ARTICLE IN PRESS
JID: YCVIU [m5G;February 15, 2017;3:41]

Computer Vision and Image Understanding 000 (2017) 1–17

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Temporal city modeling using street level imagery

Ken Sakurada[a,*], Daiki Tetsuka[b], Takayuki Okatani[b,c]

[a] Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, Japan
[b] Graduate School of Information Science, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai-shi, Miyagi, Japan
[c] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan

ARTICLE INFO

Article history:
Received 9 December 2015
Revised 23 January 2017
Accepted 30 January 2017
Available online xxx

Keywords:
Change detection
SfM
CNN
City-scale
Vehicular imagery
City map

ABSTRACT

Estimation of the temporal changes to a city is useful for city management, disaster recovery operations, and understanding natural phenomena. When several types of data are available for this task, the optimal type should be chosen depending on the changes that need to be detected. However, data of the desired type are not always available, particularly historical data. In this study, we propose two methods for detecting changes in a city, which can be used in complement to process available data types and detect changes in selected targets. The first method estimates the presence of buildings by comparing street-level images and a 2D city map of buildings created at different points in time. This method uses the Structure from Motion (SfM) technique to reconstruct a point cloud of the structures of the city, and matches the point cloud with the 3D building structures recovered from its 2D map. While 2D city maps are available for most cities, most are not very accurate. Therefore, this method is designed to overcome these inaccuracies and thus is widely applicable. On the other hand, the method cannot detect the following types of scene change: wall paintings, buildings that were reconstructed and closely restored to their previous shape, pedestrians, cars, and vegetation. The second method uses a pair of street-level images that are roughly aligned with GPS data collected at different points in time to detect such scene changes. This method uses the features of a convolutional neural network (CNN) in combination with superpixel segmentation to address inaccurate image alignment and it also enables change detection with pixel-level accuracy. Additionally, the second method is scalable for large-scale estimation because it can quickly detect scene changes by merely using an image pair without performing large-scale SfM. The authors consider the proper use of these two methods to enable temporal city modeling in various situations. We experimentally apply these methods to cities damaged by the tsunami that struck Japan in 2011 and the results show their effectiveness.

1. Introduction

This paper describes two novel methods for estimating temporal change in a city using street-level images. The first method estimates the existence of buildings by using street-level images and a 2D city map taken from a different time period. The second method detects scene changes from a street image pair taken at different points in time.

Temporal city modeling is one of the most important tasks for managing a city, recovering from disasters, and understanding natural phenomena. Accurately understanding the changes in a city requires detecting of changes between data collected at different times. However, data obtained at a specific point in time, especially historical data, are not always available. Indeed, for the cities damaged by the tsunami that struck Japan in 2011, no street images of any type (e.g., Google Street View) are available before the tsunami; the only available data are 2D city maps. Consequently, detecting changes in textures (e.g., wall paintings) and objects not present on a city map (e.g., pedestrians, cars, and vegetation) inevitably requires the comparison of images taken at different points in time, if they are available. As an example, if the shape of a reconstructed building has remained largely unchanged, it is difficult to detect the changes resulting from its reconstruction from the street images and 2D city map data taken at different time points, because the shape of the 2D city map is neither sufficiently accurate to detect small changes nor does it contain textural information.

Therefore, we propose two methods for detecting city changes depending on the available data types and the target changes. Fig. 1 shows our strategy for different types of input data. We can estimate building-level changes when the street image for time t and 2D city map for time t' are available. Fig. 2 presents visual images of a city prior to the tsunami and the subsequent changes

* Corresponding author.
E-mail address: sakurada@nagoya-u.jp (K. Sakurada).

http://dx.doi.org/10.1016/j.cviu.2017.01.012
1077-3142/© 2017 Published by Elsevier Inc.

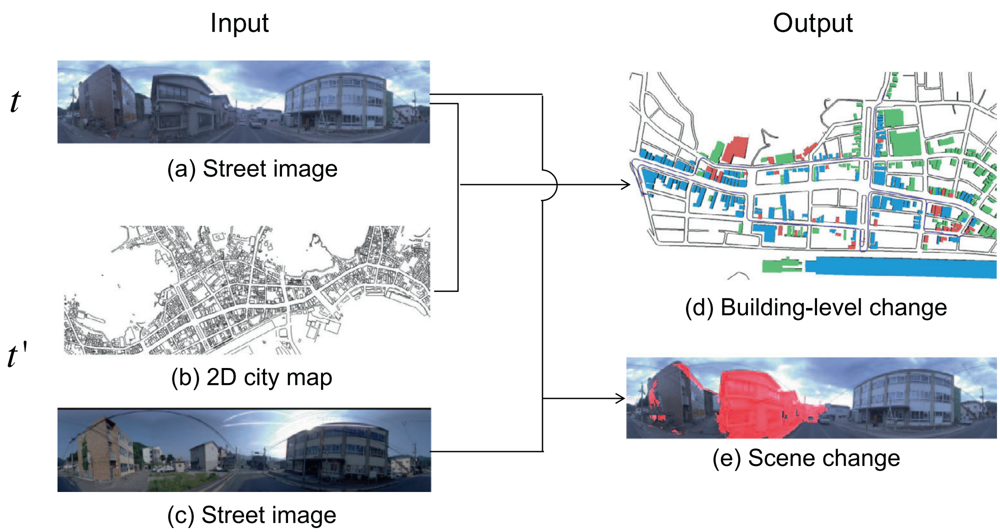

Fig. 1. Inputs and outputs in our strategy. When (a) the street image for time t and (b) a 2D city map for time t' are available, we can estimate (d) building level changes. When street images for both time (a) t and (c) t' are available, we can estimate (e) the scene change.

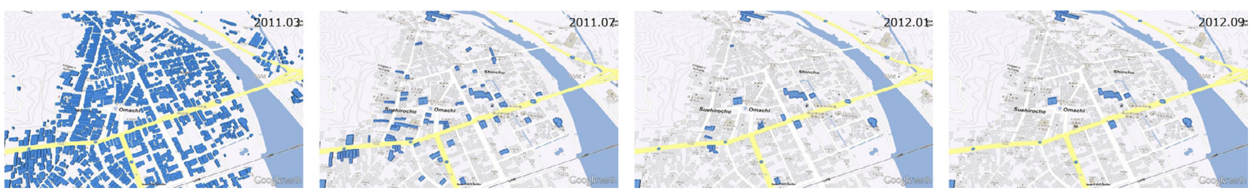

Fig. 2. Visualization of temporal changes in a city. The left image shows buildings before the disaster; the other images show the results obtained by our method using several image sequences captured at different times. These images visualize how buildings are initially removed by the disaster and then remaining buildings are gradually demolished in recovery operation.

subsequently caused by the tsunami, as obtained by this method. When street images for both time t and t' are available, we can estimate the scene change. The proper use of the two methods enables temporal city modeling for a variety of situations.

More specifically, the method for estimating changes at the building-level reconstructs the point cloud of city structures using the Structure from Motion (SfM) technique, and matches the point cloud with the 3D building structures recovered from the map. There are several difficulties faced when using this approach, including the inaccuracy in the recovered building structures, large differences in observation and thus in the point cloud size of individual buildings, and the mutual dependency of the existence of buildings due to potential occlusions. These problems are addressed by developing a model that represents the generation of a point cloud that sequentially utilizes SfM, a building wall observation model, and a greedy iterative approach to address the mutual dependency.

Most previous approaches for scene change detection require a 3D scene model and/or pixel-level registration between different temporal images. These approaches (as well as the first method) are computationally expensive for estimating city-scale changes. The second method does not have these problems because it uses the features of a convolutional neural network (CNN) in combination with superpixel segmentation. Comparison of CNN features produces a low-resolution map of scene changes that is robust to illumination changes and viewpoint differences. Superpixel segmentation of the scene images is integrated with this low-resolution map to estimate precise segmentation boundaries for each change. Hence, it can quickly detect scene changes from an image pair aligned using GPS data, making it possible to process an entire city with a single workstation.

The specific targets of our study are cities that were seriously damaged by the tsunami resulting from the Great East Japan Earthquake, which occurred on March 11, 2011. (Most of the buildings in these cities are at least five meters high, an assumption included in our building-level change detection method.) This event can be considered an unprecedented disaster in human history, as the tsunami forced many modern cities to change their structures in such a short time frame. Hence, visualizing the damage they incurred and the recovery process has many applications. Looking beyond this study, it should be noted that there is considerable opportunity to employ computer vision techniques in these types of disaster-related applications. Moreover, the development of computer vision techniques would also benefit from targeting the tsunami-damaged areas of Japan, because of the variety of damage and recovery that has occurred. These events provide several ideal test cases for enabling the development of new techniques. Both methods are applied to the tsunami-damaged areas and our experimental results demonstrate their effectiveness.

The paper is organized as follows. In Section 2, we summarize related work. Section 3 presents the building-level change detection method, which uses street images and a 2D map. Section 4 describes the scene change detection method, utilizing an image pair. Section 5 concludes this study.

2. Related work

Many studies have examined the problem of detecting building-level or scene changes by using either or both images and maps (Agarwal et al., 2009; Crandall et al., 2011; Crispell et al., 2012; Ibrahim Eden, 2008; Košecka, 2012; Pollard and Mundy, 2007; Pollefeys et al., 2008; Snavely et al., 2007; Zhang et al., 2010, 2007). These studies can be classified into several categories depending