The 6th International Conference on Construction Engineering and Project Management (ICCEPM 2015)
Oct. 11 (Sun) ~ 14 (Wed) 2015 • Paradise Hotel Busan • Busan, Korea
www.iccepm2015.org

Robust Real-time Object Detection on Construction Sites Using Integral Channel Features

Jinwoo Kim[1] and Seokho Chi[2]

Abstract: On construction sites, it is important to monitor the performance of construction equipment and workers to achieve successful construction project management; especially, vision-based detection methods have advantages for the real-time site data collection for safety and productivity analyses. Although many researchers developed vision-based detection methods with acceptable performance, there are still limitations to be addressed: 1) sensitiveness to the shape and appearance changes of moving objects in difference working postures, and 2) high computation time. To deal with the limitations, this paper proposes a detection algorithm of construction equipment based on Integral Channel Features. For validation, 16,850 frames of video streams were recorded and analyzed. The results showed that the proposed method worked in high performance in terms of accuracy and processing time. In conclusion, the developed method can help to understand useful site information including working pattern, working time and input manpower analyses.

Keywords: Construction Site Information, Site-monitoring, Construction Equipment Detection, Integral Channel Features

I. INTRODUCTION

To support successful construction project management, it is important to monitor and understand the performance and work-in-progress of construction equipment and workers. The on-site information such as locations of entities, work types, and working cycle can be used for the productivity analyses and safety assessment (Gong and Caldas, 2010; Azar and McCabe, 2012; Memarzadeh et al., 2013).

Such site information was typically collected manually on-site by direct observation or survey/interview-based methods. (Navon and Sacks, 2007; Rebolj et al., 2008; Golparva-Fard et al., 2009; Gong and Caldas, 2010; Chi and Caldas, 2011). However, due to their limitations such as time-consuming, error-prone and expensive processes, the demand for automatic site data collection has rapidly grown.

Automatic and real-time site monitoring systems are introduced for efficient, fast, and reliable information collection. One of the potential approaches is a vision-based data collection method: vision-based methods can automatically analyze various types of data including locations of entities, working progress, or sites' environmental conditions. (Chi et al., 2009; Gong and Caldas, 2010; Chi and Caldas, 2011; Park et al., 2011; Azar, 2011; Memarzadeh et al., 2013; Golparvar-Fard et al., 2013).

For instance, Chi and Caldas (2012) presented an image-based safety assessment framework on construction sites using real-time spatial risk identification. Golparvar-Fard et al. (2013) proposed vision-based action recognition algorithms of equipment that can provide fundamental information for productivity analyses.

To produce reliable and usable information for project-related decision making with the vision-based methods, automated and real-time detection of entities in in different working postures in a single camera is an essential prerequisite (Azar and McCabe, 2012; Memarzadeh et al., 2013; Golparvar-Fard et al., 2013). For example, a dump truck that is being loaded needs to be identified as the same dump truck that is traveling to the dump site.

To satisfy such prerequisite for the in-depth analyses, a range of different image processing algorithms has employed to the construction application: background subtraction algorithms (Chi and Caldas, 2011), HOG (Histogram of Oriented Gradient) Cascade and Blob-HOG algorithms (Azar and McCabe, 2012), HOG and HSV algorithms with background subtraction (Park and Brilakis, 2012), and HOG-Color algorithms (Memarzadeh et al., 2013). Although the research showed the acceptable and promising performance, there are still some key challenging issues: 1) sensitiveness to the shape and appearance changes of moving objects in difference working postures, and 2) high computation time. To deal with the limitations, this paper proposes a robust and real-time object detection method based on Integral Channel Features.

II. PRELIMINARY STUDY

A typical object detection method can be divided into two categories: unsupervised and supervised learnings (Richards, 1993). The unsupervised learning is usually used when the correct answers or results (i.e., object lable) cannot be provided. The supervised learning performs with the correct results through the algorithm training processes. The unsupervised learning can answer Object A is Object B in different image frames, but have difficulties to answer Object A is a "backhoe" since the correct label information is not provided. The supervised learning enables the latter prediction cases. The supervised learning method is generally faster and more accurate and able to handle variation of object postures and their background (Conalek, 2011).

[1] Graduate Student, 35-427 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea, jinwoo92@snu.ac.kr (*Corresponding Author)
[2] Ph.D./Assistant Professor, 35-304 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea, shchi@snu.ac.kr

In object detection based on the supervised learning, the performance of a detector is affected by the feature representation (Dollar et al., 2009). The feature is a cue of objects such as edges, colors, shape ratios, and others (Szeliski, 2010; Nixon and Aguado, 2012). To achieve high performance of the detectors, feature selection is one of the most important tasks during the training stage and thus the selected features should represent the characteristics of construction equipment and site conditions accurately and reliably.

On construction sites, it is not an easy task to extract uniform representative features due to a range of scale of workers and equipment, different view-points, working pose-variation, outdoor illumination condition (Chi et al., 2009; Gong and Caldas, 2010; Chi and Caldas, 2011; Park et al., 2011; Azar, 2011; Memarzadeh et al., 2013; Golparvar-Fard et al., 2013). The scale of equipment is not constant at different distances from a camera. The view-point is changed with the object movements. The pose-variation is large while equipment keeps working. To overcome the challenges, the authors employed Integral Channel Features.

Dollar et al. (2009) proposed the Integral Channel Features for pedestrian detection and the method significantly outperformed the previous methods. The Integral Channel Features consists of the multiple image channels and each channel is computed by transforming input images with the linear or non-linear functions. More specifically, the original image, which is usually represented with Red, Green, and Blue (RGB) channels, can be transformed to different channel representations such as gray channel, gradient channel, Gabor filter channel, Gaussian filter channel, and others. The examples are illustrated in figure 1. Multiple channel representation of objects supports to a variety of object characteristics. Thus, the research used the Integral Channel Features for the equipment detection.

CIE-LUV color space, Gradient Histogram and Gradient Magnitude were selected for Integral Channel Features.

A. LUV Color Space

LUV color space which is adopted by the International Commission on Illumination (CIE) is one of the most popular color spaces. The LUV color space consists of three channels: L channel represents the intensity of color, and U and V channels represent red-green and blue-plum colors respectively. Unlike other color spaces such as RGB (Red-Green-Blue), CMYK (Cyan-Magenta-Yellow-Key), and others that varies with a specific device (e.g. camera, scanner, etc.), the LUV color space is device independent (Phil Cruse, 2015). It can also fully separate the grayscale intensity from images.

Thus, the LUV has potential to avoid device-variant and illumination effects on sites.

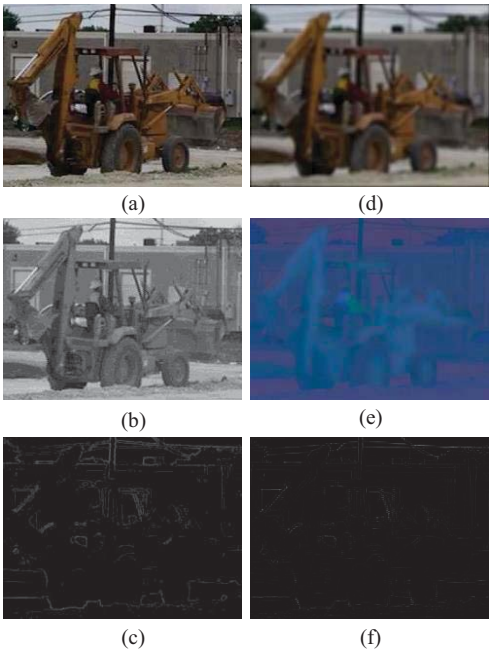


Figure 1. Various Channel Representation of Objects: (a) Original image, (b) Gray channel, (c) Gradient channel, (d) Gaussian channel, (e) LUV channels, (f) Laplacian of Gaussian channel.

B. Gradient Histogram & Magnitude

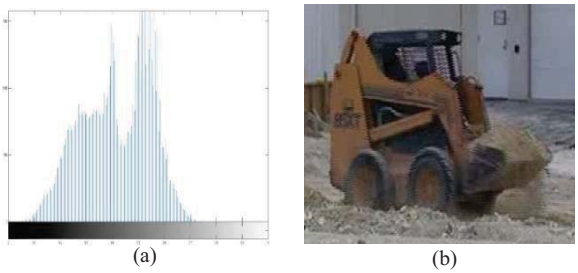
Dollar et al. (2009) defined a Gradient Histogram (Figure 2) as a weighted histogram where bin index is determined by gradient angle and weight. Gradient Histogram represents local shapes of objects and it has little variance to local geometric and photometric transformation such as translations and rotations.

Gradient Magnitude is the scale of gradient and it represents the difference of consecutive pixel values in an image. In other words, Gradient Magnitude describes edges of objects because edges are the positions where the image pixel values exhibit sharp variation. Thus, Gradient Magnitude is a scale-invariant feature and the edge information is not changed by moderate-rotation and scale-change of the objects.

Gradient Histogram and Gradient Magnitude were selected for this research. In detail, total six orientations of Gradient Histogram and one Gradient Magnitude were used (Figure 3).

C. Classifier Model

For classifier models, Random Forests are explored. Random Forests are one of the most popular classification models, which have a following characteristic; they construct a combination of weak classifiers other than a strong classifier so that classification accuracy can be significantly improved (Breiman, 2001). Each weak classifier has a weak classification ability which is better than a random classifier, but as a combination, a high accuracy can be acquired. In contrast, the strong classifier has a strong classification ability, but processing time is higher than the weak classifiers.



(a) (b)

Figure 2. (a) Gradient Histogram. (b) Original image of steer loader

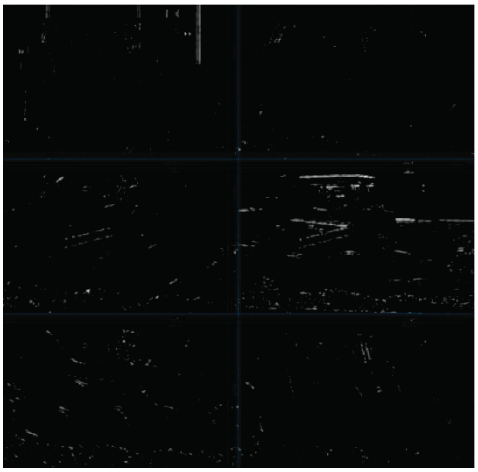


Figure 3. Gradient Channels of Six Orientations

III. Experiments

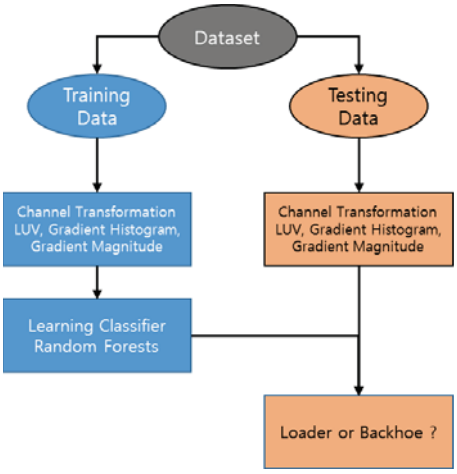


Figure 4. Framework of Experiments

A. Data Preparation & Description

The experiment design is illustrated in Figure 4. Video streaming data was collected from real construction sites. The number of total frames was 16,850 with 720 x 480 resolution. Steer loader and backhoe were the main equipment. The dataset was divided into a training dataset of 11200 frames and a testing dataset of 5650 frames. To show the potential of the proposed algorithm, the collected data was taken with different illuminations and view-points. Steer loader and backhoe were selected for testing equipment because they have large shape changes and pose-variation which are the key challenges for equipment detection.

B. Performance Evaluation Criteria

To analyze and evaluate the performance of the applied algorithm, the rules of the pattern analysis statistical modeling and computational learning (PASCAL) visual object classes challenge were adopted. The rules required that the matching portion of the detected and the ground truth areas ($B_p \cap B_{gt}$) should be more than 50% of the union of the detected and the ground truth areas ($B_p \cup B_{gt}$).

$$a_0 = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \geq 0.5$$

For further performance evaluation, missing rate versus FPPW (False Positive Per Window) curve, which is called a Detection Error Trade-off curve, was used. Missing rate explained a ratio of the number of objects which were not detected among all recorded objects. FPPW is the number of false alarms per one image window, which explains prediction errors.

IV. Results Discussion

A. Results

Figure 6 shows the detection results. Green boxes represented the predicted location of objects. From the results, the proposed algorithm was able to detect steer loaders and backhoes regardless of scale changes, different view-points, different postures, and illumination changes. Backhoes were detected although they were recorded in different view-points and scales and steer loaders were also detected with different view-points and working postures. The number above the green boxes are the detection scores. Based on the detection scores, it is available to ignore a low score of detection (e.g. lower than 1).

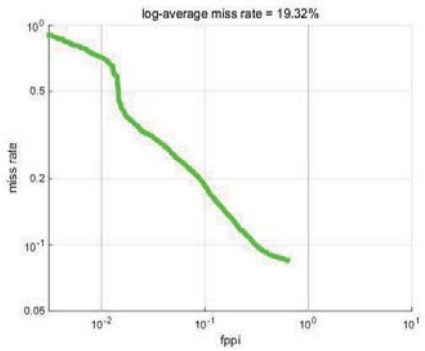
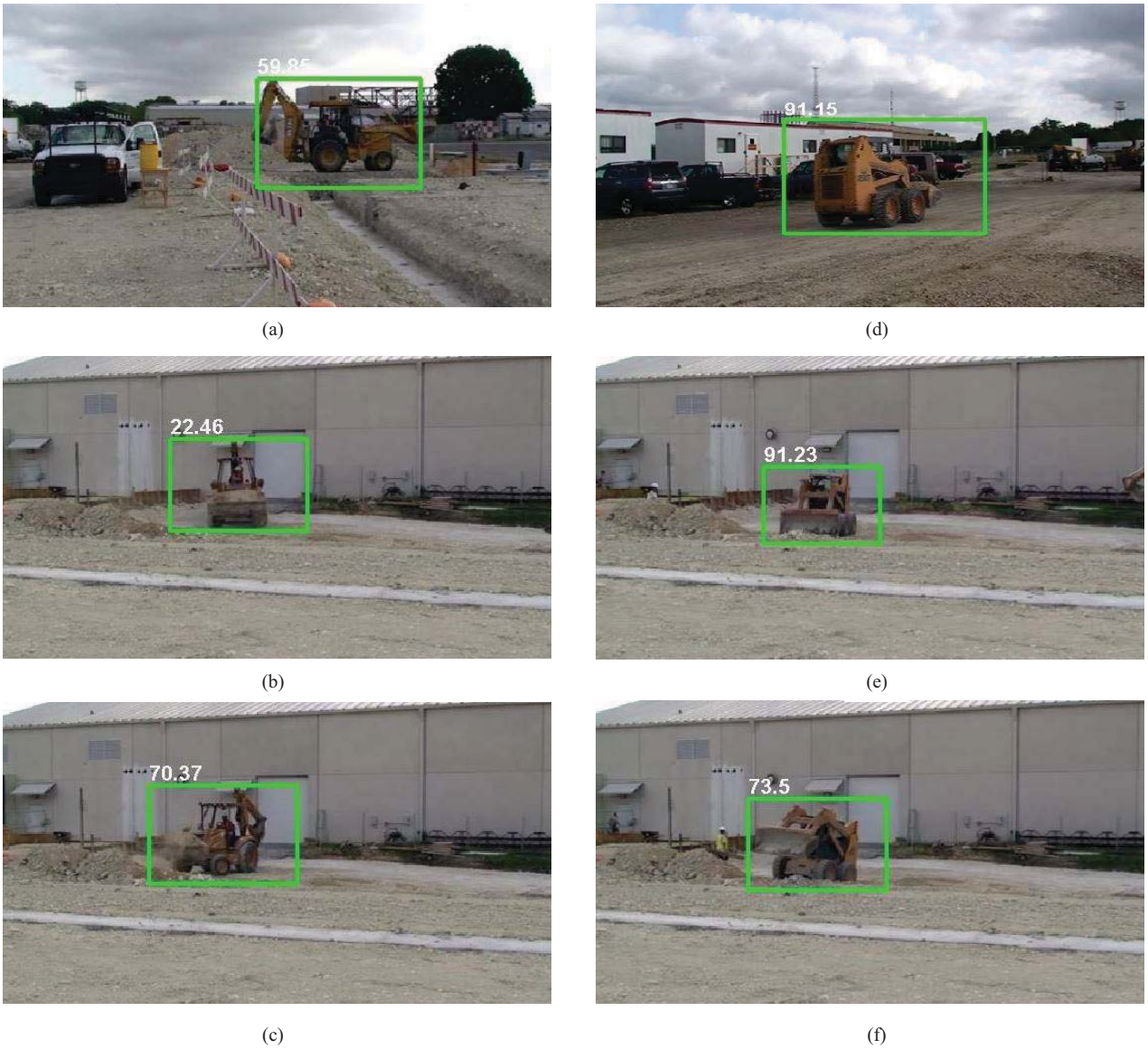


Figure 5. Detection Error Trade-off



(a) (d)

(b) (e)

(c) (f)

Figure 6. Detection Results: (a) The side of backhoe, (b) The front of backhoe, (c) The front-side of backhoe, (d) The back-side of steer loader, (e) The front of steer loader, (f) The front-side of steer loader and pose-variation of steer loader.

For numerical analyses, FPPW curve was plotted in Figure 5. The detection rate for steer loaders and backhoes was 81.68% with the 0.03 second processing time per each image. This is the promising result that the proposed algorithms have high potential for real-time application on construction equipment detection.

B. Discussion

This research showed acceptable and promising results for the key challenges. View-points, scale changes, poses and illumination variations had little effects on the performance of the algorithm.

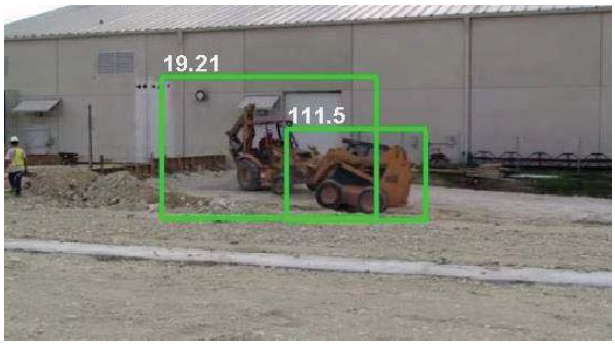
As shown in Figure 7, the proposed algorithms was also quite robust to partial and severe occluded cases. Such occlusion problems, however, can be handled better by multiple-camera monitoring network.

V. CONCLUSION

This paper presented an automated and real-time detection method based on Integral Channel Features. The results showed that the proposed algorithm worked in high performance in terms of accuracy and processing time. Although construction equipment and sites have various image processing challenges due to scale, view-point, illumination, and pose changes, the proposed algorithms showed acceptable results. With the high performance detector, reliable and usable data for productivity analyses and safety assessment are expected to be produced.

Despite of the acceptable performance, the detectors were not trained for all types of equipment. Future research needs to include a range of construction equipment for training and investigate specific object identification and tracking methodologies.

(a)



(b)

(c)

(d)

Figure 7. Detection Results: (a) Simultaneous detection of backhoe and steer loader, (b) Partial occlusion of backhoe and steer loader, (c) Simultaneous detection of backhoes, (d) Severe occlusion of backhoes.

VI. Acknowledgement

This research was supported by a grant(14SCIP-B079691-01) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement(KAIA) and the National Research Foundation of Korea (NRF) Grant (No. 2015R1A5A7037372) funded by the Korean Government (MSIP).

REFERENCES

[1] Brilakis, I., et al. (2011). "Automated vision tracking of project related entities." *Advanced Engineering Informatics* 25(4): 713-724.
[2] Chi, S., et al. (2009). "A Methodology for Object Identification and Tracking in Construction Based on Spatial Modeling and Image Matching Techniques." *Computer-Aided Civil and Infrastructure Engineering* 24(3): 199-211.
[3] Chi, S. and C. H. Caldas (2012). "Image-Based Safety Assessment: Automated Spatial Safety Risk Identification of Earthmoving and Surface Mining Activities." *Journal of Construction Engineering and Management* 138(3): 341-351.
[4] Chi, S. and C. H. Caldas (2011). "Automated Object Identification Using Optical Video Cameras on Construction Sites." *Computer-Aided Civil and Infrastructure Engineering* 26(5): 368-380.
[5] Criminisi, A., et al. (2011). Deicision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning.
[6] Ethem A., et al. (2014). Introduction to Machine Learning 3rd Edition, The MIT Press.
[7] Golparvar-Fard., et al. (2013). "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers." *Advanced Engineering Informatics* 27(4): 652-663.
[8] Gong and Caldas (2010). "Computer Vision-Based Video Interpretation Model for Automated Productivity Analysis of Construction Operations". *Journal of Computing in Civil Engineering* 24(3), 252-263.
[9] Hwang, S., Park, J., Kim, N., Choi, N., Kweon, I. S. (2015). " Multispectral Pedestrian Detection: Benchmark Dataset and Baseline". *In Proc. of CVPR 2015.*
[10] Memarzadeh.., et al. (2013). "Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors." *Automation in Construction* 32: 24-37.
[11] Park, M.W. and I. Brilakis (2012). "Construction worker detection in video frames for initializing vision trackers." *Automation in Construction* 28: 15-25.
[12] Park, M.W., et al. (2012). "Three-Dimensional Tracking of Construction Resources Using an On-Site Camera System." *Journal of Computing in Civil Engineering* 26(4): 541-549.
[13] Park, M.W., et al. (2011). "Comparative study of vision tracking methods for tracking of construction site resources." *Automation in Construction* 20(7): 905-915.
[14] P. Dollar, Z. Tu, P. Perona, and S. Belongie (2009). "Integral channel features," In *BMVC*, 2009.
[15] P. Dollar, C. Wojek, B. Schiele, and P. Perona. (2012). "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. on PAMI*, 2012.
[16] P. Viola and M. J. Jones. (2001). "Rapid Object Detection using a Boosted Cascade of Simple Features," In *Proc. of CVPR* 2001.
[17] P. Cruse. (2015). http://www.colourphil.co.uk/lab_lch_colour _space.shtml
[18] Rezazadeh Azar, E. and McCabe, B. (2012). " Automated Visual Recognition of Dump Trucks in Construction Videos." *Journal of Computing in Civil Engineering*, 26(6), 769 – 781.

[19] N. Dalal and B. Triggs. (2005). "Histograms of oriented gradients for human detection," In *Proc. of CVPR* 2005.
[20] Yang, J., et al. (2010). "Tracking multiple workers on construction sites using video cameras." *Advanced Engineering Informatics* 24(4): 428-434.
[21] Richard Szeliski. (2010). Computer Vision: Algorithms and Applications. Springer, New York, 2010.
[22] Mark S. Nixon and Alberto S. Aguado. (2012). Feature Extraction &Image Processing for Computer Vision 3th Edition.
[23] Richards, J. A., Remote. (1993). Sensing Digital Image Analysis: An Introduction 2nd Edition.