. If the Gaussian assumption of data distribution about the reference no longer holds (though this is rare in real world), ATD is inappropriate for similarity assessment and other existing measures should be adopted instead.

Figure IV.1: (a) WorldView-2 pan-sharpened natural color image (12/10/2009) of downtown Rome, Italy showing a rooftop with four facial orientations, leading to different shading appearances. Colored pixels represent corresponding ground truth ROIs for data collinearity test. (image courtesy of DigitalGlobe) (b) shows labeled ground truth of a rooftop and each color represents one side of the roof.
the DOS processing are depicted in Fig. IV.2. Note that the original radiance is always larger than the adjusted radiance due to the positive contribution of the upwelled radiance. Mathematically, two separate straight lines passing through the origin are then fitted to two image data sets in a least-squared sense. As plotted in the same figure, the solid line denotes the fitted line for original data while the dashed line is fitted to post-DOS data. Brighter lines on the green-blue plane are the projections of the fitted lines. It is obvious to see that the adjusted radiance data preserve much better collinearity than the original data, with $R^2$ values - 0.69 for post-DOS data and 0.20 for original data. The proposed radiometric conversion is justified because the collinearity of the radiance data with respect to the origin of a homogeneous material.

Figure IV.2: Collinearity test of the image data before and after DOS. The colors of the data points correspond to ROIs in Fig. IV.1b. The marks '+' and '.' denote the original and adjusted radiance data respectively. The data points in RGB space are also projected onto the green-blue plane. $R^2$ is 0.69 for the post-DOS data and 0.20 for the original data. The dark pixel spectrum extracted from WorldView-2 image scene is shown as the inset.
Figure IV.3: Detection of roof and road/parking lot. (a) GeoEye-1 pan-sharpened natural color image (06/04/2010) of a typical urban scene in southeast Rome, Italy, with a color composition of bands 3, 2 and 1. Yellow cross and green diamond indicate positions of selected reference pixel points for roofs and roads, respectively. (image courtesy of GeoEye) (b) labeled ground truth: red denotes rooftop and blue denotes roads/parking lots.

Figure IV.4: For the GeoEye image, color-coded spectral similarity scores given by five spectral similarity measures with respect to the red pitched rooftops. Negative values of SCM are truncated to zeros and then subtracted from 1. (a) $\text{SAM}$. (b) $\text{SID}$. (c) $1 - \max(0, \text{SCM})$. (d) $\text{ED}$. (e) $\text{ATD}(3)$. (f) $\text{ATD}(5)$. (g) $\text{ATD}(7)$. (h) $\text{ATD}(9)$.

is greatly improved after DOS and the radiance data have been calibrated.
B. Spectral Similarity Measures Comparison

Referring to Fig. IV.3a, a 484×484 cropped portion of a GeoEye-1 pan-sharpened ToA radiance image is used to demonstrate the applicability of \textit{ATD} and evaluate its performance compared with other spectral similarity measures. There are two reasons why this particular scene is chosen: this is a typical urban scene covering various types of objects such as buildings, paved roads, parking lots, trees, lawns, vehicles, etc.; on the other hand, within-class spectral variations for rooftops and roads/parking lots are visible. Two important applications of spectral similarity measures in urban scenes, detection of buildings and detection of roads/parking lots, will be examined for this scene.

About a dozen of building complexes with reddish rooftops are the objectives that we would like to separate from irrelevant objects and background clutter for purpose of, say, image segmentation. The reference vector is acquired by averaging over a few representative pixels on a typical red rooftop (see Fig. IV.3a). As shown in Fig. IV.4, warm color and cool color represent high or low spectral similarity scores respectively. (n.b., the scores of each spectral similarity are scaled if possible to help a better visualization.) All reddish roofs are expected to be rendered as dark blue, because low spectral similarity score indicates high spectral resemblance. In this case, this type of roofing is assumed to be uniform. All similarity measures yield somewhat anticipated outputs. Referring to Figs. IV.4a to IV.4c, \textit{SAM}, \textit{SID}, and \textit{SCM} all produce strong spectral separation between targets and background with high local contrast, while \textit{ATD} measures have weaker contrast between foreground and background. \textit{ED} misses partial rooftops in some reddish-brown buildings due to the shading issue. As a whole, those measures that exclusively extract spectral direction perform best.

The GeoEye-1 image is used again to find spectrally similar pixels with respect to paved roads and connected parking lots. Referring to the green diamonds labeled in Fig. IV.3a, the reference vector is acquired by averaging over four representative pixels on paved roads. Here all road pixels are expected to appear as dark blue with other pixels having warmer colors. Referring to Figs. IV.5f and IV.5g, \textit{ATD}(5) and \textit{ATD}(7) both give reasonably good separation such that major roads, as well as residential roads, are assigned low scores while vegetation appears red or yellow, and other types of roofs are blue, but brighter. As \(w\) becomes larger and larger, see Figs. IV.5d \((w = 1)\) to IV.5h \((w = 9)\), \textit{ATD} gives a weaker and weaker contrast between roads and rooftops (e.g., flat gray and white rooftops), but a better separation from vegetation (e.g., trees and lawns). Apparently paved roads and gray or white rooftops have similar spectral profiles, meaning after brightness normalization their spectra bear a strong resemblance. Consequently, the grayish roofing close to the right side of Fig. IV.3a is almost indiscernible from roads for \textit{ATD}(9), see Fig. IV.5h, as well as \textit{SAM} (Fig. IV.5a) and \textit{SID} (Fig. IV.5b). \textit{SCM} yields an unsatisfactory result since severe spectral confusions arise.
Figure IV.5: For the GeoEye image, color-coded spectral similarity scores given by five spectral similarity measures with respect to the roads. Negative values of SCM are truncated to zeros and then subtracted from 1. (a) SAM. (b) SID. (c) $1 - \max(0, \text{SCM})$. (d) ED. (e) ATD(3). (f) ATD(5). (g) ATD(7). (h) ATD(9).

Figure IV.6: ROC curves for roof detection.
Next, given each spectral similarity measure, we threshold its score and generate a ROC curve against the ground truth data of either roof or road/parking lot (Fig. IV.3b). ROC curves of the detection of both urban objects for all spectral similarity measures are plotted in Figs. IV.6 and IV.7, respectively, and the corresponding area under curve (AUC) values are illustrated in Fig. IV.8. \textit{SAM}, \textit{SID}, and \textit{SCM} all produce very high AUCs in roof detection, while their performances deteriorate in segmenting roads and parking lots, which verifies previous visual inspection. This is because the pixel intensity differences of red rooftops are caused by non-uniform illumination and can be compensated by brightness normalization. However, the spectral heterogeneity of roads/parking lots is not caused by scene geometries, but rather distinct materials.
Table I: Computritional time in millisecond (averaged over 10 runs).

<table>
<thead>
<tr>
<th></th>
<th>SAM</th>
<th>SID</th>
<th>SCM</th>
<th>ED</th>
<th>ATD(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. IV.4</td>
<td>29</td>
<td>93</td>
<td>44</td>
<td>8</td>
<td>42</td>
</tr>
<tr>
<td>Fig. IV.5</td>
<td>30</td>
<td>90</td>
<td>44</td>
<td>8</td>
<td>46</td>
</tr>
</tbody>
</table>

Missing spectral magnitude knowledge weakens the separability of these measures. ATD(5) produces slightly lower AUC value in roof detection but higher AUC value in road detection because both spectral direction and magnitude are taken into account in a balanced manner. ATD has peak AUC values at \( w = 3 \) for roof detection and \( w = 7 \) for road detection, which can also be verified by ROC curves (see green curve in Fig. IV.6 and blue dashed curve in Fig. IV.7). There is always a balance to strike between competing goals like separating spectrally similar objects, either in terms of spectral direction or spectral magnitude. In this scene, \( w = 5 \) is optimal for detecting roofs and roads/parking lots simultaneously.

C. Discussion

In conclusion, SAM and SID have similar performance in measuring similarity due to the common magnitude normalization process. SCM was not very powerful in consistently detecting urban structures. ED is one of the most commonly adopted spectral similarity measures. It is, however, unable to utilize the spectral direction information. ATD has a good balance of performance with sufficient flexibility and can be adapted to complex scenes at a low computational cost of a few trial-and-error tests for selecting the optimal parameter \( w \). ATD yields satisfactory results such that non-uniform illumination of the same object is compensated and truly spectrally different objects are separated. Based on the ROC curve analysis, the optimal \( w \) value generally falls between 2 and 7.

Although the selection of \( w \) is not strictly scene-dependent, we envision that a user would do a few trials to determine the optimal value of \( w \) for a given scene and given application. However, an optimal solution of \( w \) may be difficult to find if the class data do not distribute along the direction of reference.

With regard to the computational time, all spectral similarity measures run fairly fast on the test image, (see Table. I). The tests were conducted using Matlab on a PC with a 2.67GHz CPU. Since the spectral similarity measurements work on a per-pixel basis, parallel computing can be easily deployed without extra cost, simply by partitioning the image.

V. Conclusions

A modified dark object subtraction method using deeply shadowed areas common to urban scenes is employed to simplify radiometric conversion and reinforce the collinearity of image data with respect to the origin in multidimensional spectral space and thus account for the illumination effect. Commonly used
single spectral similarity measures may not perform well at differentiating objects with similar spectral profiles, because they work on either spectral direction or magnitude, but not both. Though hybrid measures combining multiple measures are possible, they also inherit drawbacks from combined single measures. Our spectral similarity measure, ATD, is designed to behave similar to MD and to naturally and adaptively exploit relatively complete information. An examination of ROC curve results indicates that ATD is capable of evaluating spectral similarity in high-resolution urban scenes. The promising results demonstrate that ATD can provide a consistently reliable measurement of the spectral similarity in high-resolution scenes. Future work involves assessment of its application to more complicated scenes and methods for integrating ATD with image segmentation techniques and other practical applications.

VI. Acknowledgment

The authors would like to thank the support from NSF grant IIP-0917839 and AFOSR grant FA9550-11-1-0348. They also want to thank GeoEye foundation and DigitalGlobe for generously providing satellite images for this work.

REFERENCES


