

Massive City-scale Surface Condition Analysis using Ground and Aerial Imagery

Ken Sakurada¹ Takayuki Okatani¹ Kris M. Kitani²

¹Tohoku University, Japan

²Carnegie Mellon University, USA

Abstract. Automated visual analysis is an effective method for understanding changes in natural phenomena over massive city-scale landscapes. However, the view-point spectrum across which image data can be acquired is extremely wide, ranging from macro-level overhead (aerial) images spanning several kilometers to micro-level front-parallel (street-view) images that might only span a few meters. This work presents a unified framework for robustly integrating image data taken at vastly different viewpoints to generate large-scale estimates of land surface conditions. To validate our approach we attempt to estimate the amount of post-Tsunami damage over the entire city of Kamaishi, Japan (over 4 million square-meters). Our results show that our approach can efficiently integrate both micro and macro-level images, along with other forms of meta-data, to efficiently estimate city-scale phenomena. We evaluate our approach on two modes of land condition analysis, namely, city-scale debris and greenery estimation, to show the ability of our method to generalize to a diverse set of estimation tasks.

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 make it possible to accurately understand 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 meter) and visibility may drastically change depending on the weather. For example, clouds may obscure the visibility of the land surface (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, captured from the ground-level can obtain higher resolution images of vertical structures and have better access to



Fig. 1. Aerial images affected by weather condition (Left: March 11, 2011, Right: March 31, 2011). The land surface might be covered by clouds and illumination conditions change drastically in aerial image.

information about covered areas. They are also less affected by weather conditions. In the same token however, street view images are constrained to the ground plane and a single image has limited physical range. 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, while leveraging the wide coverage capabilities of macro-level images and detailed resolution of micro-level images. The strategy proposed in the work uses macro-level imaging to learn land condition correspondences between land regions that share similar visual characteristics (e.g, mountains, streets, buildings, rivers), while 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 region correspondences and surface conditions, our proposed method generates detailed estimates of land surface conditions over the entire city.

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 the 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 generalize to different types of large-scale phenomena, we ground our proposed approach in a real-world application of post-Tsunami city-scale damage estimation. In regions affected by such disasters, it is extremely hard to efficiently assess the large-scale impact of a natural disaster. Technologies that enable fast and efficient city-scale estimates of damage can be extremely helpful for expediting aid to seriously damages areas. The approach describe in this paper can also be used for long-term analysis by monitoring and tracking recovery efforts.



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

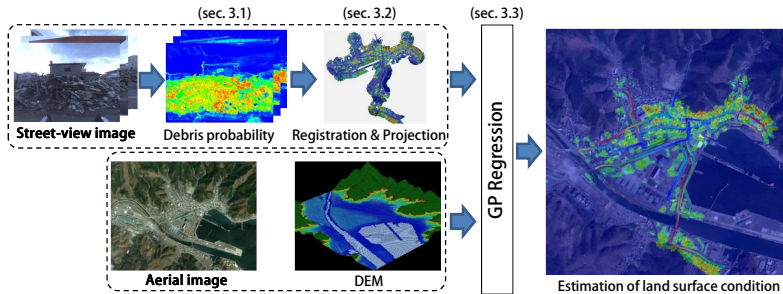


Fig. 3. Data flow diagram of city-scale estimation of land surface condition. Our approach efficiently integrates both micro (street-view) and macro-level (aerial) images along with other forms of meta-data to estimate city-scale land surface condition.

2 Related work

There has been significant advances in the state-of-the-art techniques for quantitative geometric interpretations of large-scale city scenes. Methods for city-scale 3D reconstruction have been proposed using thousands of images gathered from Internet images [1, 2]. Similar techniques have been proposed for images captured by a vehicle-mounted camera [3, 4] or aerial images [5–7]. Street-view images have also been combined with aerial images for the purpose of improving 3D reconstruction, where 3D point clouds have been projected to the ground plane and aligned with edges of buildings detected from aerial images [8] or building maps [9]. There has also been work using aerial and street view images taken several months or decades apart [10–14] to understand temporal changes of a scene. The focus of these previous approaches are on a quantitative geometric interpretation

of the scene where local visual features are matched directly to estimate camera pose using epipolar geometry [15]. In this work we aim to push beyond a purely geometric understanding of the scene towards a more qualitative understanding of city conditions. For instance, we not only interested in the 3D geometry of a building but would also like to know the condition of the building or the condition of the ground surrounding a building.

There also has been work focused on the qualitative estimation of land condition over large-scale environments. In the field of remote sensing, coarse land surface conditions have been estimated using aerial color images, aerial infrared light and aerial microwave sensing [16–21]. Color aerial images have been applied to land condition estimation for vegetation monitoring [22–24], land cover mapping, and flood risk and damage assessment [25, 26]. For example, forest maps [27–29] are an important source of information for monitoring and reducing deforestation, allowing environmental scientists to know how forested areas increase or decrease in over the entire earth.

Apart from aerial imaging using color cameras, many other modes of sensing have been proposed for estimating coarse large-scale land surface conditions. Digital elevation map (DEM) [27], Spectroradiometer (MODIS), high resolution radiometer (AVHRR) and Synthetic Aperture Radar (SAR) have been proposed to improve accuracy of estimating large-scale land surface condition. However the resolution of satellite-mounted MODIS and AVHRR only measure surface conditions over a very rough resolution – typically over a cell size of a several hundred meters. As such, these works do not utilize street-level sensing which are too detailed for their estimation task. However, in this work we are interested in a much high resolution estimate of land conditions on a cell size closer to 20 meters wide.

Our proposed work fills a void between detailed geometric reconstructions of city-scale structures and coarse qualitative estimation of land conditions. We use known techniques to provide an accurate geometric model of the city and use state-of-the-art object recognition results carefully registered to the scene geometry to understand the qualitative conditions of the entire city.

3 Large-scale estimation of land surface condition

Our framework integrates aerial and street-view images to estimate land surface conditions. In this section, we explain the details of the proposed method contextualized for post-Tsunami debris detection. Although the following explanation takes debris as an example, the method is generally applicable to other types of land surface conditions. The proposed method consists of the following three steps;

- (i) Debris detection on perspective street-view image. (sec.3.1)
- (ii) Projection of debris probabilities on street-view images to the ground using building contours. (sec.3.2)
- (iii) Estimation of debris over an entire city by integrating the projection result with all other data (e.g. aerial image, DEM) using a Gaussian process.(sec.3.3)

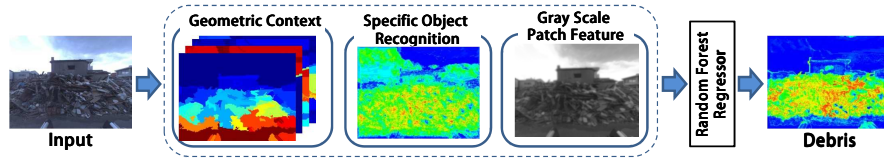


Fig. 4. Data flow diagram of debris detection. As features of debris, the probabilities of geometric context, specific object recognition and patch features are employed.

In the first step, the probability map of debris is calculated for each street-view image. Then, using the camera parameters for the street-view image, the probability map is projected onto the ground plane registered to a corresponding part of the aerial image. This projection method takes the existence of building walls into consideration. Finally in order to complement the estimation results obtained from street-view images, the projected probability map is integrated with the information obtained from aerial images and DEM using Gaussian process regression model.

3.1 Debris detection

We developed a method to calculate the probability map of debris (Fig. 4). The debris model is learned from a hand-labeled training image. The debris in the images are irregular, complicated in shape and appearance. Therefore, we exploit Geometric Context [30] as geometric feature and pixel-wise object probability [31] as an appearance feature. Geometric Context estimates the probabilities that a super-pixel belongs to seven classes. We chose four of the seven classes, "ground plane", "sky", "porous non-planar" and "solid non-planar", and used the probabilities of them as debris features. The pixel-wise object probability p_{object} is calculated using [31], Lab, HOG[32], BRIEF[33] and ORB[34]. The feature vector of debris is as follows.

$$\mathbf{x} = (p_{\text{ground}}, p_{\text{sky}}, p_{\text{porous}}, p_{\text{solid}}, p_{\text{object}}, m_{\text{patch}}, v_{\text{patch}})^T, \quad (1)$$

where p_{ground} , p_{sky} , p_{porous} and p_{solid} are the probabilities of "ground plane", "sky", "porous non-planar" and "solid non-planar", respectively. In addition to these probabilities, mean m_{patch} and variance v_{patch} of grayscale patch (5×5) are added to the features.

We evaluated the accuracy of our debris detector. We made two datasets for the evaluation. Figure 5 shows an example of the datasets and detection results. Each dataset consists of fifty images of debris. The images in two data set were taken in different date and time. We compared random forest [35], logistic regression [36] and support vector machine [37]. Figure 6 shows the F_1 -scores of the debris detections. We chose the random forest as our debris detector for all experiments because the score of random forest regressor is the best.

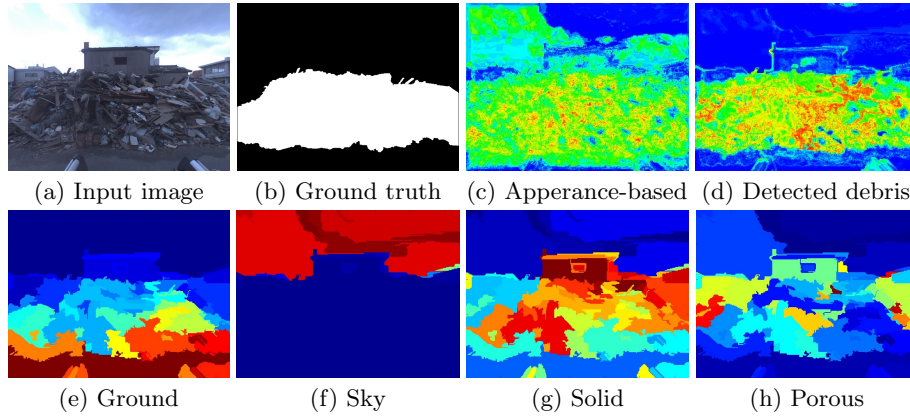


Fig. 5. Inputs and outputs of debris detection. First rows: (a) input image. (b) hand-labeled ground truth of debris. (c) result of specific object recognition. (d) final result of debris detection. Second rows: probability of geometric context (e) ground plane, (f) sky, (g) solid non-planar, (h) porous non-planar. Color denotes probability of each class, with blue corresponding to 0 and red to 1.

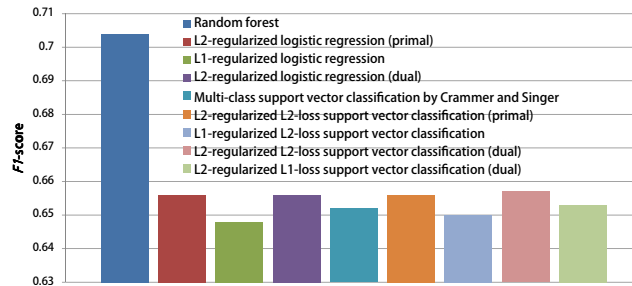


Fig. 6. F_1 -score of debris detection.

3.2 Projection of debris probabilities onto the ground

The debris probability explained in the previous section is the probability map on the street-view image. In order to integrate this probability map with the aerial image, the debris probability is projected onto the ground plane. Figure 7 shows the data flow diagram of projection of street-view image to the coordinate of the aerial image. The projection requires camera parameters of each street-view image. First, we performed Structure from Motion (SfM) to acquire the camera trajectories. We employ a standard SfM method [15, 38, 39] with extensions to deal with omni-directional images [4]. The estimated camera trajectories are fitted to the GPS trajectory by similarity transformations in a least squares sense.

Dividing the ground plane into a grid, we project the debris probability to the grid using projection matrix of each image. In this projection, we use the 3D models of the buildings that are generated from a 2D map of the city (Sec. 4.2).

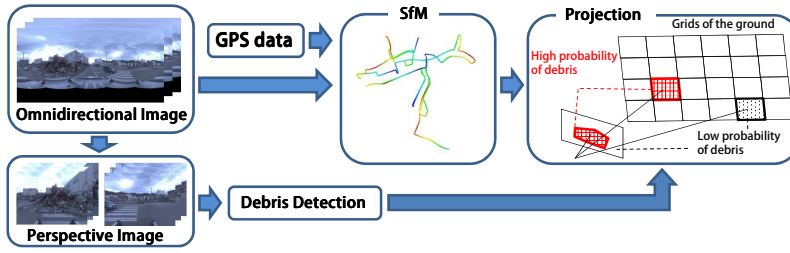


Fig. 7. Data flow diagram of the projection onto the ground plane. SfM is performed using omnidirectional street-view images. The street-view camera poses are registered to a common coordinate with aerial images and other forms of meta-data using the GPS data. After debris detection, the debris probabilities are projected to the ground plane.

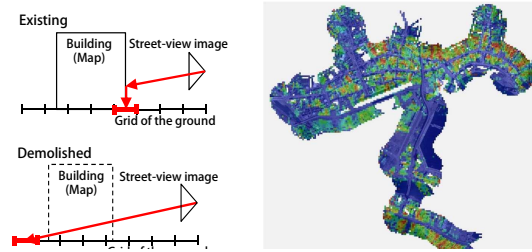


Fig. 8. Projection of probabilities on street-view images to the grids of the ground plane using building information. Left: The probabilities on a street-view image are projected to a building wall if the building is on the projection path, otherwise it is projected to the ground directly. Right: Example of projection results (top-view). The area unobserved from street-view images is shown in white.

To be specific, the debris probability is projected to a building wall if the wall is on the projection path, and otherwise it is directly projected to the ground, as shown in Fig. 8.

3.3 Integration using Gaussian Process regression

The projected debris probability map obtained up to now has no information for some areas because of occlusions or the lack of street-level images, as shown in Fig. 8. Estimating debris probability map from only an aerial image is difficult due to its low-resolution, occlusion or weather conditions. To mutually complement the street-view images and the aerial image, we used a Gaussian process regression model[40]. The main idea here is that similar geographical location tend to have similar debris probability. In the case of Tsunami-disaster, Tsunami continuously spreads from seashore to hill side, which means the damage caused by Tsunami has strong correlation with the location, especially with the elevation.

As described in the previous section, the debris probability of each grid $p_{s,i}$ ($i = 1, \dots, n$) is estimated from the street-view images. For each grid, its feature vector \mathbf{x}_i is defined as follows.

$$\mathbf{x}_i = (x_i, y_i, z_i, p_{a,i})^T \quad (2)$$

where (x_i, y_i) is a center position of each grid, z_i is a elevation of each grid calculated from DEM and $p_{a,i}$ is debris probability of each grid estimated from aerial image using pixel-wise object recognition[31]. The column vector \mathbf{x}_i for all n grid are aggregated in the $4 \times n$ training inputs matrix X , and the training outputs $p_{s,i}$ are collected in the vector \mathbf{y} .

\mathbf{x}_i contains $p_{a,i}$ as the visual feature of the i th grid. Although $p_{a,i}$ is a scalar, due to the pixel-wise object recognition[31], it summarizes the visual information of the i th grid. Compared to using general visual feature descriptors, such as SIFT[38], directly in the feature vector \mathbf{x}_i , $p_{a,i}$ saves computational resources required in the following calculation of covariance function. The covariance function for Gaussian process regression in the proposed method is as follows.

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|W(\mathbf{x}_p - \mathbf{x}_q)|^2\right) + \sigma_n^2 \delta_{pq} \quad (3)$$

where W is the weight diagonal matrix, l is the length-scale, σ_f^2 is the signal variance, σ_n^2 is the noise variance and δ_{pq} is a Kronecker delta which is one if $p = q$ and zero otherwise. Test input \mathbf{x}_* is each grid feature vector and test output is debris probability of each grid f_* .

The key insight to note here is that the output of the aerial image regressor $p_{a,i}$ enforces a correlation between parts of the scene that look similar. If two parts of the scene belong to an open field, the per-pixel response of the aerial object detection regressor will produce a similar response. The DEM also works in a similar manner to draw correlations between regions with similar elevation. The location feature enforces local smoothness over the final estimate of debris over the city. When the feature vectors \mathbf{x}_i are used to compute the covariance function, regions that are similar in appearance and elevation will be constrained to have similar target values (debris estimates generated by high resolution debris regressor computed on the street images). In this way, the Gaussian process regression model is able to propagate local estimates of debris to the entire map. This regression mechanism is what allows our model to effectively estimate debris over the entire city from only a sparse set of street view debris estimation results.

4 Experimental results

In order to evaluate the effectiveness of our proposed approach for estimating large-scale land conditions, we perform two experiments. Our first experiment is a comprehensive ablative analysis to examine the benefit of integrating micro and macro-level imagery for city-scale land condition estimation. In addition to color imaging, we also evaluate the contributions of two other modes of data,

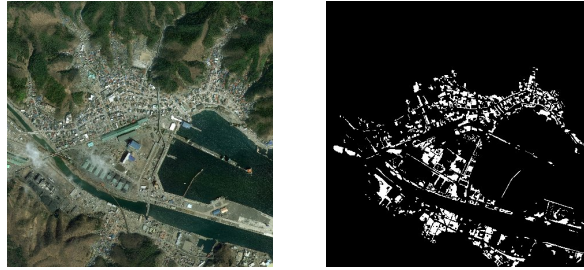


Fig. 9. Estimation target area in Kamaishi on March 31st, 2011 (left) and its hand-labeled ground truth of debris area (right). White area shows debris area.

namely, a digital elevation map (DEM) and building occupancy maps (BOM). In our second experiment, we focus on estimating the amount of greenery and vegetation across the entire city of Kamaishi. We use the exact same approach as the debris estimation described in this paper and apply it to greenery estimation. Our results show that our approach is not limited to post-disaster analysis but can easily be applied to other modes of land condition analysis.

We created the ground truth labels used for the following evaluation by many hours of manual labeling of regions on the aerial images. Ground truth data of debris and greenery were generated by visual inspection by comparing the aerial image against the street-view images available on Google Earth. Many hours of ground truth labeling confirms that the manual inspection of large-scale land conditions is not a practical solution for real-world applications.

4.1 Our data

Our experiment includes two image-based input modalities and two sources of city-scale meta-data, which are described below.

Street images. We have been creating image archives of urban and residential areas damaged by Great East Japan Earthquake in 2011. The target area is 500 kilometers long along the northern-east coastal line in Japan. The images were captured every three to four months by a vehicle having an omni-directional camera (Ladybug 3 and 5 of Point Grey Research Inc.) on its roof. The image data accumulated so far amount to about 20 terabytes. The target of this experiment is the entire city of Kamaishi, Japan (over 4 million square-meters). For the experiments, we chose the two image sequences captured on April 26th, 2011 (one month after the Tsunami) and August 17th, 2013 (two years and five months after the Tsunami). The debris can often be seen in the earlier images, while they tend to disappear in the later images as the recovery operation proceeds.

The street images are used for appearance-based recognition of ‘stuff’ [41] described in Section 3.1. The results of pixel-wise regression are then projected onto the ground plane as an input feature for our city-scale GP regressor.

Aerial images We downloaded aerial images from Google Earth for March 31st, 2011 and May 13th, 2012. We chose these dates to match up the timestamp of the street images.

We used the aerial images for appearance-based recognition of ‘stuff’ categories using the same method describe in Section 3.1 but applied to the entire aerial image as a comparative baseline. We used the aerial images of May 13th, 2012 as the labeled training data and test on the March 31st, 2011 aerial image. Figure 9 shows an example of the hand-labeled ground truth of the debris area on the aerial images.

Digital Elevation Map (DEM). We obtained the DEM information freely available from the Geospatial Information Authority, under the Ministry of Land, Infrastructure, Transportation and Tourism in Japan. The mesh resolution of the DEM is 5×5 square-meters and contains the elevation level for each grid location. The elevation is used directly as a feature for the city-scale GP regression.

Building Occupancy Map (BOM) The BOM provides building contours. We obtained the data from Zenrin Company. The building contour data used for this experiment was made before the earthquake. We used the BOM to prevent ‘stuff’ from being projected onto the ground over building location.

4.2 Ablative Analysis

We examine the effects of each input data type on the overall performance of our proposed approach. Figure 10 shows the estimation results of the debris amounts in the entire city on April 26th, 2011 and August 17th, 2013, respectively. The lines on the aerial images are the camera trajectories. Figure 11 shows the performance of our debris detection by PR-plot and $F1$ -score using different combination of input data. The results indicate that using aerial images alone yields low performance because the appearance of land conditions can change significantly over time due to changes in imaging conditions. When compared to the independent use of aerial images, our results indicate that street images are more accurate for estimating city-scale debris. Furthermore, when both aerial and street images are combined we obtain better performance as the aerial information helps the city-scale GP regression to generalize to across similar looking city regions.

Additionally, we evaluated the effects of each input data type and different number of street-view images in three different streets. Figures 12 (i)-(iii) show input aerial images of target areas. Figure 13 shows precision-recall curve (Left) and $F1$ -scores (Right, recall=0.5) of the debris area detection. We examined the detection performance for a specific area. The effect of different input type is different depending on the area-condition. For example, aerial image could cause errors in occluded areas we mentioned in the introduction, and DEM information could cause errors because its elevation includes height of building. However, street-view images basically provide detailed and accurate information of land-surface condition.

We also tested the effect of street coverage. Figure 14 shows the $F1$ -score of different number of street-view images. In this experiment, we randomly sam-

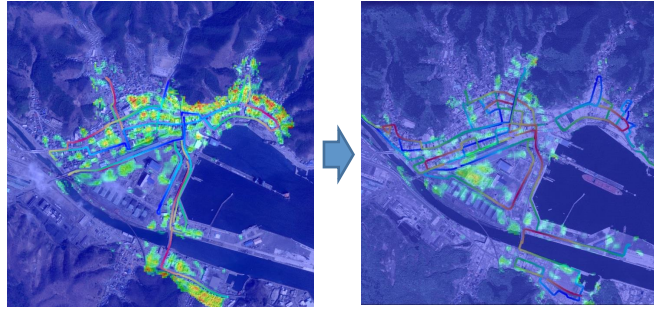


Fig. 10. City-scale **Debris** Probability in Kamaishi before and after the recovery operation (Left: April 26th, 2011, Right: August 17th, 2013). In the earlier images, there are much debris in the entire city, however, most of them have been cleaned up in later images. The city-scale temporal change is estimated and visualized accurately by our approach. Color denotes probability of debris, with blue corresponding to 0 and red to 1.

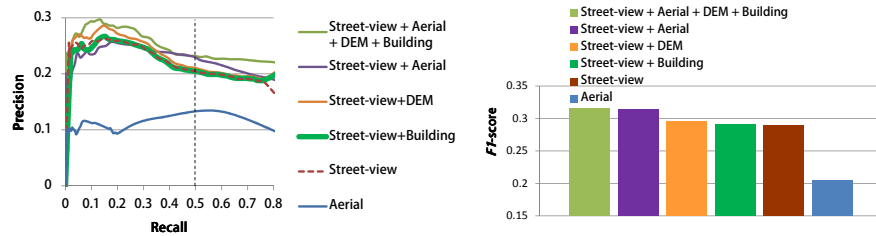


Fig. 11. Precision-recall curve of the debris area detection whose ground truth is Fig. 9. These figures show that the integration of street-view image with aerial image is efficient to estimate city-scale land surface condition.

pled street-view images. The accuracy improves as we add more images, but it quickly saturates. This indicates our algorithm needs only sparse street-view images. The sparse sampling requirement of our algorithm is beneficial in many other applications, for example, large scale citizen science or journalism in which images captured at the scene are sent to cloud computers to analyze city scale condition.

4.3 Extensions to City-Scale Vegetation Estimation

We applied our method to vegetation detection, to show how our approach can generalize to other modes of land condition estimation. Figure 15 shows an example of vegetation estimation in street-level images. The green vegetation detected in the street-view images is estimated using the same pixel-wise object recognition method [31].

Figure 16 shows the results of vegetation estimation for the entire city. By observing the vegetation heat map for the entire city, it is clear that most of

the vegetation has been washed away by the Tsunami. There is also a sharp contrast between the wide spread distribution of debris and the lack of vegetation in the time period directly after the Tsunami. By 2013 however, we can see a large increase in the number of regions covered by vegetation. Our successful vegetation detection indicates that our proposed method can indeed generalize to different types of targeted estimation of city-scale land conditions.

5 Conclusion

We presented a unified framework for robustly integrating image data taken at vastly different viewpoints to generate large-scale estimates of land surface conditions. The proposed strategy uses macro-level imaging to learn land condition correspondences between land regions that share similar visual characteristics, while micro-level images are used to acquire high resolution statistics of land conditions. For the validation of our approach, we conducted experiments to estimate the amount of post-Tsunami damage over the entire city of Kamaishi, Japan. The experimental results show that our approach can effectively integrate both macro (aerial) and micro-level (street view) images, along with other forms of meta-data, to estimate city-scale phenomena.

Furthermore, we showed that our detection method can be successfully applied to vegetation estimation. The results indicate our method can generalize well to many kinds of applications to estimate city-scale phenomena by replacing the detector target, for example, human flow, real-estate and dirt quality. These types of image data are available from many kinds of data sources, such as camera equipped mobile devices, surveillance cameras and car-mounted video recorders, or aerial-vehicle-mounted cameras. Our approach provides an effective and robust method for integrating different kinds of data to estimate city-scale phenomena.

For future work, we plan to improve the estimation accuracy of our approach. Our method has relatively low absolute precision because (i) the grid size is too large due to limitation of computational resources and (ii) the estimated camera poses have errors due to GPS errors. We believe that we can solve the first problem using large-scale Gaussian process [40]. The GPS issue can be addressed with [8, 9, 42] while taking temporal changes into account. Furthermore, in the case of extreme calamities, methods will be developed to take into consideration the complete disappearance of the buildings due to disasters.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Numbers 25135701, 25280054.



Fig. 12. Target streets of the evaluation of each input data type and different number of street-view images in our proposed approach.

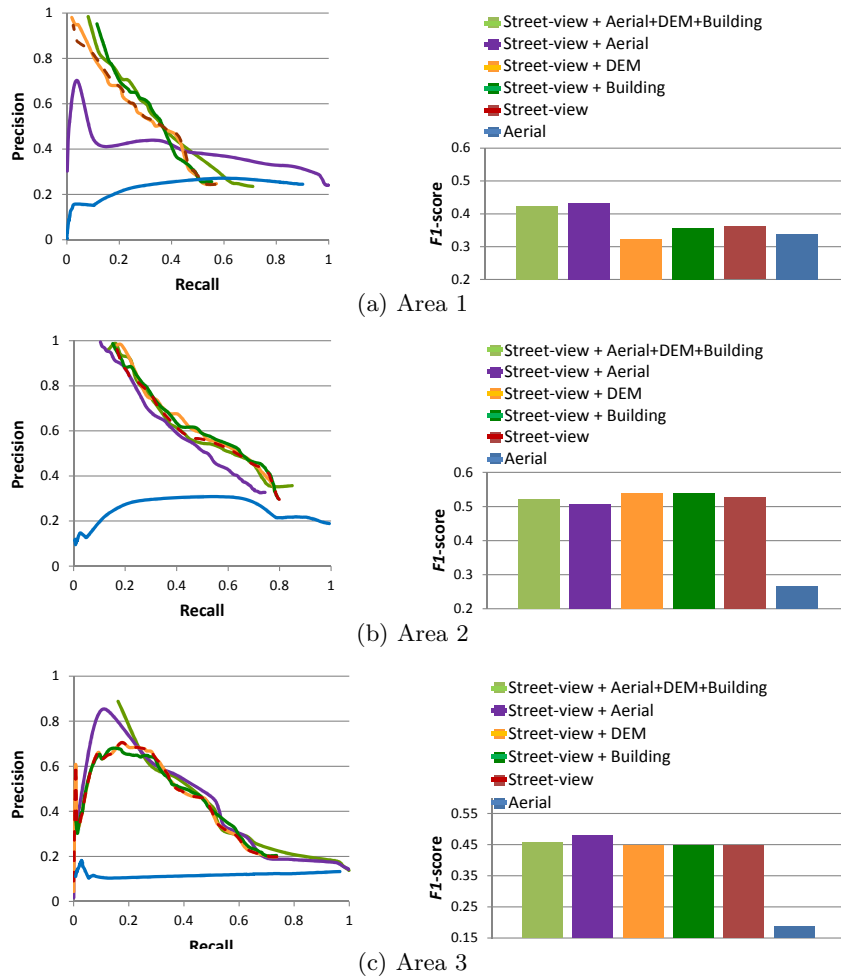


Fig. 13. Precision-recall curve (Left) and $F1$ -scores (Right, recall=0.5) of the debris-area detection. These plots show that the integration of street-view image with aerial image is effective to estimate the condition of land surface.

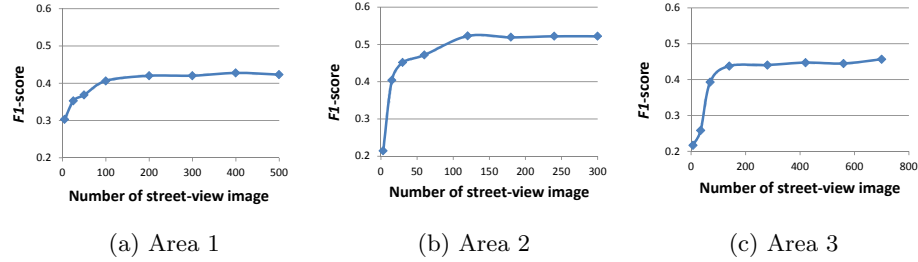


Fig. 14. *F1*-scores of different number of street-view images. Our algorithm does not require large number of images.

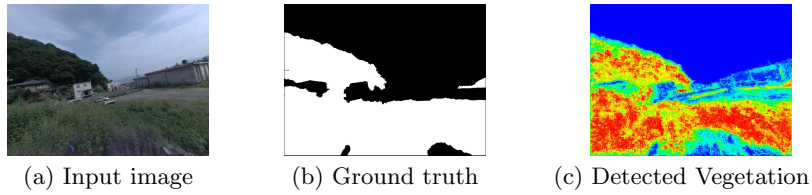


Fig. 15. Green vegetation detection. (a) input image. (b) hand-labeled ground truth of green vegetation. (c) probability of green vegetation. Color denotes probability of green vegetation, with blue corresponding to 0 and red to 1.

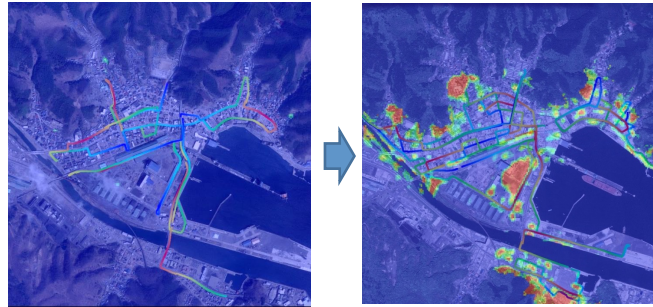


Fig. 16. City-Scale **Vegetation** Probability in Kamaishi before and after the recovery operation (Left: April 26, 2011, Right: August 17, 2013). In contradiction to the debris (Fig. 10), there was no green vegetation due to Tsunami-damage in April, 2011, however, the vegetation in entire city has grown and recovered until August, 2013. Our approach can estimate and visualize such changes in natural phenomena over massive city-scale landscapes. Color denotes probability of green vegetation, with blue corresponding to 0 and red to 1

References

1. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the World from Internet Photo Collections. *IJCV* **80** (2007) 189–210
2. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a day. In: *ICCV*. (2009) 72–79
3. Pollefeys, M., Nistér, D., Frahm, J.M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., Towles, H.: Detailed Real-Time Urban 3D Reconstruction from Video. *IJCV* **78** (2008) 143–167
4. Torii, A., Havlena, M., Pajdla, T.: From Google Street View to 3D City Models. In: *ICCV Workshops*. (2009) 2188–2195
5. Lin, C., Nevatia, R.: Building Detection and Description from a Single Intensity Image. *Computer Vision and Image Understanding* **72** (1998) 101–121
6. Suveg, I., Vosselman, G.: Reconstruction of 3D building models from aerial images and maps. *ISPRS Journal of Photogrammetry and Remote Sensing* **58** (2004) 202–224
7. Zebedin, L., Klaus, A., Gruber-Geymayer, B., Karner, K.: Towards 3D map generation from digital aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing* **60** (2006) 413–427
8. Kaminsky, R., Snavely, N., Seitz, S., Szeliski, R.: Alignment of 3D Point Clouds to Overhead Images. *CVPR Workshops* (2009) 63–70
9. Strecha, C., Pylvanainen, T., Fua, P.: Dynamic and Scalable Large Scale Image Reconstruction. In: *CVPR*. (2010) 406–413
10. Huertas, A., Nevatia, R.: Detecting Changes in Aerial Views of Man-Made Structures. In: *ICCV*. (1998) 73–80
11. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image Change Detection Algorithms: A Systematic Survey. *Transactions on Image Processing* **14** (2005) 294–307
12. Pollard, T., Mundy, J.L.: Change Detection in a 3-d World. In: *CVPR*. (2007) 1–6
13. Schindler, G., Dellaert, F.: Probabilistic temporal inference on reconstructed 3D scenes. In: *CVPR*. (2010) 1410–1417
14. Taneja, A., Ballan, L., Pollefeys, M.: City-Scale Change Detection in Cadastral 3D Models Using Images. In: *CVPR*. (2013) 113–120
15. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision* Second Edition. Cambridge University Press (2004)
16. Li, F., Jacksona, T.J., Kustasa, W.P., Sch muggea, T.J., Frenchb, A.N., Cosha, M.H., Bindlish, R.: Deriving land surface temperature from Landsat 5 and 7 during SMEX02/SMACEX. *Remote Sensing of Environment* **92** (2004) 521–534
17. Weng, Q., Lu, D., Schubring, J.: Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment* **89** (2004) 467–483
18. Schowengerdt, R.A.: *Remote Sensing: Models and Methods for Image Processing*. (2006)
19. Martinez, J., Letoan, T.: Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. *Remote Sensing of Environment* **108** (2007) 209–223
20. Weng, Q.: *Remote Sensing of Impervious Surfaces*. CRC Press (2010)
21. Lu, D., Hetrick, S., Moran, E.: Impervious surface mapping with quickbird imagery. *International journal of remote sensing* **32** (2011) 2519–2533

22. Hall, A., Louis, J., Lamb, D.: Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images. *Computers & Geosciences* **29** (2003) 813–822
23. Berni, J.A.J., Member, S., Zarco-tejada, P.J., Suárez, L., Fereres, E.: Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From an Unmanned Aerial Vehicle. *IEEE Transactions on Geoscience and Remote Sensing* **47** (2009) 722–738
24. Delenne, C., Durrieu, S., Rabatel, G., Deshayes, M.: From pixel to vine parcel: A complete methodology for vineyard delineation and characterization using remote-sensing data. *Computers and Electronics in Agriculture* **70** (2010) 78–83
25. van der Sande, C., de Jong, S., a.P.J. de Roo: A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment. *International Journal of Applied Earth Observation and Geoinformation* **4** (2003) 217–229
26. Herold, M., Liu, X., Clarke, K.C.: Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering & Remote Sensing* (2003) 991–1001
27. Gong, P., Pu, R., Chen, J.: Mapping Ecological Land Systems and Classification Uncertainties from Digital Elevation and Forest-Cover Data Using Neural Networks. *Photogrammetric Engineering & Remote Sensing* **62** (1996) 1249–1260
28. Solberg, A.H.S.: Contextual Data Fusion Applied to Forest Map Revision. *IEEE Transactions on Geoscience and Remote Sensing* **37** (1999) 1234–1243
29. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.: High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342** (2013) 850–853
30. Derek Hoiem, Alexei A. Efros, Martial Hebert: Geometric context from a single image. In: *ICCV*. (2005) 654–661
31. Li, Cheng, Kitani, Kris M.: Pixel-level Hand Detection in Ego-Centric Videos. In: *CVPR*. (2013) 3570–3577
32. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *CVPR*. (2005) 886–893
33. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF : Binary Robust Independent Elementary Features . In: *ECCV*. (2010)
34. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB : an efficient alternative to SIFT or SURF. In: *ICCV*. (2011) 2564–2571
35. BREIMAN, L.: Random Forests. *Machine Learning* **45** (2001) 5–32
36. Mitchell, T.M.: *Machine Learning*. (1997)
37. Vladimir Naumovich Vapnik: Estimation of Dependences Based on Empirical Data. (2006)
38. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *IJCV* **60** (2004) 91–110
39. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle Adjustment - Modern Synthesis. In: *ICCV*. (1999) 298–372
40. Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press (2005)
41. Forsyth, D.A., Malik, J., et al.: *Finding pictures of objects in large collections of images*. Springer (1996)
42. Cham, T.J., Ciptadi, A., Tan, W.C., Pham, M.T., Chia, L.T.: Estimating Camera Pose from a Single Urban Ground-View Omnidirectional Image and a 2D Building Outline Map. (In: *CVPR*) 366–373