Computer Vision and Image Understanding 146 (2016) 86–98

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

# Hybrid macro–micro visual analysis for city-scale state estimation

Ken Sakurada[a,*], Takayuki Okatani[a], Kris M. Kitani[b]

[a] Graduate School of Information Science, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai-shi, Miyagi, Japan
[b] Robotics Institute, Carnegie Mellon University, 5000 Forbes Ave Pittsburgh, PA 15213, USA

ARTICLE INFO

*Article history:*
Received 17 March 2015
Accepted 29 February 2016

*Keywords:*
City-scale
Vehicular imagery
Aerial imagery

ABSTRACT

We address the task of estimating large-scale land surface conditions using overhead aerial (macro-level) images and street view (micro-level) images. These two types of images are captured from orthogonal viewpoints and have different resolutions, thus conveying very different types of information that can be used in a complementary way. Moreover, their integration is necessary to enable an accurate understanding of changes in natural phenomena over massive city-scale landscapes. The key technical challenge is devising a method to integrate these two disparate types of image data in an effective manner, to leverage the wide coverage capabilities of macro-level images and detailed resolution of micro-level images. The strategy proposed in this work uses macro-level imaging to learn the extent to which the land condition corresponds between land regions that share similar visual characteristics (e.g., mountains, streets, buildings, rivers), whereas micro-level images are used to acquire high resolution statistics of land conditions (e.g., the amount of debris on the ground). By combining macro- and micro-level information about regional correspondences and surface conditions, our proposed method is capable of generating detailed estimates of land surface conditions over an entire city.



## 1. Introduction

We address the task of estimating large-scale land surface conditions using overhead aerial (macro-level) images and street view (micro-level) images. These two types of images are captured from orthogonal viewpoints and have different resolutions, thus conveying very different types of information that can be used in a complementary way. Moreover, their integration is necessary to enable an accurate understanding of the changes in natural phenomena over massive city-scale landscapes.

Aerial images are an excellent source for collecting wide-area information of land surface conditions. However, it may come at the cost of a lower resolution (i.e., number of pixels per square meter) and visibility may drastically change depending on the weather. For example, clouds may obscure the extent to which the land surface is visible (Fig. 1). A more important limitation of aerial images is that they are limited to a vertical (top-down) perspective of the ground surface, such that areas occluded by a roof or highway overpass are not visible to the camera (first and second row of Fig. 2) making it difficult to estimate land conditions in covered areas.

Street-view images, on the other hand, are captured at ground level and can provide higher resolution images of vertical structures and have unobstructed access to visual information in covered areas. They are also less affected by weather conditions. However, street-view images are constrained to the ground plane and a single image has limited physical range. Moreover, it is also labor intensive to acquire street-level images of large land surface areas (i.e., millions of square meters).

The key technical challenge is devising a method to integrate these two disparate types of image data in an effective manner, to leverage both the wide coverage capabilities of macro-level images and detailed resolution of micro-level images. The strategy proposed in this work uses macro-level imaging to learn the extent to which the land condition corresponds between land regions that share similar visual characteristics (e.g., mountains, streets, buildings, rivers), whereas micro-level images are used to acquire high-resolution statistics of land conditions (e.g., the amount of debris on the ground). By combining the macro- and micro-level information about regional correspondences and surface conditions, our proposed method generates detailed estimates of land surface conditions over an entire city.

* Corresponding author. Fax: +81 22 795 7015.
*E-mail address:* sakurada@vision.is.tohoku.ac.jp (K. Sakurada).

http://dx.doi.org/10.1016/j.cviu.2016.02.017
1077-3142/

**Fig. 1.** Aerial images affected by weather conditions (Left: March 11, 2011, Right: March 31, 2011). For example, the land surface may be covered by clouds, thereby drastically changing illumination conditions in an aerial image.

**Fig. 2.** Examples of aerial and street-view images. There are many cases in which aerial images and street-view images provide complementary information about the land surface condition. For example, areas obscured by the building roof (top and second row) and stacked objects (bottom row) are best viewed from the street.

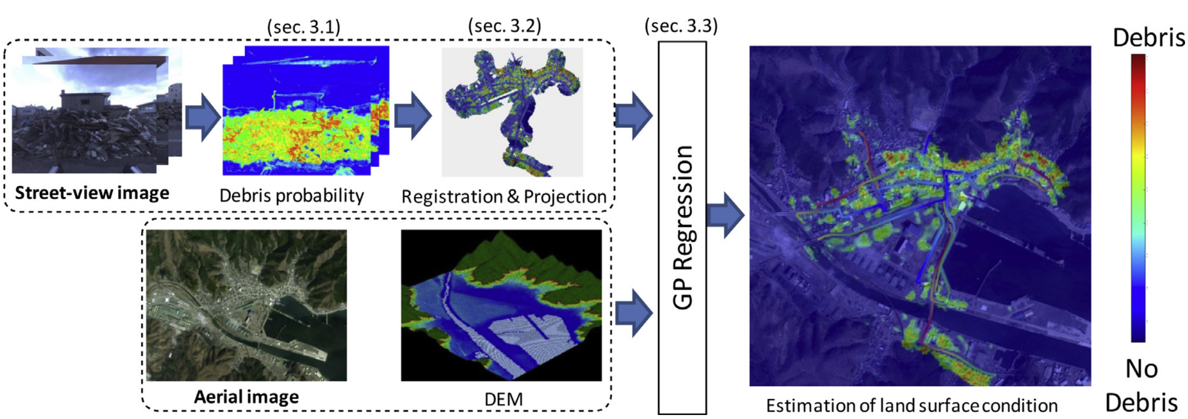


**Fig. 3.** Data flow diagram of city-scale estimation of land surface condition. Our approach efficiently integrates both micro (street-view) and macro-level (aerial) images along with other forms of meta-data to estimate the city-scale land surface condition.

The technical contribution of this paper is a novel procedure for generalizing from a sparse set of visual recognition results to a large-scale land condition regression estimate. The proposed system carefully brings together state-of-the-art algorithms for semantic scene understanding, structure-from-motion and non-parametric regression to generate a massive city-scale land condition probability map (Fig. 3). To the best of our knowledge, this is the first work of its kind to use sparse image-based street-level object recognition results to extrapolate the surface conditions of an entire city (over 4 million square meters).

Although our method can be generalized to accommodate different types of large-scale phenomena, we ground our proposed