MULTISCALE OBJECT-SPECIFIC ANALYSIS: SCALE PROBLEMS AND MULTISCALE SOLUTIONS

Ola Hall1, Geoffrey J. Hay2 and Danielle, J. Marceau2

1Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, 02139 Massachusetts, ohall@mit.edu.
2Geocomputing Laboratory, Département de Géographie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J, ghay@sympatico.ca, danielle.marceau@umontreal.ca

Abstract

Based on a series of empirical studies beginning in the 1970’s, it was noted that remote sensing data suffered from the scale and aggregation problem. It was further recognized that there was no unique or ‘optimal’ spatial resolution for detecting the different sized, shaped, and spatially arranged entities represented in a remote sensing image of a complex scene. Today within the Earth sciences, it is strongly recognized that landscapes exhibit distinctive spatial patterns associated to different processes at different scales. Consequently, multiscale approaches are required for modern landscape analysis. It is within this context that the Multiscale Object-Specific Analysis (MOSA) framework was developed. In this paper we review the background, foundations, and recent developments of MOSA. We begin with the original definition of Object-Specific Analysis (OSA) and Object-Specific Upscaling (OSU), and continue with the recent integration of Marker Controlled Watershed Segmentation (MCS) as an automatic feature detector.

INTRODUCTION

With the recognition that landscapes can be modeled as complex systems consisting of heterogeneous spatial components, non-linear interactions, emergence, self-organization and scale-dependence, scale and scaling issues represent increasingly important areas of research among the earth sciences (Wu and Hobbs, 2002). Conceptually, scale represents the window of perception, the filter, or the measuring tool through which a landscape may be viewed or perceived (Levin, 1992). Thus, changing the scale changes the patterns of reality, which has obvious implications for understanding the dynamics of any environmental system. In the context of remote sensing, scale corresponds to spatial resolution, which refers to the ability of a sensor to record and display fine spatial detail as separated by its surroundings. Remote sensing imagery also encapsulates two important aspects of scale: grain and extent. Grain represents the finest distinction that can be made in an observation set (i.e. spatial resolution), while extent corresponds to the span of all detected entities (Allen and Hoekstra, 1992) (i.e. the total area within an image swath).

Recently, resolving the challenge of multiscale analysis has been introduced as essential to the understanding of complex systems, such as landscapes. The rationale for this is that landscapes are known to exhibit distinctive spatial patterns associated to different processes at different scales. Therefore, landscape analysis performed at a single scale is insufficient for understanding multiscale patterns and processes. In addition, there is no way to a priori define a unique ‘optimal’ scale for analysis (Marceau et al., 1994; Wu et al., 2000). Presently, a small number of techniques that allow for multiscale analysis exist (for a
review see, Hay et al., 2003). Here we will focus on a novel framework, referred to as \textit{multiscale object-specific analysis} (MOSA) that is specifically developed for remote sensing data. Consequently, the objective of this paper is to provide a description of MOSA along with a brief description of the ‘scale’ concept in remote sensing.

**BACKGROUND TO OBJECT-BASED MULTISCALE APPROACHES**

Significant interest in scale issues in remote sensing began with a series of empirical studies in the mid seventies (for a detailed review see Marceau and Hay, 1999). The prime conclusion reached from several independent studies was that a change in spatial resolution could significantly affect classification accuracies, and that in many cases, the use of successively higher spatial resolution data resulted in lower overall classification accuracy (Marceau and Hay, 1999). This was generally explained with an increase in within-class spectral variability that confused per-pixel classifiers.

On the basis of these findings, Marceau (1992) undertook a review encompassing two decades of thematic mapping studies involving remotely sensed imagery, and revealed the following observation: within a given classification consisting of several land-cover/land-use types, there were often considerable inconsistencies in the results obtained from one class to another. In an attempt to evaluate which factors were responsible for the classification accuracies of nine land covers, using the grey-level-co-occurrence matrix method, it was found that 90\% of the variability in the classification results was explained by the \textit{window size} employed for the co-occurrence computation. Furthermore, the best classification accuracy for each class was achieved with \textit{different window sizes}.

Searching for a broad framework to link scale and automated feature extraction results, Marceau (1992) hypothesized that a parallel existed between inconsistency in the results obtained when classifying remote sensing data and the serious problems involving the manipulation of arbitrarily defined areal data known as the modifiable areal unit problem (MAUP). Later, Marceau et al. (1994) argued that the acquisition of remote sensing data is a particular case of an arbitrary uniform spatial sampling grid used to obtain measurements about geographical entities that induces the scale and aggregation effect responsible for haphazard analysis results. To confirm this, they conducted an empirical investigation to verify the impact of spatial resolution and aggregation level on classification accuracy of forest data. Their results indicated that per-class accuracies were considerably affected by changing scale and aggregation level. This led to the conclusion that remote sensing data are not independent of the sampling grid used for their acquisition, and that neglecting the scale and aggregation level can produce haphazard results having little correspondence with the geographic entities of the scene. They also noted that there is no unique spatial resolution appropriate for the detection and discrimination of all geographical entities composing a complex natural scene, and advocated that classification based on the use of a unique, or ‘optimal’ spatial resolution should be replaced by a \textit{multiscale} approach.

**Solutions to the scale problem in remote sensing**

Since the MAUP was well documented in the social sciences, possible ways to overcome its effects had already been proposed. Fotheringham (1989) listed five different significant solutions: the derivation of optimal zoning systems, the identification of basic entities, the development of spatial statistics (as opposed to traditional statistics), sensitivity analysis, and the search for fluctuations in variables and relationships with scale.
The optimal spatial resolution approach had been inspired by the search for optimal zoning systems in the social sciences. Openshaw (1984) proposed an approach where the selection of areal units becomes an integral part of the goal of a particular spatial analysis. The richness of this approach is that the spatial units are not only defined to fulfill the particular needs of the analysis, but also to be geographical meaningful. Other paths that have been explored include traditional parametric statistical analysis. In remote sensing, it is increasingly acknowledged that traditional statistics have severe limitations since they do not take into account the properties of locational data. Such data are generally spatially auto-correlated, non-stationary, non-normal, irregularly spaced, and discontinuous, while standard statistical pattern-recognition techniques assume independent and random data, and a normal distribution. Another solution proposed to overcome the MAUP is to simply recognize its existence and to report the sensitivity of analytical results to variations in both the scale and level of aggregation. To do so requires that the data used can be aggregated and that results can be obtained for higher levels.

In a novel structural approach to quantify the three dimensional image-texture of a forest canopy, Hay (1993) and Hay and Niemann (1994) incorporated location-specific primitives, and an object-specific variable sized and shaped moving kernel to regularize images based on objects spatial characteristics. The importance of this work was its introduction of varying sized and shaped object-specific filters to extract individual entities from remote sensing imagery. As reported by Fotheringham (1989), the identification of basic entities perhaps provides the clearest way out of the MAUP because it is a product of aggregation that represents meaningful real-world entities. In the following section, the development of these ideas into MOSA framework will be described.

MULTISCALE OBJECT-SPECIFIC ANALYSIS

MOSA is composed of three primary components: object-specific analysis (OSA), object-specific upscaling (OSU), and marker-controlled watershed segmentation (MCS). The original formulation of OSA/OSU is described in detail in Hay et al. (2001). MCS was integrated as an automatic feature detector for use with object-specific output, but was not part of the original formulation (Hall, 2002; Hall and Hay, 2003; Hall et al., 2004). OSA is a multiscale approach that automatically defines unique spatial measures specific to the individual image-objects comprising a remote sensing scene (Hay et al., 2001; Hay et al., 1997). These object-specific measures are then used in a weighting function (OSU) that automatically upscale an image to a coarser spatial resolution, by taking into account the spatial influence of the image-objects composing the scene at a finer spatial resolution. Then MCS is applied to these upscaled data to automatically segment them into topologically discrete image-objects that correspond to visually defined image-objects (Hall et al., 2004).

Foundations of OSA/OSU

The term ‘image-object’ (Hay and Niemann, 1994) refers to individually resolvable entities located within a digital image that are perceptually generated from high-resolution pixel groups. High-resolution corresponds to a situation where real-world objects are modeled by many individual pixels (Woodcock and Strahler, 1987). The term low-resolution represents the spectral integration of many (smaller) real-world objects within a single pixel. As a result, an image-object tends to be composed of spatially clustered pixels with high autocorrelation because they are part of the same object (Hay et al., 2002). In a satellite sensor image, each pixel exhibits both high- and low-resolution duality.
An underlying premise of OSA/OSU is that all pixels are exclusively considered as high-resolution. Thus pixels are used to define the size of the image-object they are part of. In practice, this is done by evaluating the variance within an iteratively growing (round) window centered over each pixel until a series of heuristics are met (Hay et al., 2001). Based on these heuristics, a threshold is defined as the kernel reaches the image-object’s edges. The window size at this location corresponds to the objects known size. When this threshold is reached, corresponding mean and variance values are recorded for the pixel under analysis. This process is repeated for all pixels within the original image, resulting in corresponding variance ($V_I$), area ($A_I$), and mean ($M_I$) images (Figure 1b-d). This constitutes OSA.

OSU represents an upscaling procedure that incorporates the area values defined for each pixel as part of a weighted resampling scheme. Data are upscaled to a coarser resolution based on either a user-defined value, or a resampling heuristic based on the relationship between pixel size, image-objects, and a generic point-spread function property of sensors. When OSA is iteratively applied to the resulting mean image, dominant landscape objects start to visually emerge at each new iteration. At the first OSA iteration every pixel is assessed within a progressively growing window until a local maximum variance threshold is reached. When this process is applied to the entire image the first image-set (IS$_1$) is

![Image Set](image.png)

Figure 1: This image set presents the two first scale domains (SD1-2) of MOSA processing. Figure 1a is the original subset of a IKONOS-2 (1m spatial resolution) scene acquired over Norra Djurgården, Stockholm, Sweden in August 2001. Figure 1b-d illustrates the Variance, Area and Mean images of the first scale domain. Figure 1 e-f illustrates the result after MCS has been applied. Note that OSA/OSU detects structures in semi-managed grassland in image-left. Figure 1g-k illustrates the same processing but for the second scale domain. Note how image-objects start to merge into larger objects as scale domains increase.
defined, consisting of $V_1$, $A_1$, and $M_1$ (as described above). In the second iteration, each pixel in the newly generated $M_1$ is assessed until a local minimum variance threshold is reached. The resulting images become the second image-set (IS$_2$) (i.e. $V_2$, $A_2$, and $M_2$) where they represent the beginning scale of all newly emergent image-objects (Figure 1b-d). Thus, odd-numbered OSA iterations define the ‘end’ of objects, while even-numbered iterations define the beginning of the next emergent object that each pixel is part of. All $M_i$ generated from even-numbered OSA iterations are then selected for upscaling. The whole process is iterated until the number of pixels is too small for further processing. Conceptually, this form of processing will eventually result in the entire satellite scene being represented by a single pixel. The resulting image sets form a nested hierarchy composed of variance, area and mean images that have membership in a unique scale domain (SD$n$). Within each SD$n$, each image shares the same grain and extent. Furthermore, each SD$n$ is also a member of a scale-domain super set (SDS) that represents the entire range of OSA and OSU evaluated within a unique satellite image (Hay and Marceau, 2004).

**Additional developments – automatic object delineation**

One of the limitations to the early OSA/OSU framework was the lack of an automatic multiscale feature detection. Thus, in a collaborative effort, Hall et al. (2004) suggested that marker controlled watershed segmentation (MCS) could be integrated with the object-specific framework suggested by Hay et al. (2001). Therefore, once OSA/OSU processing was completed, MCS was used as a feature detector and applied to the resulting images. MCS is a watershed transformation technique that detects regional similarities as opposed to an edge-based technique that detects local changes (Meyer and Beucher, 1990). The key characteristic of MCS is the ability to reduce the problem of over-segmentation found in the original watershed transformation by placing markers or ‘seeds’ in user specified areas, from which discrete watershed perimeters can be defined. The general procedure involves three steps:

1. Enhancing variations in image intensity with a gradient operator (i.e. an edge detector) resulting in a gradient image.
2. Defining a relevant marker set and incorporating it into the gradient image.
3. Applying the watershed transformation to this modified gradient image.

In recent studies that integrated OSA, OSU and MCS (Hall and Hay, 2003; Hall et al., 2004), Variance images were used as gradient images, rather than haphazardly choosing a gradient operator and (static) kernel size from which to generate gradient images. Then unique markers were generated for each scale domain by automatically defining regional minima in the corresponding Area images, as these datasets explicitly represent object-specific information. The resulting markers were then embedded in the gradient image using a simple operator, and the Matlab watershed algorithm was applied. Each polygon in the resulting watershed images (W$_i$) was then labeled with a value equal to the average of the corresponding $M_i$ pixels located within its perimeter.

While these recent results were significantly better than haphazardly generating a gradient image and using the regional image minima as markers, it was recognized that there was still a problem with over-segmentation, even if it was all object related. In a new effort to reduce over-segmentation, while producing watersheds that visually correspond well to image-objects, pre-processing of the object-specific data with a median filter is incorporated, new gradient images are generated instead of using the Variance images, and
a more object-specific approach to automatically define markers is developed (Hay and Marceau, 2004).

In this new step a 3 x 3 median filter is applied to all images (V1, A1, M1, and O1). Recall, that in OSA processing the maximum high-resolution content is maintained in all images, including image noise (if present). In the object delineation portion of MOSA, there is no longer interest in single pixels, but rather pixel groups that represent unique objects. When this is taken into consideration, along with the fact that the smallest object-specific kernel resides within a 3 x 3 pixel window, and that edges larger than this is preserved due to the characteristics of the median filter, this is an effective and straightforward approach for defining pixel groups and excluding noise.

Rather than using the variance images as gradient images (G1), new gradient images are generated for each scale domain by subtracting the Mean image from the corresponding resolution Upscale image (U1) and defining the absolute value of the result. This method for generating new gradient images is similar to the technique in mathematical morphology where external contours can be created by defining the difference between the original and the dilated image. For the gradient images produced for this paper the following equations were used according to Hay and Marceau (2004).

\[
G_2 = \text{abs}(O_1 - M_2) \quad (1)
\]
\[
G_4 = \text{abs}(U_1 - M_4) \quad (2)
\]

Where G represents the newly generated gradient images, abs represents the absolute value, O1 represents the original digital image M represents the newly generated mean images, and U1 represent the newly generated Upscale images. This is applied to each new median filtered dataset (that was originally generated by OSA/OSU).

Object markers are then generated by combining regional minima from the corresponding variance and area images using a logical AND operation. This results in a binary image, where values equal to one represent regional minima. Variance minima values represent areas of low heterogeneity that conceptually correspond to object centers. Area minima indicate that the object heuristics for the pixel being assessed were met within a small analyzing kernel and also correspond to object centers. Based on an extensive visual analysis of the OSA/OSU results, it became evident that the local Area minima represent both image-object centers and the edges between two or more image-objects. Exclusively using markers derived from Area minima - as done in earlier studies (Hall and Hay, 2003; Hall et al., 2004) - does not provide optimal results. Fortunately only image-objects are composed of both (relatively) small area, and small variance values. Thus to ensure that image-objects, rather than edge-objects are defined as markers, the AND logical operator is applied to the regional Area minima and the regional Variance minima datasets. This produces a combined binary marker dataset.

To define individual image-objects, the new (combined) marker sets were embedded within the corresponding gradient image. More explicitly, the location of each marker set was defined within the appropriate gradient image using the Matlab `imimposemin` function. This function modifies the intensity image using morphological reconstruction so the intensity image only has regional minima wherever the binary (marker) image is nonzero. By applying the watershed algorithm (Vincent and Soille, 1991) to each ‘embedded’ image, watershed boundaries separating image-objects are generated by this algorithm (Figure 1e and 1j).
Each pixel in the Mean images represents a member of a newly detected image-object. Since these images are generated from average values calculated within unique threshold kernels, they represent the dominant image structure defined at a specific spatial resolution within a unique scale domain (Hay et al., 2001). Therefore, each newly defined - though empty - watershed polygon is used as a mask to generate a value equal to the average of the corresponding Mi pixels located within its perimeter (Figure 1f and 1k). In essence, each watershed polygon now spatially represents the average grey-tone, and areal extent of a unique image-object. This step is referred to as ‘object labeling’ and represents the automatic delineation of discrete multiscale image-objects in each scale domain. This procedure was also applied by Hall and Hay (2003) and Hall et al. (2004). In all cases, the new methods can be automatically implemented using data inputs that continue to be met by the Variance, Area and Mean images.

CONCLUDING REMARKS

Wu and Hobbs (2002) have argued that scale and scaling are at the top of the research agenda in Landscape ecology. However, questions yet to be addressed include how to determine appropriate scales for understanding unique landscape patterns and processes, and how to scale up and down across heterogeneous landscapes. Based on a number of empirical studies (Hay and Niemann, 1994; Hay et al., 1997; Marceau, 1992; Marceau et al., 1994) and theories from Complex systems and Hierarchy theory (Allen and Hoekstra, 1992), we recommend Multiscale Object-Specific Analysis as a unique multiscale modelling and methodological framework for understanding and analyzing remotely sensed landscapes. MOSA represents an integration of Object-specific Analysis (OSA), Object-specific Upscaling, and Marker Controlled watershed Segmentation. This framework for automatic multiscale scene generation and feature extraction allows dominant image-objects to emerge at their respective scales. Most important, it requires no a priori scene information.

The present research agenda for MOSA involves automatic classification of image-objects, linking of image-objects through scale, and continued, evaluation of algorithm performance in relation to different landscape types.

ACKNOWLEDGEMENT

This research has been funded by a post-doctoral fellowship awarded to Dr Ola Hall from the Knut and Alice Wallenberg Foundation, by a NSERC (Natural Sciences and Engineering Research Council of Canada) and a GREFI (Groupe de recherche en écologie forestière interuniversitaire) post-doctoral fellowship to Dr Geoffrey Hay, and by a NSERC research grant awarded to Dr Danielle J. Marceau.

REFERENCES


