Construction of synthetic spectral reflectance of remotely sensed imagery for planning purposes

Tal Feingersh*, Eyal Ben-Dor, Juval Portugali

Department of Geography and the Human Environment, University of Tel-Aviv, 39040, Israel

Received 19 July 2004; received in revised form 15 October 2005; accepted 17 November 2005

Abstract

Urban and environment development plans commonly lack spectrally based value-added information layers such as expected albedo, emissivity and temperature of the planned landscapes. These can be integrated into plans in order to assist in using specific materials or in the way new landscapes and urban spaces are designed. In contrast to existing space-borne remotely sensed imagery from which information layers as such can be extracted using atmospheric correction tools, development plans are set on paper, in a geographic information system (GIS) or as perspective “artistic images” at best. This paper describes a new software tool within the environment for visualization of images (ENVI 4.1) software, for automatic simulation of such multispectral reflectance images, given thematic maps of planned landscapes and associated spectral signatures.

We discuss issues related to the image generation process, the method of spectral signature integration, and to quality assessment measures. An example is provided. We assess the simulated output quantitatively using a pixel-based “goodness-of-fit” measure and by calculating Pearson’s correlation coefficients. Results show that simulation of images based on local neighborhood spectral mixtures, have all, mean total-goodness-of-fit measures amounting 99%, and have a general positive linear correlation of around 0.86 with real data. A class-wise correlation of a simulated image with a real reference image shows that large image segments of homogenous land-cover classes such as vegetation classes, inland waters and some soils, match about 80–90% of corresponding real data. On the other hand, simulated data will match only 20–40% of real values for highly textured land-cover classes with relatively small spatial extent over the image, such as for built-up areas. We conclude with two prospective applications related to the validation of classification algorithms, and to planning exercises.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Image simulation; Spectral mixing; Value adding; Planning

1. Introduction

Remotely sensed imagery taken over time, for a given place, creates what is generally called a multi-temporal data set (Lillesand and Kiefer, 1994). In such images we see mainly two phenomena: point-wise dynamics and regional-dynamics. The former means a land-cover entity at a given location may change spectrally over time (i.e. be dynamic), in a linear fashion (such as time in a calendar) or in a cyclic fashion (such as vegetation cycles through seasonal change) (Kainz and Raza, 1999). The latter means that static spectral objects can change their spatial distribution over time. Examples are agricultural land turning to be build-up area, floods that cover dry lands, etc. With respect to both observations there has been much work recently (Li and Yeh, 1998; Viovy, 2000; Ji et al., 2001; Rashed et al., 2002; Lavelle et al., 2002). The mere fact that we collect ground truth for classification of imagery stems from temporal change (of land, of atmosphere, of illumination conditions, etc.). A recent focus on time in the context of remote sensing (i.e. change-detection), relies on methods to quantify change (e.g. Rashed et al., 2002) and model it.
1.1. Recent experience

The first good example of the above focus is the recent UrbEx experiment (Urban Expansion monitoring) by the European Space Agency (ESA) and the World Wildlife Fund (WWF), where multi-temporal earth observation imagery were classified and subtracted to define change (Fernandez-Prieto et al., 2002). UrbEx outlines loss of natural landscape due to urban development in Italy in the last 15 years. To be more precise, the ESA looks for illegal construction, roads, expansion of ports, degradation of wetlands, destruction of dunes, coastal erosion and water pollution by industrial sewage. These indicators are highlighted and validated by ground survey of the WWF (UrbEx, 2003). Fernandez-Prieto et al. (2002) use optical SPOT panchromatic data to classify density levels within the city, and couple these with ERS SAR interferometry-based coherence maps. The latter maps generally urban versus non-urban land-use, and inland water bodies.

At an international level, efforts are also made, to understand urban dynamics. The directorate general Joint Research Center (JRC) of the European commission, launched in 1998 the MURBANDY (Monitoring Urban Dynamics) project which expanded later to become MOLAND (Monitoring Land-use/Land-cover Dynamics) based on earth observation data, socio-economic data and data sets describing transportation networks for 40 cities around Europe for the last 50 years (Kemper et al., 2002). Once data are collected it is used once to compute so-called “sustainability indicators” of urban and regional development (Lavelle et al., 2002). At the last stage the same data are input to a Cellular Automaton (CA) spatial dynamics engine, to model past and future urban trends. A detailed description of the data, and the technical basis for MOLAND’s urban growth model can be found at the project’s site (MOLAND, 2003). The link to remote sensing in MOLAND is expressed by using earth observation (EO) data as some of the input and by digitizing and interpretation of imagery.

1.2. The spectral dimension in spatio-temporal simulation

Image generators allow currently the integration of visual appearance with shading and topography in a virtual reality fashion. Such visualizations lack an important aspect that is necessary for spatio-temporal environmental research. That can be generally defined as point-related information. Dynamic changes at the pixel and sub-pixel levels, the ability to identify materials in images and assess their physical conditions quantitatively — all these do not exist in image simulators. Dynamic behavior of materials with relation to aging, climatic changes or seasonal changes and the study of their physical conditions and chemical composition require spectral information to the image at any location (Roberts and Herold, 2004). A multispectral or hyperspectral image, coupled with a cause-and-effect simulator of a cellular automaton (CA) type, will add that extra dimension of materials, their exact identification and their physical condition under pre-defined conditions. Therefore, it is expected that a true environmental study of the effects of changes on their vicinity, will be possible only by integrating spectral information with a dynamic map.

Methods and projects for reconstruction of historic data on urban morphology, growth and (possibly) future development are being developed still today. The link between the social and the empirical-regional urban-dynamics approaches can be made via CA models, that reflect the “self-organization” nature of cities today (Portugali, 1999). A good example to that is Vilaró’s CA-based land-use model (Vilaró, 1999) that integrates socio-economic transition rules to increase reliability of urban evolution.

Although spectral data of man-made materials are being collected at several cities for modeling urban dynamics (Ben-Dor et al., 2001; Ben-Dor, 2001; Segl and Roessner, 1999; Roberts and Herold, 2004), very little is done in order to actually integrate time-dependant spectral signatures and remote sensing imagery into CA models, in a way that iterations of CA are done spectrally, rather than thematically, or even in combination. Such integration can improve our quantitative understanding of physical parameters and site-related environmental changes of the city.

We would like to stress here the added value of the spectral dimension to simulated maps. The first and foremost application that will benefit from such data is automated image interpretation by classification algorithms. Such data can generate full-scale ground truth images, from independent, non-spectral data sources, such as scanned maps. Therefore it can provide a true reference for validation — per classified pixel. Such data today are very difficult to achieve, and are consuming time and resources. A second application that can benefit from such a process is the simulation of future environmental scenarios related to development alternatives. Quantitative analysis of long-term effects of pollution, water management schemes and intervention in natural habitats can be made. One can also use this process to visualize naturally occurring phenomena with very slow dynamics — in an image-like spatially dynamic fashion.

We start, therefore, from the point where recent studies have reached (i.e. past-to-present nominal representation of dynamics), and offer a method for integration of a spectral dimension to nominal visualizations for applications we suggest later. We first generate a synthetic image by reconstructing an existing image, in order to validate and discuss the synthetic result. Then we use the same method to simulate a forecasted scenario of the current state.

2. Data processing

Multispectral image generation requires at least two types of information, namely a classification map and a spectral library. A classification map will tell ‘what is where’, and a spectral library will tell ‘how things look like’. The former is a thematic description of a real input image, in which a land-cover class is defined for each of the image pixels, based on a decision rule. This rule relates to spectral samples
of each of the land-cover classes (hence “classification”). The second input is a collection of so-called spectral signatures, which represent relative reflectance intensity of all classes available in the classification map. The processing required to combine these two inputs into a synthetic image is described in the following sections.

2.1. Data description

The civilian Landsat program was started by NASA in the 1972 and resulted in the launch of several generations of imaging sensors at the visible and near-infrared parts of the electromagnetic spectrum. The latest of these sensors was launched on 1999 onboard the Landsat-7 satellite, and was called “Enhanced Thematic Mapper+” (ETM+), indicating better spatial and spectral resolution configuration with comparison to earlier sensor generations. ETM+ images are acquired on a 16-day basis repeatedly, having eight image bands corresponding to different wavelengths. Bands 1–5 and 7 cover the visible and near-infrared spectral range (0.45–2.35 μm). Band 6 is a thermal band (10.4–12.5 μm) and band 8 is panchromatic (0.52–0.9 μm). Both bands 6 and 8 are not used in this study since their spectral and spatial resolutions differ from those of the rest.

The image we use in this study is a subset of the city of Tel-Aviv, Israel, acquired on 21 May 2000, with a spatial resolution of 28.5 × 28.5 m. Fig. 1 shows a general location of the case study area. Radiometric values represent surface reflectance ranging from 0 to 1. The selected subset covers part of a recreation park along the Yarkon stream, the Tel-Aviv University campus and adjacent residential and industrial areas. The park includes areas of bare soil, grass, trees and a small lake.

2.2. Maximum-likelihood classification

The image was classified using the maximum-likelihood rule, as described in Richards (1994). Samples for classification were collected by image interpretation, resulting in 10 land-cover classes; (1) inland lakes & rivers, (2) sand, (3) bright sandy alfisol (“soil 1”), (4) regular alfisol (“soil 2”), (5) trees, (6) grass, (7) old (bright) asphalt, (8) new (dark) asphalt, (9) low leafy vegetation (“vegetation type 3”), and (10) built-up area. Image spectra that did not match any of the classes were classified as “un-known”.

The classification we use was tested using ground truth data, collected independently by observation and analysis of the image. By independent we mean that ground truth did not overlap, nor was adjacent to any of the training samples used for that classification. Calculated using a standard error matrix, total overall accuracy of the classification equaled 95%, indicating a good match with reference ground truth data. The KHAT statistic (a KAPPA estimate) equaled in our case 0.93 indicating high classification accuracy, in which the chance agreement defined by omission and commission errors (Congalton, 1991) is very minor. This data set forms our first input for the simulator, being a spatial-nominal component.

2.3. Spectral signatures

The second component for image generation is, in our case, the reflectance intensity, per image band, for each of the land-cover classes we want to generate. Spectral reflectance properties of all land-cover classes (ground truth) can be collected in the field using a hand-held spectrometer, from an existing spectral library, or from a real image (by interpretation). We chose the latter. For each land-cover class we collected several dozens of samples from the real image, and computed the mean reflectance intensity per ETM+ band as a representing spectral signature for that class. The resulting spectral library is illustrated in Fig. 2.

2.3.1. Linear spectral mixing

Spectral mixture modeling is the process of deriving mixed signals from pure end-member spectra while spectral unmixing aims at doing the reverse, deriving the fractions of the pure end-members from the mixed pixel (Stein et al., 1999). Since cities consist of a large variety of materials and of

![Fig. 1. (a) General location (geographic projection) and (b) the Tel-Aviv test area.](image-url)
relatively small sized objects, with relation to spatial resolution of most sensors, there exist in cities a relatively high percentage of mixed pixels (Segl and Roessner, 1999).

From a scene generation perspective, non-linear mixing models are less likely to apply to the urban environment, since they simulate such physical relationships as of a material covering another material, having some transparency, or spectral reflectance from mixtures of liquids, etc. (Ali, 2002; Clark, 1999). In contrast, linear models assume that materials in the field of view are optically separable so there is no multiple scattering between components. The combined signal is simply the sum of the fractional area times the spectrum of each corresponding component, and therefore such mixing is called area mixture. The two assumptions underlying this model are that end-members are Lambertian reflectors and that they are homogeneous in composition. Since in reality land-cover is a continuous phenomena, and digital images are their discrete representation, some mixing of spectral responses occur, within each image pixel. This idea is expressed as

\[ \rho = F_i S_i + \epsilon \]  

where \( \rho \) is the observed reflectance at a given pixel, \( F \) is the fraction of the area that land-cover class \( i \) covers in that pixel, and \( S \) is \( i \)'s respective spectral signature. A composition of mixed reflectance will generally include some random error (\( \epsilon \)). This residual error will be the difference between the measured and the modeled DN in each band (Van der Meer and de Jong, 2001). Residuals of all bands in a pixel can be averaged to give a root mean square error calculated from the difference of the modeled overall spectrum and the measured overall spectrum over a limited wavelength range. \( \epsilon \) can be the result of accumulating errors that relate to sensor and atmospheric parameters. A discussion on those is outside the scope of this work.

Previous work on remote sensing of urban environments observe that a representing “urban object” will be as big as 2–5 m in size (Roberts and Herold, 2004). This means that in the case of an ETM+ image, a pixel of about 30 × 30 m will result in mixing of 5–15 objects in general. The problem that arises from this observation is that classification results of a maximum-likelihood classifier are concrete decisions of only one land-cover class per pixel. If we represent each class in the nominal map by its corresponding spectrum, we get in turn a spectrum of one class only at each pixel. The result of such a process is what we call a “Naïve image” where each pixel has one “pure” spectral response, and where no mixing occurs. In search for ways to estimate fractional memberships \( (F_i) \) as weighting factors of spectra \( S_i \), in order to achieve a more realistic output, we adjust to concrete-decision classification methods such as the one we used. We estimate weight by looking at neighbors’ class attribute at the classification map and account for their topological relationship with the pixel in question (Hatna, 2004). Given a rectangular sliding window centered over a pixel of the nominal classification map, neighbors of class \( i \) influence the \( i \)th weight of that central pixel by a pre-defined convolution operator (kernel) that serves as mediator for their respective influence. The result is a convolved weight factor, per pixel, per class, that serves as component \( F_i \) for linear mixing (Eq. (1)).

In cases of medium resolution imagery as input (as in the case of Landsat ETM+ images), the size of the sliding window should not exceed 3 × 3 neighborhoods, as common urban objects of size 2–5 m are not distinguishable any more (Roberts and Herold, 2004) and therefore spectral relationships between pixels do not propagate far. In cases of high-resolution input, these sliding windows can become larger. The result of such operation is an image the size of the input (real) image, with \( i \) values of weights per pixel (\( i \) being the number of classes to mix). Hence the weight factor \( F_i \)

\[ \begin{array}{ccc}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 \\
\end{array} \]  

\[ \begin{array}{ccc}
1/2 & 1 & 1/2 \\
1 & 4 & 1 \\
1/2 & 1 & 1/2 \\
\end{array} \]  

\[ \begin{array}{ccc}
1 & 2 & 1 \\
2 & 4 & 2 \\
1 & 2 & 1 \\
\end{array} \]  

\[ \begin{array}{ccc}
0 & 1 & 0 \\
1 & 4 & 1 \\
0 & 1 & 0 \\
\end{array} \]  

Fig. 3. Four kernels used for convolution in the synthesis process.
in Eq. (1) is solvable for all pixels and land-cover classes by using the following convolution operator:

\[
F_i = \frac{1}{\sum(k_{x,y})} \sum_{x=1}^{n} \sum_{y=1}^{m} k_{x,y} C(i)_{x,y}
\]

where \( F \) is the output weight of the \( i \)th class, defined for the pixel centered at the sliding window, \( k \) is the kernel coefficient located at column \( x \) and row \( y \) of the \( n \) by \( m \) kernel, and \( C \) is the binary representation of the presence of class \( i \) at the same location.

The key to achieve reasonable spectrally mixed output in terms of visual consistency with a realistic view, is in the definition of kernel values used in the convolution stage. Four test kernels were defined for the purpose of defining \( F_i \), resulting in four synthetic images, respectively. These kernels are illustrated in Fig. 3. The central pixel of these kernels is the one being considered for mixing. Kernel (a) allows equal weights for all eight neighbors and results consequently in smoothing of spatial data. Kernel (b) considers the distance between the centers of cells, with respect to the center of the kernel, and weights the eight neighboring spectra inversely. This results in four-connected neighbors (i.e. connected to the center by line) having about 40% more influence on the mixed output. The central cell has the strongest influence on the output and weights the corresponding spectrum four times. Kernel (c) weights four-connected neighbors twice as much as neighbors connected to the central cell by point only, and finally kernel (d) considers only the central cell and four-connected neighbors, where cells connected by point are disregarded. In the rest of the paper we refer to these kernels as “four convolution operators” or “four options”, respectively.

3. Validation

It is tempting to assume that any deviation of the synthesized image from the real image is in fact a representation of the error component \( \epsilon \) in Eq. (1). However, this is usually not the case, baring in mind that common non-fuzzy type of classification algorithms (i.e. concrete-decision classifiers) by themselves are mere simplifications of a more complex reality. Being as such, additional error is introduced to the existing \( \epsilon \). In addition to that, any misclassification will also be part of an overall error of a simulation output. In a sense, outputs resulting from linear mixing as defined in Eq. (1) are lacking the \( \epsilon \) as it appears. This means that images generated with correct mixtures should be, in fact, perfect copies of real data, having no atmospheric or sensor abstractions. The emphasis is on correct mixtures. In our case, since convolution operators estimate the mixture per pixel, an error measure \( \epsilon' \) will be the replacement

Fig. 4. Processing steps: (a) ETM+ image, (b) classification by maximum likelihood, (c) simulated image (convolution operator #2), and (d) a corresponding GoF image (band #1). Note the lowest GoF values along the Yarkon stream (see also Fig. 1). Black pixels in (b–d) are those classified as “un-known” and were not considered in the convolution process. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
of ε, and the result of the two main stages in the generation process: classification and convolution. This error can be quantified to give a measure of success and to help in fine-tuning the generation process. This allows in turn verifying that constructed imagery is indeed a true representation of future realistic environments. We use two quality measures of agreement between the generated image and its corresponding real reference image, namely a goodness-of-fit measure and correlation analysis.

3.1. The goodness-of-fit measure (GoF)

The measure is made for situations where simplification of spatial reflectance is an inherent part of linear mixing, introduced by concrete-decision classifiers. Adapting this measure from the application of segmentation of textured images, as defined in Delves et al. (1992) and in Feingersh (2000), this measure relates once to each pixel by itself, and then to the image as a whole. Let ρᵣ be the real reflectance value at a given pixel and ρₛ the synthesized reflectance value at the corresponding image location. The goodness-of-fit test, GoF, is made for each pixel, to give

\[
\text{GoF} = 1 - \frac{\sqrt{(\rhoᵣ - \rhoₛ)^2}}{\rhoᵣ + \rhoₛ} = 1 - \text{ND}
\]  

where the normalized root-squared-difference can be simplified to a factor we call ND for later purposes. GoF ranges from 0 to 1, with 1 defined as identity. An imaginary synthesized ρₛ that is an inverted image of ρᵣ, or alternatively a constant zero image, will result both in zero fit, as the right-hand part of expression (3) reaches unity. GoF images are quite intuitive in the sense that they highlight good match brightly and poor match in dark tones. In its mathematical sense, GoF images highlight the location of erroneously classified pixels, and consequently a poor spectral reconstruction. GoF images should ideally be free of any structure appearance, and show purely texture in the image. Any structure in GoF images indicates a "mis-simulation" of spectral signal, indicating under/over-estimation of the true mixed signal.

Since GoF images have a fit image for each input spectral band, one can get the difference of global fit between bands and locate local failures to reconstruct the correct spectral mixture. Suppose there are n columns by m rows in a GoF image, then a numeric measure of the total-goodness-of-fit (TGF)
per image band, can be reached by normalizing the sum of normalized root-squared-difference reflectance values by the number of pixels in the image, and subtracting it all from unity, such that

\[
TGF = 1 - \left( \frac{\sum_{x=1}^{n} \sum_{y=1}^{m} ND}{nm} \right)
\]

In order to relate fit values to classes, GoF image layers were masked by their corresponding class’ spatial distribution in the classification map, and first moment statistics were defined. The closer class means approach unity the better the overall match. However, the bigger the standard deviation of a given class, the worse is the overall consistency of the simulated class, suggesting that a better classification could take place or that texture is significantly under-estimated.

### 3.2. Spectral correlation

A complementing measure of quality of the synthesized output image is a correlation (r) made between the synthetic output and its corresponding reference. A scatter-gram representing DN values of a given reconstructed band are plotted against its counterparts’ reference values. At best, results form an ideal strong positive correlation (r = 1) and a corresponding unit determination coefficient ($R^2$). Any deviation from the origin will stem from a bias in synthesized reflectance values. Additionally, higher dispersion of values will indicate usually the lack of texture in (synthesized) spectrally homogeneous regions, or alternatively errors in the classification stage, or both. On a lower hierarchical level, it can be interesting to evaluate the performance of each synthesized land-cover class separately. This can be done by isolating those pixels that correspond to class $i$ in input and output data, and analyzing them separately.

### 4. Results and discussion of the Tel-Aviv case study

The six reflective bands of a Landsat-7 ETM+ image of Tel-Aviv were cut around the area of interest illustrated in Fig. 1b. Classification and convolution followed, using the four convolution operators defined above. Fig. 4 compares the reference image, its corresponding classification and the major outputs of this simulator, as demonstrated for convolution operator #2. Finally GoF and correlation were calculated for the assessment of performance.

Both assessment measures were analyzed in view of several aspects: overall performance per image band, per land-cover class, and finally per convolution operator.

TGF values for simulated outputs as whole (all bands together) are basically equal to unity in all cases. Variations start at the fifth digit after the decimal point (e.g. 0.99993) and are considered negligible. In order to find finer differences, we therefore assess performance by calculating the GoF of classes and bands separately. The summary of performance by the GoF measure, per image band and per class is illustrated in Fig. 5 and Fig. 6, respectively.

Considering GoF values in Fig. 5 we observe that all bands show relatively high performance, scoring as high as 0.91 in average (unity being identity). The general fit of the green, red and near-infrared bands (2, 3 and 4, respectively) perform slightly better than bands 1, 5 and 7, with mean GoF values 4% bigger in general (Fig. 5). Treating standard deviations as consistency indicators (e.g. Fig. 5b) their corresponding standard deviations are relatively low (standing at about 5% in average), suggesting that their match is relatively consistent. Bands 1, 5 and 7 have bigger standard deviations, exceeding 7% in average, and indicating generally lower consistency. This band-wise trend appears to be linear ($R^2 = 0.971$) and is repeated, although less clearly, when referring to class performance. This can be explained possibly by a better signal to noise ratio (SNR) in these bands (in the reference image), which results in less influence of noise on the signal.

The performance of image reconstruction per land-cover class was evaluated for convolution operator 2 only (Fig. 3b), being the one with the most favorable results statistically (Fig. 1b) and visually (Fig. 4). In order to isolate performance per class, the corresponding GoF image was masked 10 times (once per class) and only specific class areas were analyzed. We chose to estimate an overall GoF measure per class by averaging GoF values over all bands, per masked subset, as
seen in Fig. 6a. According to this mean GoF measure, spatial homogeneity of classes does not necessarily play the expected dominant role in its influence on spectral match. For example, still waters (the lake in the park) are a relatively homogenous spectral entity per band, and still perform worse than all other classes. On the other hand, the class “buildings” is visually quite textured (see Fig. 4), but has a mean GoF value which is amongst the highest overall, reaching roughly 93% match, and scoring almost 10% better than water. The general trend of consistency of mean GoF measures per class, is opposite to their respective score, as illustrated in Fig. 6b. A decrease of some 8% in consistency is observed when changing from “still water” to “soil 1”. Crossing these results with the class map at Fig. 4b, one can generally observe that a relationship exists between the size of class objects in the image and their performance and consistency. Those forming relatively small objects perform worse than others. This may be explained by the fact that smaller objects contain a larger percentage of mixed pixels, which in turn decrease the overall match. Results of the correlation determination coefficient $R^2$ are summarized in Figs. 7–10 and represent performance of bands, classes and convolution operators.

When analyzing band performance by correlation ($r$) and determination ($R^2$) coefficients, we see that determination decreases by 5% at bands 3 and 4. This is apparent from their mean performance over all simulations, generated by the four convolution operators. When comparing band performance based on their correlation with class-wise corresponding real bands, band 3 is indeed of lower performance, but band 4 has a relatively high performance, scoring about 60% agreement. The variation between bands may be explained by the relative spectral change that occurs in these bands. To realize that, we suggest the reader to review the collection of spectral end-members in Fig. 2 above. We believe that the “red-edge” that occurs in bands 3 and 4 can cause linear mixing to affect the performance of these bands, as the convolution operation affects spectral change more drastically and therefore reduces overall band performance.

As done before, class-wise performance is evaluated based on convolution operator #2 only, as it visually expresses one of the best matches. Therefore, we conceive class performance an estimate, in this case, which represents the true performance of classes throughout all convolution operators. In order to isolate performance per class, the synthesized image was masked 10 times (once per class) and only the masked values were analyzed. In Fig. 8 we do see a relation between output class performance and class texture. A summary of assessment statistics on class performance is listed in Table 1.

Classes that form large image segments, and have a spatially homogenous nature score high with correlation determination coefficients, ranging 77–87% from water to grass, respectively. Classes with smaller spatial coverage are more affected

### Table 1

<table>
<thead>
<tr>
<th>Land-cover class</th>
<th>$n$</th>
<th>$R^2$</th>
<th>GoF</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buildings</td>
<td>1925</td>
<td>0.382</td>
<td>0.927</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td>Grass</td>
<td>720</td>
<td>0.88</td>
<td>0.931</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>New asphalt</td>
<td>2152</td>
<td>0.385</td>
<td>0.904</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>Weathered asphalt</td>
<td>319</td>
<td>0.223</td>
<td>0.925</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>Sand</td>
<td>19</td>
<td>0.558</td>
<td>0.916</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>Soil 1</td>
<td>287</td>
<td>0.599</td>
<td>0.939</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>Soil 2</td>
<td>191</td>
<td>0.803</td>
<td>0.923</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>Still water</td>
<td>60</td>
<td>0.768</td>
<td>0.846</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>Vegetation (type 3)</td>
<td>271</td>
<td>0.856</td>
<td>0.892</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>Trees</td>
<td>1413</td>
<td>0.823</td>
<td>0.879</td>
<td>0.085</td>
<td></td>
</tr>
</tbody>
</table>

$n$ is the number of pixels classified as class $i$, $R^2$ is the calculated correlation determination coefficient of that class with reference data, $\mu$ and $\sigma$ are the mean and standard deviation of the GoF image, respectively.

![Fig. 10. Synthesis performance according to the correlation determination coefficient ($R^2$). Naive corresponds to no convolution, and columns 1–4 correspond to convolution operators 1–4, respectively (see discussion at Section 2.3.1).](image-url)
by the convolution operator and therefore score between 55 and 60% in performance. Finally asphalts and buildings are performing worse, with up to 40% performance maximum. Correlations between real and simulated data for water, grass and buildings are illustrated in Fig. 9.

The reason for performance of each of the latter three classes, to our opinion, is due to different reasons. Weathered ("old") asphalt forms a small group of small, elongated segments that are highly sensitive to the mixture ratios, and performs the worse of all classes. Buildings are a classic example of the influence of under-estimation of texture, and new asphalt performs low because it entails a classification error in this case. The error is obvious when we assess visually Fig. 4(b–d) again. The "Yarkon" stream that runs through the bottom of the scene is erroneously classified as asphalt. Since convolution is a subsequent stage of classification, any misclassification error propagates forwards to convolution and to the final synthetic output. Surprisingly, this trend is not in line with the GoF results (class-wise). Let us note at this point that classes that scored high in the GoF test had the largest standard deviations.

Overall performance of convolution operators was evaluated by the correlation coefficient of the image as a whole, including all bands and all pixels (Fig. 10). As expected, the option of simulating imagery without mixing at all performs worse of all options, and scores 50% agreement. All options that are based on convolution operators perform significantly better by some 20–25%. In the range of 69–76% agreement ($R^2$), option 4 performs relatively low, and options 2, 3 and 1 perform better, respectively, reaching 76% determination at best. Samples of these convolution outputs and their corresponding correlation values are illustrated in Fig. 11.

5. Prospective applications

As we strive to improve texture and signal reconstruction, we already showed that synthetic spectral image generation by convolution operators, in the case of Tel-Aviv, reconstructs most of the original signal of a corresponding real image. As we discussed above, the added value of spectra to existing thematic simulators will be in its ability to map precisely the spatial distribution of materials and objects, their mixture abundance, and to relate these (sometimes, given additional data) to their corresponding physical attributes (approximate material weathering, pollution concentrations, albedo, short-wave absorption, vegetation indices, etc.). In the following, we suggest two prospective applications that can benefit from such a simulation tool.
5.1. Ground-truth images for classification algorithms

Reference data are generally needed to validate classification results. In many cases samples of ground truth are available by image interpretation, from the field or from libraries of spectra. In most cases, ground truth will contain some error that is referred to as “provider’s error” or “error of omission”. These are basically the result of interpretation errors and inclusion of mixed pixels to classification training samples. Ground truth images that provide the known mixture of known spectral end-members for each pixel of the image, are difficult to get and are time and effort consuming. To stress this point, the only way to generate such a truth image at present, is to visit every pixel of the remotely sensed image, geo-reference it, validate its corresponding land-cover components and estimate their respective fractional cover for that pixel. Alternatively, assume an image classified by two methods: \( j \) and \( k \). The tool we describe in this paper can generate a complete ground truth image based on \( j \) for which all spectra in all pixels are known exactly for their mixture fractions, as these were defined by known convolution operators. Classification method \( k \) can now use the ground truth image to calculate the true error of classification (rather than an estimate) and an error matrix based on the complete data set, rather than on spatially restricted samples. Spectral image synthesis therefore offers the availability of per pixel ground truth.

5.2. Cellular automata (CA) for urban and environmental planning

Cellular automata (CA) are discrete dynamical systems whose behavior is completely specified in terms of a local relation (neighborhood). A discrete cellular automaton can be thought of as a simplified universe (Itami, 1994). Space is represented by a uniform grid, with each cell containing one or more bits of data; time advances in discrete steps and the laws of the “universe” are expressed such that at each step, each cell computes its new state from that of its close neighbors. So, the system’s rules are local and uniform. Without the cellular structure, there would be infinitely many possible connections between components. Still, many possibilities can imply change in CA, by changing what is sometimes called “totalitarity of rules” (Rucker and Walker, 1997), weighting influence of neighbors according to their topological relationship with the candidate cell, or switching between rules in certain conditions.

Although several new studies and international projects focus on the contribution of the CA approach to geo-problems (Deadman and Brown, 1993; White and Engelen, 1997; Cheng, 1999; Vilaro, 1999; Cheng and Masser, 2002; Rabino and Laghi, 2002), no work is done, to our best knowledge, that acts directly on spectral data (i.e. multispectral or hyperspectral remotely sensed imagery) in a CA fashion. That is, even most recent projects are limited to translation of images to GIS grids of classified cells or vector files, and simulate dynamics on these.

The result of using the spectral simulation in this context is illustrated in Fig. 12. We generate two imaginary scenarios of urban development: the “grey” scenario and the “green” scenario. The first forecasts expansion of impervious surfaces on expense of open urban areas (soils, water bodies and various vegetation types), while the green scenario forecasts the unlikely opposite.

For the grey scenario we initiated from the classification used in Fig. 4 above and applied twice a \( 3 \times 3 \) local-minima filter on land-cover classes that form naturally urban open spaces. For the green scenario we initiated from the same classification but applied the local-minima filter on land-cover...
classes that form impervious surfaces. These include buildings and new and weathered asphalts. The grey scenario is illustrated in Fig. 12a and the green scenario in Fig. 12b. Spectral simulation of these future scenarios was the result of the same process we defined above for spatial—spectral reconstruction. Extracted sample spectra from an image simulated from the grey scenario are given in Fig. 12c. The change in the abundance of relevant land-cover classes is illustrated in Fig. 13.

Looking at the changes we clearly see that the green scenario generated an increase of roughly 50% for vegetation and soil classes (as well as for weathered asphalt) with comparison to the current real situation, that impervious surfaces were reduced from 60 to 40%, and that open spaces increased in size from 40 to 60%. In contrast, the grey scenario generated a varying decrease for soil, grass and trees ranging from 45 to 65%, an increase of about 25–40% in impervious

Fig. 14. Quantitative added value products derived from two imaginary development scenarios. The central column indicates the current state, the left column expresses the green scenario, and the right column expresses the grey scenario. Lines of figures from top to bottom indicate reflectance (band 5), SAVI, albedo and absorbed short-wave energy ($E_a$), respectively. See full discussion in the text.
surfaces and a decrease of 20% of open spaces. We now wish to add value to this product by extracting quantitative measures that can be derived from the spectral dimension of this simulation.

Multiple added value quantitative products can be derived from the spectral content of the image. Richter (2005), Roberts and Herold (2004), Lugassi (2002), Parlow (2002) and Voogt and Oke (2003) provide a set of possible products with relation to urban areas, to name but a few. To illustrate this point we will compare between the current reconstructed state (Fig. 4) and the two scenarios defined above in terms of the following quantitative aspects: the Soil-Adjusted Vegetation Index (SAVI) defined by Huete (1988), albedo and overall absorption of short-wave energy. Richter (2005) indicates the SAVI’s high compatibility for calculations of energy fluxes. It is expected will compare between the current reconstructed state (Fig. 4) to urban areas, to name but a few. To illustrate this point we provide a set of possible products with relation to urban areas, to name but a few. To illustrate this point we will compare between the current reconstructed state (Fig. 4) and the two scenarios defined above in terms of the following quantitative aspects: the Soil-Adjusted Vegetation Index (SAVI) defined by Huete (1988), albedo and overall absorption of short-wave energy. Richter (2005) indicates the SAVI’s high compatibility for calculations of energy fluxes. It is expected will compare between the current reconstructed state (Fig. 4) to urban areas, to name but a few. To illustrate this point we provide a set of possible products with relation to urban areas, to name but a few.

Three main factors may explain the differences observed when comparing outputs of the image simulation tool with the reference image (e.g. Fig. 4). These differences in fact limit the accuracy of this tool: texture under-estimation, the use of CA models and the use of convolution operators for spectral mixing.

Simulation of texture is currently discarded. It shows vaguely as a by-product of spectral mixing. That is why large objects of land-cover classes such as the lake and the grass park are texture-less from a distance of 1 border-pixel inwards. Attempts are being made to improve the texture appearance of the synthesized images (e.g. by using means and standard deviations of the reference image as indicators), but although improving correlation coefficients, these efforts do not show distinct visual improvements and are therefore not included in this software version. The use of CA models is limited in the sense that changing the so-called “totalitarity of rules”, or weighting differently the influence of neighbors on candidate cells, or even switching between transition rules under certain conditions, all modify the final output. As a result, even using the same transition rules in a CA in two different iterations can lead to two different outputs. It should be stressed, however, that the spatial forecasts made in the Tel-Aviv case study above were completely imaginary, and no attempt was made here to generate necessarily logical forecasts in terms of urban development.

Finally, the use of convolution operators restricts the accuracy of simulated outputs in the sense that a change in kernel values will result in different spectral mixtures and therefore to different results. Even though, Fig. 9 proves that by calculating correlation coefficients, reasonable estimates of success can be used that correspond to visual impression, in terms of selection of appropriate operators.
6. Summary and conclusion

In an effort to add spectral content to spatial simulators we offered an image synthesis process based on spectral mixture of library spectra of land-cover classes. We start by constructing a real image in order to evaluate the performance of the spectral image generator. The spatial distribution of the classes is assumed to be available by pre-defined non-fuzzy maps that allow 1 class per pixel ("concrete-classifiers"). Convolution operations on local neighborhoods are used to achieve spectral mixture in a linear fashion, and to generate synthetically images that correspond to remotely sensed images. Once the synthesis method is defined, we assessed spectral—spatial reconstruction by the GoF statistic and by correlation and determination coefficients. We finalized by two example applications: generation of ground truth data for classification, and spectral simulation of imaginary spatial situations, and derived quantitative measures.

The software itself is implemented as a new menu called "simulation" in the environment for visualizing images (ENVI 4.1) of Research Systems Inc. this menu is seen in Fig. 15. As can be seen the layout of processing modules follows the chronological stages of work. First, in the spectral simulator module we define inputs (i.e. a CA or classification map, a reference image and a spectral library), convolution and image generation/simulation, and possibly additional information concerning sensor and illumination geometry. Secondly a spatio-temporal simulator module is, in fact, a cellular automaton for generation of forecasted spatial scenarios. The validation module follows with the GoF, TGF and correlation tests for the simulated image. Additional validation tools are being developed for objective evaluation of CA outputs and for the spectral library information. Analysis is a module that combines qualitative image analysis means (classification, vegetation indices, etc.) and quantitative means for estimating material concentrations and quantities, when some reference data exist. Finally, the visualization module allows generating some basic diurnal and material-anisotropy animations.

As seen in the software layout, each module performs independently. This allows flexibility in repetition of processing steps, with outputs of one processing step being the inputs of the next one. The overall processing chain summarizing the main steps described in this paper, are illustrated in Fig. 16.

As for reconstruction quality measures, since it is observed that the total-goodness-of-fit statistic TGF is basically the same for all simulation outputs and goodness-of-fit is less reliable in cases of high class-scores, we believe this test to be insensitive to changes that are readily visible. In contrast, when assessing reconstruction of the signal for corresponding classes by correlation and determination analysis, we can explain the reason for $R^2$ scores, and generally relate them either to misclassification or to under-estimation of texture of classes.

Secondly we observe that spectral simulation of imaginary situations can be assumed to correlate strongly and positively ($r \cong 0.85$) and explain more than 70% of future reality. We believe that the automatic addition of texture to the synthesis process will improve these figures. It should be emphasized that we can, rather easily, generate scenarios (and corresponding images) that are not realistic at all, as shown in the example of a cellular automaton.

We note that although we tried to isolate buildings in the classification stage, it is clear that any synthesis that is based, at some stage, on a concrete-classification result will be dependant on the accuracy of definition of classes. This means, in turn, that any effort to introduce land-use classes such as urban or built-up to a land-cover classification map will probably include soil and vegetation samples that will decrease the overall performance of the synthesis by those pixels that are misclassified. We finally conclude that the addition of the spectral dimension to conventional cellular automata simulation tools, allows new quantitative parameters to be extracted for forecasts related to the influence of urban development on the environment. Expansion of this method to 3D visualization may be the key factor to deal with complex urban environments.

Acknowledgements

This research is supported by the Sapir Fund for applicative municipal research, under management of the Pais Foundation (Mifal Ha-Pais), Israel.

References


Richter, R., 2005. Atmospheric/topographic correction for airborne imagery. (ATCOR-4 user guide, version 4.0, DRL-IB 565-02/05, DLR Wessling, Germany).


