Data-Driven Scene Parsing Method for Construction Site Monitoring

Hongjo Kim, Kinam Kim, and Hyoungkwan Kim

School of Civil and Environmental Engineering, Yonsei University;
Seoul, Korea
E-mail: hongjo@yonsei.ac.kr, kinam@yonsei.ac.kr, hyoungkwan@yonsei.ac.kr

Abstract –

This paper presents the applicability of a nonparametric scene parsing model for recognizing all objects in an image. The data-driven model labels all the pixels of query images as their object classes using the labels in similar images of the existing dataset. The model has a flexible number of parameters depending on the size of the training data, it is possible to add or remove new data without re-training the whole model. The capability of the nonparametric model would improve the monitoring performance on the construction site with updating the size of the image dataset over time.

Keywords –

Monitoring; Construction Site; Computer Vision; Image Processing; Scene Parsing; Object Recognition

1 Introduction

A lot of efforts have been made for vision-based monitoring in the construction industry [1-5]. Few monitoring methods exist to recognize numerous types of construction objects [6-9]. The previous studies recognizing multiple construction objects used parametric models requiring cumbersome model training processes. Model training processes involve preparing image training data and tuning parameter for which developers spend considerable time and cost. Developer's efforts are proportional to the number of objects of interest; therefore, recognizing holistic construction site scene is elusive.

To address the issues, a data-driven scene parsing method is proposed to recognize various construction object classes. The proposed method is nonparametric in which modeling prior information of target object's features is unnecessary. Furthermore, if a training data size is changed, only an additional data is processed to update the monitoring system, unlike parametric methods involving reconstruction of the whole system.

Benefits of the proposed construction site monitoring system are (1) capability of understanding the whole image area, (2) ease of updating the system with a changing size of training data, and (3) simplicity of tuning model parameters.

2 Data-driven scene parsing

In an image, each pixel has a brightness value within the range of 0–255. A brightness value or cluster has no intuitive meaning of an object to a person. To get object's information, an object boundary in a pixel grid should be properly segmented and then recognized based on features of the pixel cluster. Scene parsing is a term including these processes.

Among several data-driven scene parsing methods, this paper adopts a nonparametric scene parsing method suggested in [10]. The key idea of the method is transferring labels–object identities–of several images to annotate labels of a query image scene. Specifically, every pixel in a query image is annotated by searching the most similar label of retrieved images relevant to a query image from a database. Retrieving similar images to a query image is conducted by two stage filtering. The first filtering stage uses nearest neighbors to scene contents represented by GIST features (refer to [11] for details). The second stage is the dense scene correspondence filtering, which uses SIFT (scale invariant feature transform) flow matching across the filtered images at the first stage. SIFT flow is a concept similar to optical flow, which extracts a SIFT feature at every pixel of an image to calculate a matching score to a query image [12]. The retrieved images are used to transfer their labels to each pixels of a query image using a probabilistic Markov random field model for energy function formulation. In a Markov random field model, an image pixel grid is represented as an undirected graph having edge and node components. An edge represents a relation between neighbor pixels and a node signifies a pixel. The scene parsing procedure is illustrated in Figure 1.

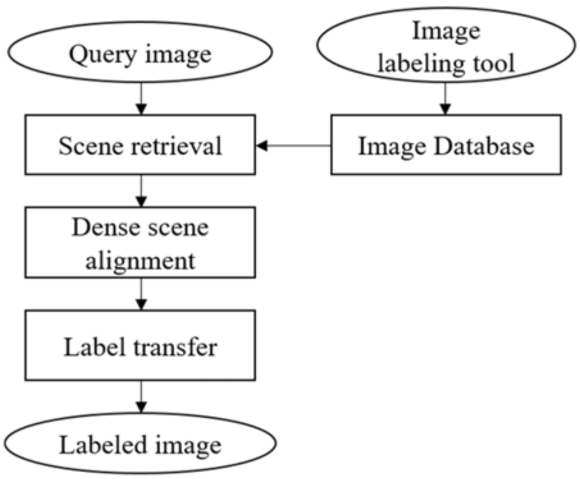

Figure 1. Scene parsing process

3 Experimental settings and results

An experiment was conducted on a desktop computer with Intel i7-4770 CPU, 32 GB RAM in the Windows 10 and 64 bit operating system. This study used the open source code for scene parsing [10] and executed it using a MATLAB programming language. The LabelMe image annotation tool [13] was used to create training image data. The 169 images were used to train the system and 42 images were tested for validation.

The performance of the scene parsing system was measured by the average pixel-wise recognition rate. The evaluation criterion in [10] was selected. The average pixel-wise recognition rate was 81.48% from the 42 test images, with a median value of 84.47% and standard deviation of 13.12%. The average processing time was 7.93 s/query image. Samples of the result are shown in Figure 2.

4 Discussion

The result shows the reasonable performance on recognizing construction site images. However, there has a room for improvement: (1) per-class recognition rates vary in a large degree by object classes. (2) some objects taking a small region in an image were failed to be identified. (3) the long processing time prevents real-time applications. Further studies are required to address these issues.

5 Conclusion

This paper proposed a vision system to identify (scene-parse) the whole image area of construction sites. The system could identify object classes using labels of images that are similar to the query image. The system needed a relatively small number of parameters. The accuracy of the scene parsing was 81.48% based on the average pixel-wise recognition rate. The scene parsing method, in its current version, is capable of recognizing static objects. Further studies, with more training data, would increase the accuracy and the number of construction object classification.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (NRF-2014R1A2A1A11052499 and No. 2011-0030040). The authors would like to thank Yonsei University and Doosan Engineering & Construction for granting access to the construction site of the Baekyang-ro renovation project and the tower construction of College of Engineering Hall 1.

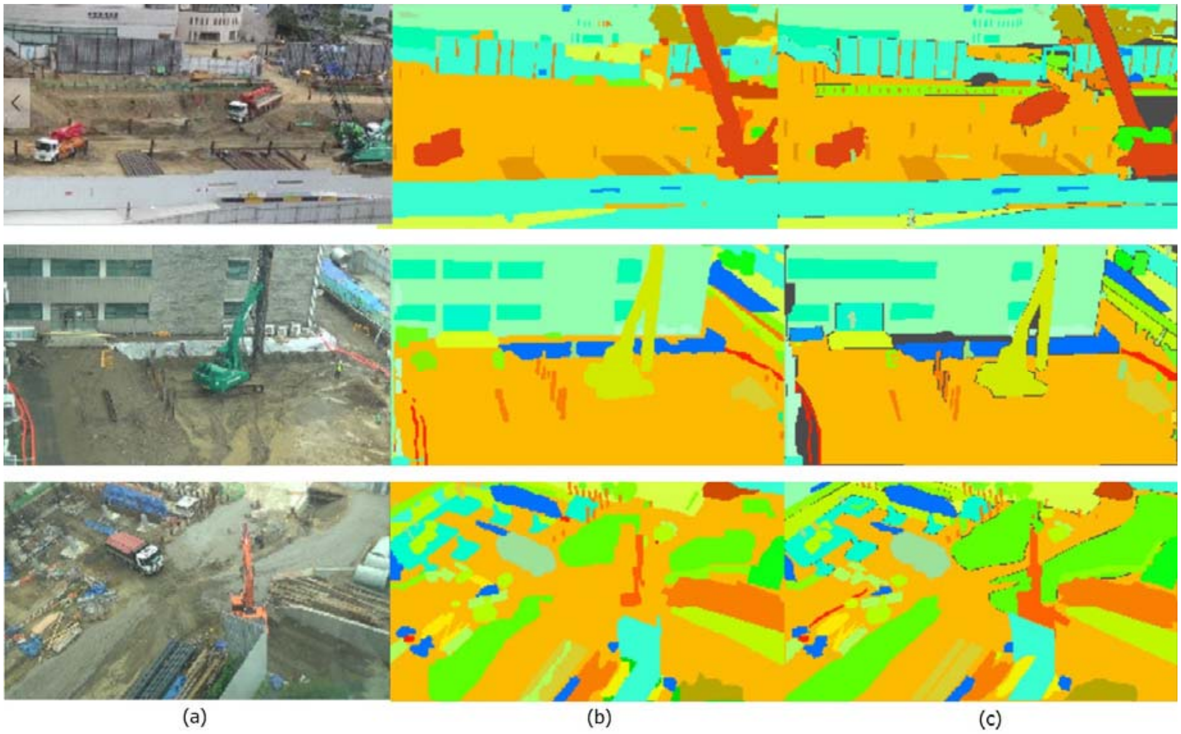

Figure 2. Samples of the scene parsing result. (a) query image, (b) labeled result, and (c) ground-truth label.

References

[1] H. Fathi, F. Dai, M. Lourakis, Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges, Advanced Engineering Informatics 29 (2) (2015) 149-161.

[2] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Advanced Engineering Informatics 29 (2) (2015) 196-210.

[3] H. Son, F. Bosché, C. Kim, As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: A survey, Advanced Engineering Informatics 29 (2) (2015) 172-183.

[4] J. Teizer, Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites, Advanced Engineering Informatics 29 (2) (2015) 225-238.

[5] J. Yang, M.-W. Park, P.A. Vela, M. Golparvar-Fard, Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future, Advanced Engineering Informatics 29 (2) (2015) 211-224.

[6] S. Chi, C.H. Caldas, Automated Object Identification Using Optical Video Cameras on Construction Sites, Computer-Aided Civil and Infrastructure Engineering 26 (5) (2011) 368-380.

[7] A. Dimitrov, M. Golparvar-Fard, Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, Advanced Engineering Informatics 28 (1) (2014) 37-49.

[8] H. Son, C. Kim, N. Hwang, C. Kim, Y. Kang, Classification of major construction materials in construction environments using ensemble classifiers, Advanced Engineering Informatics 28 (1) (2014) 1-10.

[9] H. Tajeen, Z. Zhu, Image dataset development for measuring construction equipment recognition performance, Automation in Construction 48 (2014) 1-10.

[10] C. Liu, J. Yuen, A. Torralba, Nonparametric scene parsing via label transfer, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (12) (2011) 2368-2382.

[11] A. Oliva, A. Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope, International Journal of Computer Vision 42 (3) (2001) 145-175.

[12] C. Liu, J. Yuen, A. Torralba, Sift flow: Dense correspondence across scenes and its applications, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (5) (2011) 978-994.

[13] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: a database and web-based tool for image annotation, International Journal of Computer Vision 77 (1-3) (2008) 157-173.