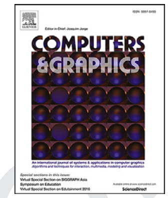
Computers & Graphics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag



Technical Section

Segmentation of aerial images for plausible detail synthesis☆

Q1 Oscar Argudo[a,*], Marc Comino[a], Antonio Chica[a], Carlos Andújar[a], Felipe Lumbreras[b]

[a] ViRVIG, Computer Science Department, Universitat Politècnica de Catalunya, Jordi Girona 1-3, Barcelona, Spain
Q2 [b] Computer Vision Center, Department of Computer Science, Universitat Autonòma de Barcelona, Edifici O, Bellaterra, Spain

A R T I C L E I N F O

Article history:
Received 31 July 2017
Revised 6 November 2017
Accepted 16 November 2017
Available online xxx

Keywords:
Terrain editing
Detail synthesis
Vegetation synthesis
Terrain rendering
Image segmentation

A B S T R A C T

The visual enrichment of digital terrain models with plausible synthetic detail requires the segmentation of aerial images into a suitable collection of categories. In this paper we present a complete pipeline for segmenting high-resolution aerial images into a user-defined set of categories distinguishing e.g. terrain, sand, snow, water, and different types of vegetation. This segmentation-for-synthesis problem implies that per-pixel categories must be established according to the algorithms chosen for rendering the synthetic detail. This precludes the definition of a universal set of labels and hinders the construction of large training sets. Since artists might choose to add new categories on the fly, the whole pipeline must be robust against unbalanced datasets, and fast on both training and inference. Under these constraints, we analyze the contribution of common per-pixel descriptors, and compare the performance of state-of-the-art supervised learning algorithms. We report the findings of two user studies. The first one was conducted to analyze human accuracy when manually labeling aerial images. The second user study compares detailed terrains built using different segmentation strategies, including official land cover maps. These studies demonstrate that our approach can be used to turn digital elevation models into fully-featured, detailed terrains with minimal authoring efforts.

1. Introduction

Publicly available Digital Terrain Models (DTM) and aerial images have opened new possibilities for using real scenarios in video games and entertainment applications. These 2.5D models are readily usable for rendering aerial views of the scenes, but lack resolution and 3D appearance for close-up views. As a consequence, artists often need to enrich DTMs with synthetic detail, such as procedural bedrock and rocks for the ground, realistic water shaders for lakes and rivers, and fully-detailed plant models for the vegetation. When done manually, a substantial amount of effort is required to locate the different elements in the aerial images and to apply a suitable detail synthesis technique to them.

In this paper, we address the problem of segmenting high-resolution (25 cm/pixel) aerial images into small sets of classes suitable for detail synthesis (Figs. 1 and 2). The *segmentation for synthesis* problem (*S4S* from now on) exhibits a number of unique issues that we summarize below. First, categories in S4S are defined according to the different detail synthesis techniques artists might want to apply. For example, in a particular desert scene for a Dakar rally game, one artist might want to distinguish rock, sand, cacti and palm trees, whereas in a tropical forest scene we could be interested in segmenting vegetation and rivers.

Second, we want to give artists the possibility to add new classes dynamically. This way artists can progressively refine the appearance of different materials, which due to their variety are hard to know in advance (e.g. forest, shrub, grass, crops, sand, bare rock, scree, water courses, inland marshes, snow...) and can decide to distinguish non-anticipated categories (e.g. deciduous forest from coniferous forest). This flexibility means that the pixel classifier should be able to work with relatively small training sets (containing examples from a varying set of classes) and that both training and classification times should be within the range of a few minutes. Notice also that we cannot assume balanced classes in the segmented exemplars, neither the exemplar class distributions to be representative of the true class distributions.

In this context, generating large and varied training sets is unfeasible. Manual image segmentation requires a substantial amount of effort. For example, a 25 cm/pixel image covering 1 km^2 contains 16 M pixels. Even when using the advanced tools found in state-of-the-art object-based image processing applications (multiresolution segmentation, superpixels, manual relabeling...),

☆ This article was recommended for publication by Prof. J. Jorge.
* Corresponding author.
E-mail address: oargudo@cs.upc.edu (O. Argudo).

https://doi.org/10.1016/j.cag.2017.11.004
0097-8493/

Fig. 1. Renders of detailed terrains created from aerial images. The segmentation of the aerial image allows for the application of different techniques depending on the soil cover. In these examples, synthetic trees, bushes and grass for vegetation areas, fractal displacement for the bare soil, rocks and gravel shader for unpaved roads, and specific shaders for snow, lakes and rivers. The last image shows an example of landscape and visual impact assessment.

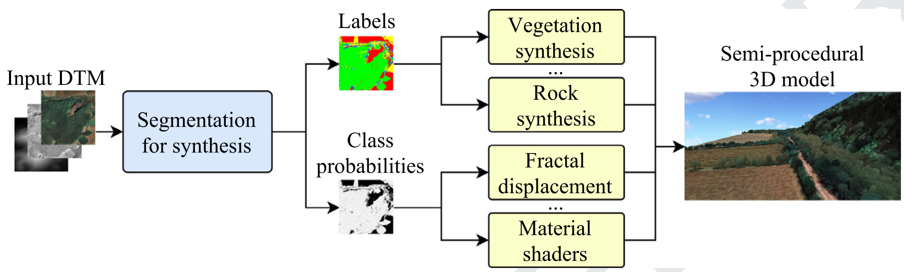


Fig. 2. We address the problem of adding realism to digital terrain models through the segmentation of aerial images (blue box) into a suitable set of classes (e.g. vegetation, rock, water). This way each class can be rendered using specific shaders and procedural content (yellow boxes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

complete segmentation of a 1 km^2 image can take several hours of an expert human classifier. Moreover, we will show that resulting labels largely depend on the labeler's judgment and thus cannot be safely taken as ground truth.

This manual labeling effort can be largely alleviated by segmenting only a collection of easy-to-label uniform regions (see Fig. 9) and taking the training examples from these regions using a suitable sampling strategy. We adopt this approach (partial, region-based classification) as the only feasible approach in the context of dynamic classes.

Although there is an extensive literature on image segmentation in remote sensing, previous approaches either require extensive training sets (unfeasible in the context of dynamic categories), use descriptors tuned for very specific categories (e.g. crops), assume balanced, representative datasets, or rely on expensive-to-train classification algorithms.

To the best of our knowledge, this is the first work exploring the optimal components of a standard image segmentation pipeline specifically tailored for segmentation-for-synthesis.

The key contributions of the paper are:

- A complete pipeline for training and inferring per-pixel labels from a dynamic, user-defined set.
- A performance comparison of state-of-the-art machine learning algorithms for S4S.
- An analysis on the contribution of different pixel descriptors (at varying resolution levels) in the classifier accuracy.
- A discussion on different strategies for sampling the training set from partially-segmented exemplars.
- A user study analyzing human accuracy when manually labeling uniform regions in aerial images. We estimate the difficulty of the regions, the expertise of the labelers, and the true label of the regions.
- A second user study demonstrating the effectiveness of our approach. We asked users to compare renders built from images segmented using either our approach or official land cover maps.

2. Previous work

Despite the specific nature of the S4S problem discussed above, we rely on the great amount of work already done in the field of aerial image analysis and on typical strategies of region analysis based on color and texture.

Classical image segmentation techniques compute per-pixel features based on image values in the pixel's neighborhood. Many different pixel descriptors have been proposed both for general images and remote sensing images. Ruiz et al. [1] compared different texture and spectral feature descriptors for pixel classification of remote sensing images. Results on different forest scenes showed that there is no universal criteria – the suitable set of features depends on the type of landscape units defined in each application. Similarly, dos Santos et al. [2] also compared the effectiveness of various color and texture descriptors for image classification. Although their task was not pixel based, the best descriptors were also dependant on the type of input dataset. Tokarczyk et al. [3] compare classical feature sets with feature banks computed with the first layers of deep networks. Their results show that features based on patches dominate over those based on individual pixels, i.e. texture holds important information in high-resolution images. However, complex feature extraction methods or even non-linear feature learning yield small or no improvement, while adding a significant computation cost. Penatti et al. [4] studied whether Convolutional Neural Networks trained for classification of everyday objects generalize to aerial and remote sensing images. They used the output of the last fully-connected layer of two networks (OverFeat and CaffeNet) as features, and compared with low-level feature descriptors. On the aerial dataset, deep features achieved the best results, but for remote sensing images they were outperformed by the low-level descriptors. Since we do not know in advance the specific classes artists will require, and training sets are expected to be too small for end-to-end learning, we use a large number of color and texture features from the literature, combined with height, slope and gradient descriptors from the DTM.

Please cite this article as: O. Argudo et al., Segmentation of aerial images for plausible detail synthesis, Computers & Graphics (2017), https://doi.org/10.1016/j.cag.2017.11.004