# Dataset Development for the Recognition of Construction Equipment from Images

**Humaira Tajeen**

A Thesis

In the Department

of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Building Engineering) at

Concordia University

Montreal, Quebec, Canada

September, 2013

**CONCORDIA UNIVERSITY**
**School of Graduate Studies**

This is to certify that the thesis prepared

By: Humaira Tajeen

Entitled: Dataset Development for the Recognition of Construction Equipment from Images

and submitted in partial fulfillment of the requirements for the degree of

**Master of Applied Science (Building Engineering)**

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Tarek Zayed

_____ Examiner
Dr. Amin Hammad (CIISE)

_____ Examiner
Dr. Osama Moselhi

_____ Examiner
Dr. Tarek Zayed

_____ Supervisor
Dr. Zhenhua Zhu

Approved by: _____
Dr. Maria Elektorowicz, GPD
Department of Building, Civil and Environmental Engineering

_____2013

_____
Dr. Christopher Trueman
Dean, Faculty of Engineering and
Computer Science

**ABSTRACT**

**Dataset Development for the Recognition of Construction Equipment from Images**

Humaira Tajeen

The construction industry, being one of the largest industrial sectors in Canada, has been continually searching for automated methods that can be adopted to monitor the productivity, consistency, quality and safety of its construction work. The automated recognition of construction operational resources (equipment, workers, materials etc.) has played a significant role in achieving the full automation in monitoring and control of the construction sites. Considering that construction equipment is one of the main operational resources in executing construction tasks, this research work is focused on automated recognition of such equipment from on-site images. In order to achieve this, it is first necessary to evaluate the construction equipment recognition performances of existing object recognition methods. The currently available object recognition datasets that are used to validate the existing recognition methods contain only limited categories of objects, where construction equipment are not included. As a result, it is unclear whether these methods could be used to recognize construction equipment from on-site images, especially considering that construction sites are typically dirty, disorderly, and cluttered. To fill this gap, this research work proposes to create a standardized dataset of construction equipment images that can be used to measure the construction equipment recognition performances of existing object recognition methods. Almost 2,000 images have been collected and compiled to create the dataset, which covers 5 common classes of construction equipment (excavator, loader, tractor,

compactor and backhoe loader). Each image has been annotated with information concerning the equipment class, identity, location, orientation, occlusion, and labeling of equipment components (bucket, stick, boom etc.). The effectiveness of the dataset has been tested on two common object recognition methods in computer vision. The recognition tests imply that the recognition methods can be adopted comprehensively for the recognition of construction equipment with the dataset developed in this research. The performances of these two methods are further compared on the basis of the recognition tests conducted in this work. The results show that the construction equipment recognition performance of existing object recognition methods can be evaluated with the dataset in a standard, unbiased, and extensive way.

**DEDICATION**

*I dedicate this thesis to my beloved parents and sister for their endless love, affection, support and inspiration in every step of my life.*

**ACKNOWLEDGEMENTS**

My deepest gratitude and solemn appreciation goes to my supervisor, Dr. Zhenhua Zhu, for his enthusiastic encouragement, close supervision, and guidance. His thoughtful advice, detail feedback and constructive criticism motivated me to make this research possible. I would like to thank him for suggesting various ideas and insightful comments at different stages of my research. I would also like to thank Dr. Osama Moselhi, Dr. Tarek Zayed and Dr. Amin Hammad for their constructive criticism, expert advice and valuable opinions regarding the work during the oral presentation.

I would like to acknowledge the financial support provided by SEED funding from the Vice President's Research and Graduate Studies (VPRS) at Concordia University, for this research. Furthermore, I take the opportunity to thank the construction engineers and superintendents, who were kind enough to allow me taking photographs from the construction sites. Last but not the least, I am indebted to my mother, whose inspiration and supports have always been with me, throughout my study. I would also like to thank my sister; her love will always be remembered. Finally, I wish to express my extended and special thanks to my husband Faisal for his support, guidance, understanding, patience, and encouragement.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| P | Precision |
| R | Recall |
| TP | True Positive |
| FP | False Positive |
| TN | True Negative |
| FN | False Negative |
| ACC | Accuracy |
| $F_1$ | F-measure |
| AP | Average Precision |
| TPA | True Positive Accuracy |
| TPR | True Positive Rate |
| P/R curve | Precision-Recall curve |
| EOI | Equipment of Interest |
| CAT | Caterpillar |
| DOM | Document Object Model |
| GUI | Graphical User Interface |
| XML | Extensible Markup Language |
| CERD | Construction Equipment Recognition Dataset |
| SIFT | Scale-Invariant Feature Transform |
| SURF | Speeded Up Robust Features |
| HOG | Histogram of Oriented Gradients |

CHAPTER 1

# INTRODUCTION

## 1.1 MOTIVATION

The construction industry has been transformed as one of the largest industrial sectors in Canada (Chutter, 2012; Historica-Dominion, 2012). Similar to other industrial sectors, the construction industry could be enormously benefited by adopting automation in its operation. The automation can facilitate a construction project to finish on time, within budget, and with high quality (Zou and Kim, 2007). The proper implementation of automation can solve the prevailing problems of low productivity and delayed project completion in the construction industry, since it can improve the speed and consistency of construction operations by reducing operation cycle time and minimizing equipment idle time (Heydarian and Golparvar-Fard, 2012; Tatum, 1989). It has the potential to perform the tasks that are beyond human capabilities in size, weight, speed, etc. (Elatter, 2008). In addition, automation can reduce the requirements of human labor and thus, can save the labor cost to a greater extent. The timely completion eliminates the possibility of exceeding total cost of a construction project. Moreover, the poor quality, which is another common phenomenon of construction works, can be overcome by replacing human labor in case of repetitive and monotonous physical work (Demsetz, 1990). Furthermore, the safety of construction workers can be enhanced by substituting them with automated facilities for difficult and tedious tasks, and in hazardous construction environments (Elatter, 2008). Thus, the construction industry can be benefited from the proper implementation of automation, since

automation in construction can significantly increase the productivity, cost efficiency, quality and safety of construction works (Heydarian and Golparvar-Fard, 2012).

In order to achieve the potential benefits, construction researchers and professionals have been working hard towards promoting the construction automation (Gong and Caldas, 2010; Heydarian and Golparvar-Fard, 2012), where construction site images have been utilized as the basis of developing a significant portion of these automation work (Brilakis et al., 2006). Considering their acceptable return on investment (ROI), high-resolution digital cameras have been increasingly employed at construction sites (Bohn and Teizer, 2010). The time-lapse images collected from the construction sites do not only record the as-built progress of the projects under construction, but also capture the daily job site activities (Nitithamyong and Skibniewski, 2004). As an example, the site images can be used to indicate the location and states of on-site construction operational resources, such as whether the materials are stored in the right place or not, whether an equipment is in idle state or in operation etc. This way, useful management information can be obtained from analyzing the construction site images, which consequently facilitates the construction engineers/managers to monitor and control sites remotely and dynamically (Nitithamyong and Skibniewski, 2004). As a result, the prevailing problems of the traditional monitoring and control could be overcome, which has been executed manually and hence slow, inefficient, labor intensive, error-prone and unreliable (Navon and Berkovich, 2006; Davidson and Skibniewski, 1995). Considering the fact that construction site images have the potential to reflect important information about construction site activities, necessary steps should be taken to automatically retrieve these information from the site images.

The construction site images could be fully utilized for the automation of construction work, if only the automatic recognition of various construction operational resources (e.g. equipment, workers, and materials) from images could be achieved. The successful recognition of such construction operational resources could facilitate many construction monitoring and control tasks to be performed in an automated and remote way (Figure 1-1). When such construction operational resources are successfully recognized, the traditional way of monitoring and control tasks of a construction project could be significantly transformed.

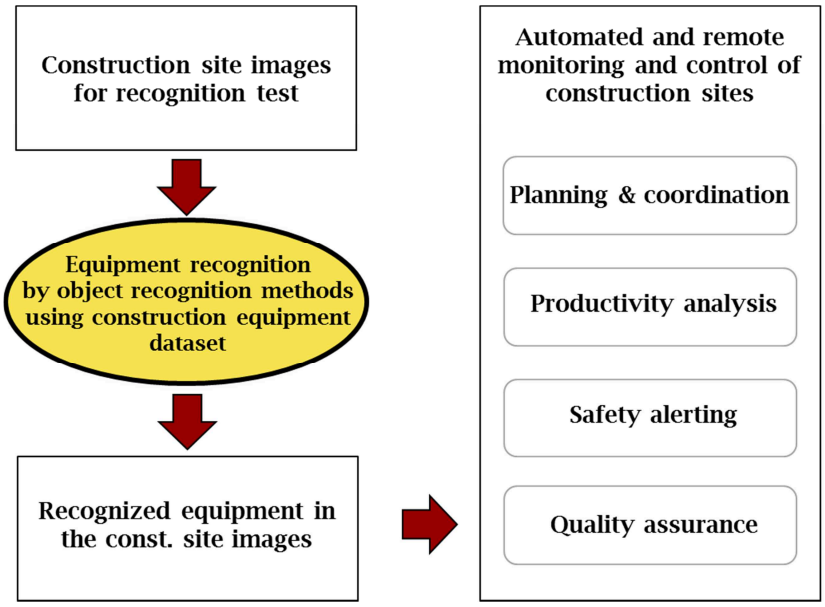



Figure 1-1: Potential applications of construction equipment recognition from images

For example, the recognition of construction equipment can facilitate the automated productivity analysis of a construction project (Azar and McCabe, 2012; Gong et al. 2011). Also, the recognition of workers can be used in order to track the location of on-site workforce; observe their performances; communicate with them when necessary,

and investigate accidents (Chi and Caldas, 2011). Moreover, the recognition of construction materials, which constitute a large portion of the total construction cost, is necessary in order to ensure proper handling, storage and availability when they are needed throughout the construction work (Kasim et al. 2012). Therefore, the ultimate goal of the automated recognition of construction resources is to facilitate the automated monitoring and control of construction projects, which eventually enables the construction engineers/managers to take any rapid, corrective decision for the improvement of a construction project, such as tracking equipment in order to minimize their idle time (Gong and Caldas, 2010), monitoring site personnel in order to ensure communication and safety, tracking construction materials in order to make them available at the right time, right place (Song, 2005) and with exact quantity (Kasim et al. 2012). Thus, the automated recognition of construction operational resources plays a significant role towards successful, cost-effective and timely project completion.

However, the automatic recognition of construction resources under real construction site conditions is not an easy task. This is for the fact that construction sites are generally characterized as dirty, untidy and cluttered with machines, tools, materials and debris. For this reason, the resources at construction sites are typically viewed with partial occlusions, and against heavily cluttered background. Therefore, the recognition of on-site construction resources has been perceived as difficult and challenging, due to the disorderly characteristics of typical construction sites.

So far, many object recognition methods have been developed by the researchers, mainly in computer vision community (as discussed in the following chapter). Also, the effectiveness of these methods has been tested through different datasets that have

been created and made publically available, until now. However, these datasets contain only limited classes of objects in natural scenes, such as pedestrians, human faces, bicycles, cars, etc. and none of them include construction equipment, as it is best known to the author. As a result, it is unknown whether or not the existing object recognition methods could be used to recognize on-site construction operation resources under real site conditions. In order to address to this question, the necessity of developing a new dataset, which will cover typical construction equipment images under realistic site conditions, has been perceived.

## 1.2 RESEARCH GOAL AND OBJECTIVES

The ultimate goal of this research is placed on the automated recognition of construction equipment from images that are captured under real site conditions. The objectives are: (1) to develop a dataset, which comprises images of different types of construction equipment (excavator, loader, tractor, compactor and backhoe loader) with a wide variety of sizes, poses, and camera viewpoints; and (2) to evaluate the performance of existing object recognition methods for the recognition of construction equipment from real site images, using the dataset developed in this work. The goal and objectives of the current research are illustrated in the Figure 1-2.

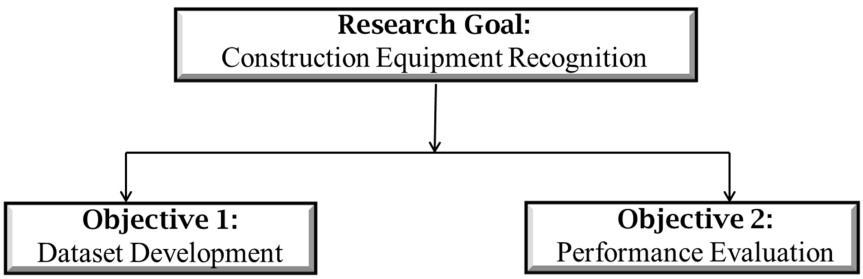



Figure 1-2: Research goal and objectives

In order to create a diverse and rich dataset, hundreds of images have been collected for each class of construction equipment. The dataset includes equipment images with illumination variations, and partial occlusions by debris, materials and other equipment at construction sites. When all the images are collected and compiled, a MATLAB-based annotation tool has been created to annotate the Equipment of Interest (EOI) in these images. The annotation includes various information about the image and the equipment contained in it. For example, it contains information about the image resolution, equipment type, location, viewing angle, occlusions, and labels of corresponding equipment components, such as bucket, stick, boom, cab, tracks, wheels etc. The images and annotations are used as ground truth to evaluate the construction equipment recognition performance of existing object recognition methods.

So far, two object recognition methods have been tested and their performances are evaluated. These methods are: 'discriminatively trained part-based model method' developed by Felzenszwalb et al. (2010) and 'a simple object detector with boosting method' developed by Torralba et al. (2004). The results show that the image dataset developed in this paper can evaluate the methods in a standard, unbiased, and extensive way. Based on the results, it is found that none of these recognition methods are absolutely perfect regarding all the performance criteria. However, the method developed by Felzenszwalb et al. (2010) performed more robustly against partial occlusions and pose variations, while the method proposed by Torralba et al. (2004) is computationally favorable as it needed less time for construction equipment recognition.

## 1.3 ORGANIZATION OF THE THESIS

This research work will be presented as follows:

- Chapter 1: Introduction; A brief introduction on the motivation of this research; i.e. the benefits of adopting automation in construction sector, the use of construction site images, the necessity of automated recognition of construction operational resources (materials, workers and equipment), the necessity for developing a new dataset for construction equipment images, and the 'gap in knowledge' identified in this respect, which drives the main force of this research work.
- Chapter 2: Literature Review; Literature search for the related work in the domain of object recognition from images, categories of existing object recognition methods, available datasets for recognition performance evaluation, typical performance metrics that are commonly used by the researchers for comparing the performances of object recognition methods.
- Chapter 3: Objectives and Scope; In this chapter, the objectives of the research work, along with the scope, is elaborated.
- Chapter 4: Development of Construction Equipment Image Dataset; This chapter describes the steps involved in developing the dataset, factors affecting the image collection process, the image annotation procedure, development of the annotation tool, examples of annotation information contained in the XML files.
- Chapter 5: Construction Equipment Recognition Tests; Detailed description of the experimental setup (hardware and software configurations of the computer), execution of 'model training' and 'recognition testing' phases for the two methods, examples of the recognition results generated by the methods tested in this work.

- Chapter 6: Results and Discussion; The performances of the methods are compared methodically and elaborately (on the basis of three common performance metrics, i.e. correctness, robustness and computation speed); an analytical discussion on the recognition performances of both the methods is made.
- Chapter 7: Conclusions and Future Work; This chapter includes the summary of the present research work, highlights its contributions, and proposes the future direction of this research.

CHAPTER 2

# LITERATURE REVIEW

Object recognition from images has been considered as a challenging task. The recognition of three-dimensional (3D) objects from images (2D) is often complicated since the appearance of a 3D object can transform drastically with the change of relative pose to the camera and the viewing angle. Also, an object may have multiple sizes and shapes. Moreover, the object in the image can appear with partial occlusions, against heavily cluttered background, and experience different environmental lighting conditions (Yang 2009; Ulrich and Steger, 2002). Considering the fact that recognizing 3D objects is more complex in nature than 2D shapes/characters, researchers in computer vision have developed many object recognition methods for recognizing 3D objects from images.

## 2.1 CATEGORIES OF OBJECT RECOGNITION METHODS

The 3D object recognition methods, which are developed so far, are distinct in nature from each other with respect to the strategy they follow for recognition. Based on the type of the recognition cues employed, these methods can be broadly classified into three categories: (1) the geometry-based category, (2) the appearance-based category, and (3) the feature-based category (Matas and Obdrzalek, 2004; Yang 2009). The geometry-based recognition methods rely on the shape or silhouette of the object. Here, other properties of the object, such as color and texture, are not used. On the contrary, the appearance-based methods typically consider the object surface reflectance

properties, such as brightness and contrast, as recognition cues. In the feature-based methods, the object visual features, such as surface patches and interest points, are used for matching (Matas and Obdrzalek, 2004). However, all the categories of object recognition methods have common characteristics regarding recognition process. Typically, recognition is performed in two phases – the 'training' and the 'testing' phase. In order to evaluate the performance of these methods, several datasets have been developed with objects in natural scenes, such as people, car, bicycle etc. Categories of object recognition methods are summarized in Figure 2-1.

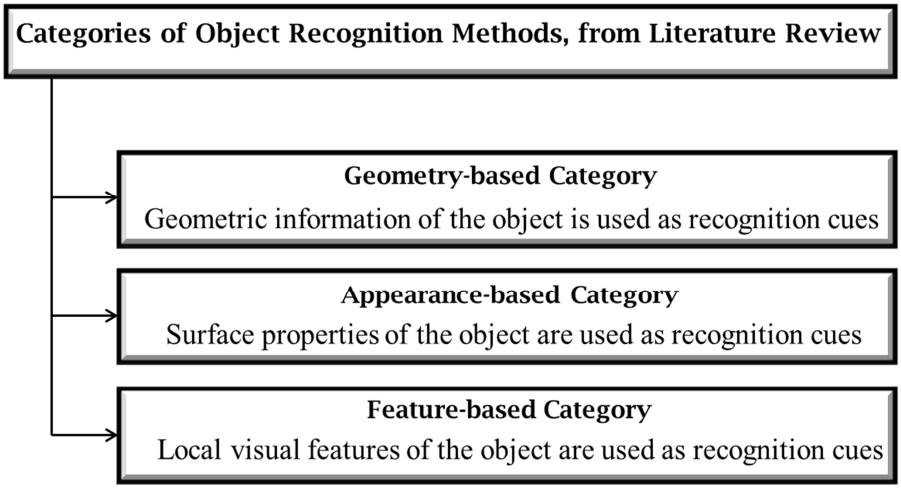



Figure 2-1: Categories of object recognition methods

### 2.1.1 GEOMETRY-BASED CATEGORY

In the geometry-based methods, 3D geometric primitives (e.g. boxes, spheres, cylinders, etc.) or 2D shapes and contours are used to represent an object, without detailed information of additional object properties such as color and texture. Then, a hierarchical organization of the primitives, shapes/contours is created. This hierarchy is used to define the model of the corresponding object. When the model is created, the

object recognition can be performed by measuring the geometric similarity between that object model and all the geometric information that can be retrieved from an image of the object (Pope, 1994). The object is recognized if its geometry is similar to the geometric information contained in the model of the conforming object (Pope, 1994).

In order to measure the geometric similarity, several methods have been developed so far, such as the hierarchical chamfer matching (Borgefors, 1988), geometric hashing (Lamdan and Wolfson, 1988), and shape-based matching (Steger, 2001). Also, the similarity can be measured using the Hausdorff distance transform (Rucklidge, 1995) or generalized Hough transform (Ballard, 1981). Rucklidge (1995) developed the method based on Hausdorff distance measure, where the distance of each pixel of the reference image is measured and compared with the pixels of the corresponding search image. Steger (2001) used a set of points and their corresponding direction vectors to construct the model, which is then compared to the search image to compute a matching score between the model and the image. The object is recognized when the matching score reaches a satisfactory level defined by the method (Steger, 2002).

The geometry-based methods have been considered more robust for the recognition of the objects with small degree of occlusions or background clutter, in particular, when compared to the appearance-based methods (Yang, 2009). Also, they are moderately invariant to small degree of lighting and viewpoint changes (Matas and Obdrzalek, 2004; Yang, 2009). To the contrary, the detection of all the geometric primitives is challenging, especially in the case of large illumination variations. As a result, these methods are not robust in the cases of large illumination variations, and heavy occlusions and/or background clutter (Matas and Obdrzalek, 2004). Moreover, the

geometry-based methods are typically restricted for recognition of the objects that have easily identifiable components, as the effectiveness of these methods is highly dependent on the reliable extraction of geometric primitives (Matas and Obdrzalek, 2004). Furthermore, the methods are often computationally expensive, especially for the recognition of objects with deformable parts (Ulrich and Steger, 2002; Pope, 1994). Overall, the geometry-based methods require long computation time to recognize objects (Pope, 1994).

### 2.1.2 APPEARANCE-BASED CATEGORY

The appearance-based methods refer to those methods that rely on object color, texture and/or surface reflectance (albedo) properties as recognition cues (Matas and Obdrzalek, 2004). Here, any geometric information of the object is not required. These methods have been developed on the concept of "remembering all possible appearances" of an object (Matas and Obdrzalek, 2004). The effective recognition is entirely dependent on retaining large number of diverse views of the object, which is usually captured by two-dimensional images of the object-of-interest from different viewpoints. In the first phase, i.e. "training" phase, an appearance model is constructed based on the set of reference images that includes the object's multiple views under different orientation and illumination conditions. The second phase is the "recall" phase, where the parts of a test image are first extracted through image segmentation, and then the recognition is performed by matching the extracted parts of the test image with the model constructed in the "training" phase (Matas and Obdrzalek, 2004).

So far, several appearance-based methods have been developed to recognize objects. For example, Murase and Nayar (1995) developed a method where image Eigen values

is used as a basis of recognizing objects with different viewpoints and illumination variation. Swain and Ballard (1991) relied on the image histograms, where an object is represented by a color histogram and recognition is performed by matching the histograms of the search image and the model image. The effectiveness of these methods has been validated for recognizing objects without occlusions or against black background (Nayar et al. 1996). More recent works in this category include: k-nearest neighbor (Zhang et al., 2007), neural networks with radial basis function (RBF) (Poggio and Edelman, 1990), support vector machines (SVM) (Sch¨olkopf and Smola, 2002), sparse network of Winnows (SNoW) (Yang et al., 2000) etc.

There are two main advantages of the appearance-based methods in comparison to the geometry-based approach. First, the methods do not require any user-provided models (Matas and Obdrzalek, 2004). The models can be automatically generated from the training images. Second, the methods show invariance under controlled variations in illumination and viewpoint conditions (Yang, 2009). However, the use of the appearance-based methods is restricted since they require complete segmentation of the object-of-interest from the background and hence they are sensitive to object occlusions and cluttered background (Matas and Obdrzalek, 2004). As a result, these methods are not always robust and they are mainly suitable to recognize rigid objects (Dorkó and Schmid, 2005). Another major limitation of this approach is that they suffer from a lack of invariance to similarity transformations, such as scale or rotation (Dorkó and Schmid, 2005). Moreover, the appearance-based methods require dealing with all variations of the object appearance, which is computationally unfavorable (Dorkó and Schmid, 2005).

### 2.1.3 FEATURE-BASED CATEGORY

The feature-based object recognition methods have been evolved in more recent times. These methods are developed on the idea that an object is represented by a set of local visual features, such as the surface patches, corners, or other interest points with intensity discontinuity. These local features are typically invariant to scale, illumination and affine transformation (Yang, 2009). In the training phase, the features are learned from the object. In the testing phase, the learned features are compared with the features extracted from the search image. Then, the number of matched visual features is determined in order to assess the presence of the object in the search image. The presence of the object, in the corresponding images, is determined if the number of matched features are adequately high (Felzenszwalb et al., 2010).

There are several visual feature detectors and descriptors that have been developed to extract these features. For example, the Scale-Invariant Feature Transform (SIFT) (Lowe, 1999), the Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005), and the Speeded Up Robust Features (SURF) (Bay et al., 2006) etc. These feature descriptors are used to learn the features from the object at the initial step to create a model of the corresponding object. For the final step, i.e. matching, researchers often employ exact nearest neighbor search, i.e., kd-trees (Freidman et al., 1977) or approximate similarity search methods, i.e. hashing-based algorithms (Grauman and Leibe, 2011). The recent emergence of 'locally invariant visual features matching' concept has been used immensely in many areas of computer vision, such as object recognition, image retrieval, and stereo matching etc. (Grauman and Leibe, 2011).

Feature-based object recognition is a powerful and robust approach, since the detection and description of local visual features are invariant to scale, illumination and/or affine transformation (Matas and Obdrzalek, 2004). Also, it is not essential that all the local features are matched for the successful recognition, since very few matches can determine the presence of the object in the search image. As a result, these methods are applicable even for the partially occluded objects and/or against cluttered background (Lowe, 1999). Additionally, no user-provided model is required in this approach since the object features can be automatically extracted and learned from a set of training images (Matas and Obdrzalek, 2004). Moreover, segmentation of objects from background is not necessary and the objects can be recognized under any unknown background (Matas and Obdrzalek, 2004). Furthermore, these methods are effective for significant changes in viewpoint and illumination conditions, since these methods rely on the principle of matching features that are invariant to scale, illumination and affine transformation (Matas and Obdrzalek, 2004). Considering all these advantages of feature-based approach over geometry- and appearance-based methods, it can be anticipated as the potential approach for the recognition of construction operational resources under real construction site conditions. Recently, some researchers have introduced these methods for the recognition of different construction objects, such as trucks (Azar and McCabe, 2012), construction workers (Park and Brilakis, 2012; Gong et al. 2011) etc.

## 2.2 CURRENT DATASETS FOR OBJECT RECOGNITION PERFORMANCE EVALUATION

The object recognition research has been experienced much advancement and many recognition methods have been developed by the researchers until now. However, most of the existing recognition methods are still sensitive to large illumination variations,

heavy occlusions and background clutter (Yang 2009). Therefore, it is necessary to evaluate and record their performances for future improvement. In order to evaluate the performance of existing object recognition methods, several datasets have been developed, such as the datasets created by the MIT (Torralba et al., 2004), UIUC (Agarwal et al., 2004), CALTECH (Griffin et al., 2007), YALE (Georghiades et al., 2001), CMU (Sim et al., 2002), etc. These datasets are created by collecting a large number of images covering limited object classes in natural settings. Take the LabelMe dataset as an example, which was developed at Massachusetts Institute of Technology (MIT). It includes the images of bicycle, bottle, apple, bookshelf, car, chair, desk, sofa, building, door, window, and the scenes viewed from offices or at streets (Russell et al. 2008).

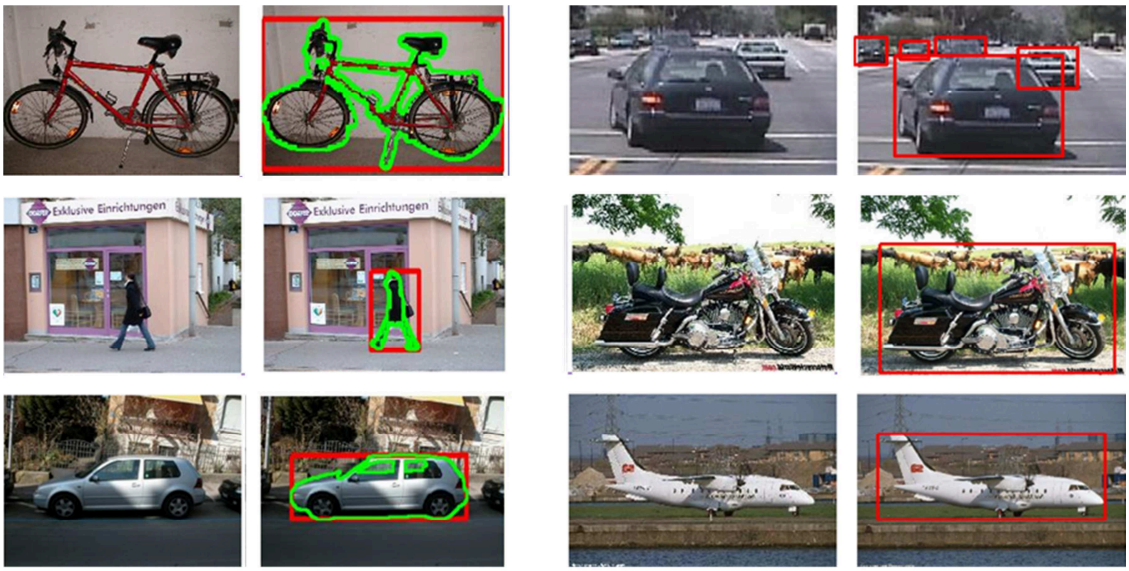
The researchers at the University of Illinois at Urbana-Champaign developed the UIUC dataset, which contains the images of cars with side views only (Agarwal et al. 2004). The CALTECH-101 and CALTECH-256 datasets that were developed at the California Institute of Technology cover multiple classes of objects. CALTECH-101 contains 101 categories of objects including aeroplanes, cars, human faces, motorbikes etc. (Fei-Fei et al., 2006). In the CALTECH-256 dataset, the successor of the CALTECH-101, the number of classes was increased from 101 to 256 (Griffin et al., 2007). The INRIA dataset was created as a part of the research work in human detection. It comprises the images of people only with the upright positions (Dalal and Triggs, 2005). The PASCAL VOC datasets were developed as a standardized collection of numerous object recognition datasets. For example, the VOC 2005 dataset contains images from other datasets, including TU-Darmstadt, Caltech, TU-Graz, UIUC and INRIA datasets. It contains 1,578 images of motorbikes, bicycles, people, and cars in arbitrary poses (Everingham et al., 2006). Again, the VOC 2006 dataset comprises 5,304 images with more object

categories, such as bicycles, buses, cats, cars, cows, dogs, horses, motorbikes, people, and sheep with random poses (Everingham et al., 2006). The recently developed PASCAL VOC dataset contains twenty visual object classes, i.e. person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, and TV/monitor (Everingham et al., 2010). Altogether, these publically available multi-class datasets played an important role towards the recent development of category-level object recognition research (Ponce et al., 2006).

In the aforementioned datasets, the annotations of the objects-of-interest are included along with the collected images. The images are manually annotated to obtain the annotation files that are used as the ground truth for recognition. For instance, the TU-Graz dataset (developed at Graz University of Technology) includes 3 object classes – bicycles, people, and cars – where boundary polygons are used to annotate the objects in the images (Figure 2-2a). The CALTECH dataset contains 4,620 annotated images of aeroplanes and motorbikes (side views), cars (rear views), faces (front views) and general background scenes. The original ground truth data is created in the form of a bounding quadrilateral, which is then converted into a bounding rectangle following the original annotations – shown in Figure 2-2b (Fergus et al., 2003).

The MIT-CSAIL dataset includes 72,000 images of objects and scenes, among which 2,873 have been annotated with boundary polygons for the corresponding object or region (Torralba et al., 2004). Figure 2-2c illustrates the ground truth annotations for the MIT-CSAIL dataset, where objects are annotated with polygons. The TU Darmstadt Dataset (formerly known as ETHZ Dataset) is created at Darmstadt University of Technology by including side views of motorbikes, cars and cows. The ground truth

annotations are provided as bounding boxes for the motorbikes, and polygons for the cows and the cars –represented in Figure 2-2d (Leibe et al., 2004). In all the datasets, the objects are commonly labeled with information about the object class, identity, pose, viewpoint etc.




(a) TU-Graz dataset (b) CALTECH

(c) MIT-CSAIL dataset (d) TU Darmstadt

Figure 2-2: Databases with ground truth annotations: (a) TU-Graz dataset, (b) CALTECH dataset, (c) MIT-CSAIL dataset and (d) TU Darmstadt dataset

Although the datasets that are currently available provide a common ground truth to evaluate the existing object recognition methods, there are several issues restricting the use of the current datasets for the recognition of construction equipment. First, the datasets only contain limited object classes in natural scenes. None of these datasets included construction equipment, as it is best known to the author. Second, the images in the datasets reflect a small range of variations regarding the pose and position of the object-of-interest in the image. Also, the view point and orientation of the objects do not seem to change largely. Most of the current datasets contain such images, where objects are presented with their stereotype poses and placed at the image centers (Ponce et al., 2006). Moreover, the images in the datasets are mostly captured with little or no occlusion and background clutter (Ponce et al., 2006). For all these reasons, it has been unpredictable whether the existing recognition methods could be used for the purpose of construction equipment recognition from on-site images. In order to answer this question, it is first necessary to create a new dataset, which will cover typical construction equipment images under realistic and diverse site conditions, such as multiple pieces of equipment working together with illumination variations and partial occlusions by debris and materials. Therefore, the creation of the construction equipment image dataset is an essential part to achieve successful recognition of construction equipment from images.

## 2.3 METRICS USED FOR OBJECT RECOGNITION PERFORMANCE EVALUATION

The performances of the existing object recognition methods can be evaluated by using the datasets, since the annotated images of the datasets offer a common ground truth. As suggested by Pope (1994), and Ulrich and Steger (2002), the performance of the

object recognition methods is measured on the basis of three performance metrics, i.e. (1) Correctness, (2) Robustness and (3) Speed. Correctness of a method represents the quality of proper implementation of the method's intended ranking and decision criteria. To evaluate the performance of a method regarding correctness, several measures are used such as – precision, recall/sensitivity, specificity, accuracy, F-measure etc. Robustness of an object recognition method denotes the level of tolerance it shows against noise, occlusions and illumination variations of the scenes. Also, speed of a method signifies the inverse of the amount of time it requires for the computation of its corresponding search space to recognize an object (Pope, 1994). In order to evaluate the performances of different object recognition methods, these metrics are widely used as common criteria for making rational judgment. The criteria that are commonly used to measure the performance of object recognition methods are summarized in Figure 2-3.

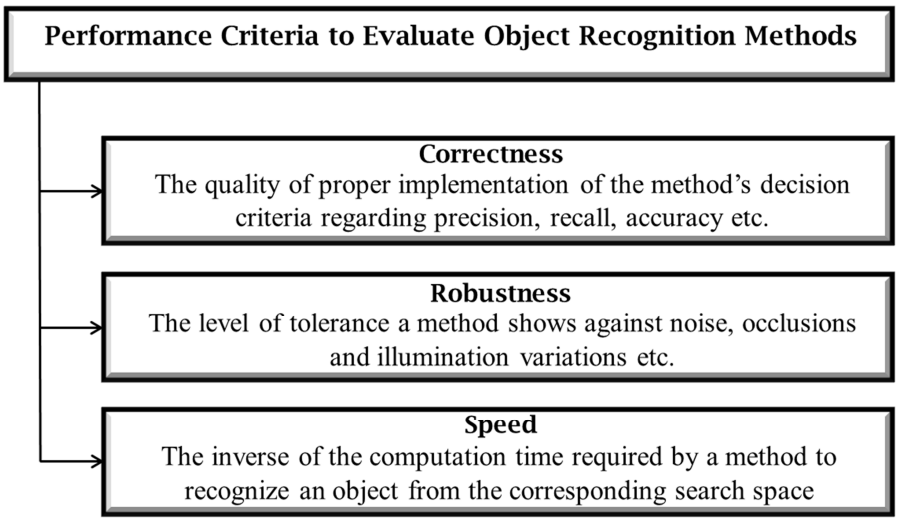



Figure 2-3: Common criteria for object recognition performance evaluation

### 2.3.1 CORRECTNESS: PRECISION, RECALL, ACCURACY, F-MEASURE ETC.

In order to measure the correctness of a recognition method, it is first necessary to calculate the true positives (TP), false positives (FP), true negatives (TN), false negatives (FN) etc. recognized by the method (Taylor, 1999). The number of positive instances that are correctly recognized as positive is called the true positive (TP), whereas the number of negative instances that are wrongly recognized as positive is called the false positive (FP). The number of negative recognition in case of positive instances is known as the false negative (FN) and if the negative labeled instances are correctly recognized as negative, they are called true negatives (TN). The FP and FN are often referred to as type I and type II errors respectively (Sheskin, 2004).

After the TP, FP, TN, and FN are calculated, the Precision (P), Recall (R), Accuracy (ACC), F-measure ($F_1$), and Average Precision (AP) could be further estimated on the basis of these values. The Precision (P) or Confidence denotes the proportion of the correct real positives among all the positive cases recognized (Powers, 2011). It can also be called true positive accuracy (TPA). High precision means many positive instances detected by the method are correct real positives, which means the number of false alarms is comparatively low. The Recall (R) or Sensitivity signifies the proportion of the correct positives recognized among all the real positive cases (Powers, 2011). It can also be called true positive rate (TPR). High recall means many of the real positive instances of the object-of-interest are correctly detected by the method, which means the number of false negatives is comparatively low. In other words, precision means how many of the retrieved results are truly relevant and recall means the how many of the truly relevant

results are retrieved. Recall and precision vary with the strictness of the method's threshold (McCann, 2011).

In general, recall and precision are inversely related, and a precision-recall (P/R) curve is commonly used to present this relationship in order to indicate the precision-recall performance of an object recognition method (McCann, 2011). It is obtained directly by plotting the precision, p(r) of a method as a function of its recall, r (Zhu, 2004). Thus, it can provide a clear picture of a method's performance towards recognizing objects. However, instead of comparing curves, a single number is often used that characterizes the performance of a method more precisely. This metric is common*ly known as the average precision (AP).* In theory, the average precision is the precision p(r) averaged across all values of recall, i.e. over the interval from r=0 to r=1, in the P/R curve (Zhu, 2004). It is also sometimes referred to as the area under the P/R curve. In practice, it is the approximated sum over the precisions multiplied by the change in recall, at every possible threshold value (McCann, 2011).

The accuracy (ACC) of a recognition method denotes the proportion of the correct real positives and negatives among all the positive and negative cases predicted by the method (JCGM, 2008; Olsen and Delen, 2008). High recognition accuracy means many positive and negative instances detected by the method is correct real positives and negatives respectively, which indicates that the number of false positives and negatives is comparatively low. The F-measure/$F_1$ score is the statistical measure of a test's accuracy (Chinchor, 1992; Powers, 2011). It can be interpreted as a weighted average of the precision and recall, which indicates that both the precision (P) and recall (R) of the test are considered to compute the $F_1$ score (Rijsbergen, 1979). The value of $F_1$ score can

vary between the range of 0 and 1. The traditional F-measure, also known as balanced $F_1$ score, is the harmonic mean of precision and recall (Sasaki, 2007). Table 2-1 summarizes the definitions of the common performance metrics that are used to represent the correctness of object recognition methods.

Table 2-1: Definitions of common performance metrics to evaluate correctness

| **Performance Metrics** | **Definition** |
|---|---|
| Precision (P) | The proportion of the correct real positives among all the positive cases predicted by a method; P = TP / (TP+FP) |
| Recall (R) | The proportion of the correct positives predicted by a method among all the real positive cases; R = TP / (TP+FN) |
| Average Precision (AP) | The precision averaged across all values of recall in the P/R curve; $AP = \int_0^1 p(r)\,dr$ |
| Accuracy (ACC) | The proportion of the correct positives and negatives among all the positive and negative cases predicted by a method; ACC = (TP+TN) / (TP+TN+FP+FN) |
| F-measure ($F_1$) | The weighted average / harmonic mean of the precision and recall; F1 = 2 x precision x recall / (precision + recall) |

### 2.3.2 ROBUSTNESS

The second evaluation criterion that plays a vital role for measuring the performances of the object recognition methods is the robustness. It signifies the degree of tolerance that an object recognition method can undertake against reasonable amount of noise and occlusion in the scene (Pope, 1994). This also includes the invariance of the method

against the change in the environmental illumination conditions (Ulrich and Steger, 2002). A method is considered to be robust if the performance of the recognition method does not degrade significantly when those tolerances are exceeded, i.e. under the cases of noise, occlusions, and illumination variations (Pope, 1994). In particular, occlusion is perceived as a total degradation of a part of the object that is considered for recognition (Caputo, 2004). If significant parts of the object are occluded, it can cause extensive degradation in the performance of the method. As suggested by Caputo (2004), robustness against occlusions can be measured by obtaining the recognition rates under different levels of occlusions in the test images. The recognition rate is then plotted as a function of the level of occlusions in order to compare the robustness performances of different object recognition methods. In general, the robustness decreases as the amount of occlusions increases in the test set (Caputo, 2004).

### 2.3.3 SPEED

Speed is another common criterion to measure the performance of an object recognition method. Typically, it is measured by the computation time that a method takes to complete the recognition tasks. The inverse of the average computation time is considered as the recognition speed of the method. The computation time required by a method strongly depends on the individual implementation procedure of the conforming method. It can remarkably vary for different methods, since these methods work on different principles (Ulrich and Steger, 2002). Though the computation power of modern computers has increased significantly in the last few decades, the necessity of fast and efficient methods is still perceived, specifically when they are adopted for automation in the industrial sectors.

## CHAPTER 3

# OBJECTIVES AND SCOPE

As noted earlier, construction site images contain a lot of information about the construction site activities. The on-site images do not only record the as-built progress of the project under construction, but also capture daily job site activities. If we could automatically retrieve the on-site information from construction site images, it could facilitate us to automate many construction applications. For example, the automated retrieval of on-site information could help us to do the job site planning and co-ordination, to automate the construction productivity analysis, to enhance the safety in construction sites by issuing proactive safety alerts, and to control the quality of construction works etc. This way, the automatic retrieval of construction site information from on-site images can be beneficial to automate many construction management applications.

In order to automatically retrieve the information from construction site images, the first and fundamental step is to automatically recognize the construction operational resources. Currently, there are many object recognition methods available that are mostly developed by the researchers of computer vision (discussed in chapter 2). These methods are widely used for recognizing generic objects in natural scenes. However, the performance of the existing recognition methods for the recognition of construction operational resources is not known, especially considering the fact that the construction sites are typically characterized as dirty, disorderly and cluttered by tools, materials and debris. Moreover, the construction resources in the site images are often captured

with partial occlusions, which make the recognition tasks even more difficult and challenging. For these reasons, it is not clear which method could be effectively and efficiently applied for the recognition of construction operational resources from construction site images.

In order to evaluate the performance of the existing recognition methods, many datasets are developed. Most of these datasets are developed for the purpose of object recognition research at different universities. Although there are many datasets available, these datasets have several limitations (discussed in chapter 2). For example, these datasets only contain limited categories of objects, where construction equipment is not considered. The datasets reflect a small range of variations regarding the pose, camera viewpoint and orientation of the objects. Considering the limitations of the existing datasets, the necessity of developing a new dataset that will cover typical construction equipment images under realistic and diverse site conditions, has been perceived. The newly developed dataset could be used to facilitate the automated recognition of construction equipment from images.

The main objective of this research work is to develop a dataset which covers the images of common construction equipment from different classes, manufacturers, models, sizes and shapes. The dataset also includes such images where construction equipment are captured with various poses, camera viewpoints, occlusions, background clutter and diverse illumination conditions. After the images are collected from the construction sites, they are manually annotated by an annotation tool developed within the scope of this research. The equipment in the images are labeled as "excavator", "loader", “tractor”, “compactor”, “backhoe loader” etc., which establish the ground truth

for construction equipment recognition. The images of the dataset along with the annotations could perform as a common ground to evaluate the construction equipment recognition performance of different recognition methods, from images under realistic site conditions. Thus, the developed dataset could provide a solid foundation to promote automated applications in construction site monitoring by selecting a suitable and appropriate recognition method.

In order to explore the effectiveness of the construction equipment recognition dataset (CERD) developed in this research, the dataset is used to evaluate existing object recognition methods. Two common object recognition methods, developed by Felzenszwalb et al. (2010) and Torralba et al. (2004), are selected for construction equipment performance evaluation since they have shown promising results for the recognition of general objects. The performances of these two methods have been evaluated with the dataset developed in this work, for the recognition of construction equipment from on-site images (discussed elaborately in chapter 5). The method of Felzenszwalb et al. (2010) was built upon the discriminatively trained deformable part models, which demonstrated efficient and successful results on the PASCAL and INRIA person datasets (Felzenszwalb et al., 2010). Consequently, it was recognized by the PASCAL VOC "Lifetime Achievement" Prize in 2010 (Everingham et al. 2010). The method proposed by Torralba et al. (2004) relied on a simple object detector with boosting. The method was implemented successfully for recognizing the objects from the MIT-CSAIL dataset, and the work was awarded the “Best Short Course Prize” at ICCV 2005 (Fei-Fei et al. 2005). Considering the above strengths, these two methods are selected as suitable candidates for the recognition of construction equipment from images under realistic site conditions.

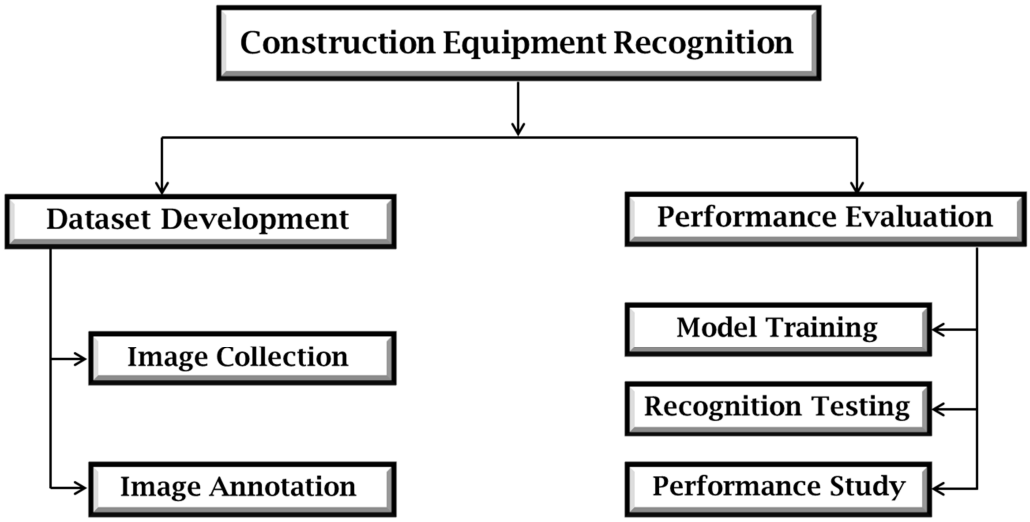



Figure 3-1: Objectives and scope of the current research

Within the scope of this research, the performances of these methods are evaluated based on the recognition results obtained from the recognition tests. A detailed description of the methods' recognition performances is provided in chapter 6. The recognition tests are performed for 5 classes of construction equipment, i.e. excavator, loader, tractor, compactor and backhoe loader. The performances of the methods are then compared on the basis of the performance metrics mentioned before – correctness, robustness and speed – for each class of equipment discretely. Based on the test results, it is found that none of these recognition methods are absolutely perfect with respect to all the performance criteria. However, the recognition tests demonstrate that the existing recognition methods have the potentials to recognize construction equipment using the dataset developed in this work. Based on the results, it is evident that the performances of the existing recognition methods can be evaluated in a standard and extensive way with the developed dataset. Thus, the effectiveness of the dataset is

assessed by the demonstration of the results obtained from the recognition tests conducted in this research. The test results also indicate that the dataset provides an unbiased foundation for comparing the performances of different object recognition methods to recognize construction equipment from construction site images under realistic conditions such as partial occlusions, illumination variations, changes in poses, orientation and camera viewpoints, multiple pieces of equipment working together etc.

# CHAPTER 4

# DATASET DEVELOPMENT FOR CONSTRUCTION EQUIPMENT RECOGNITION

The development process of the Construction Equipment Recognition Dataset (CERD) is explained in this chapter. The dataset is developed in two phases: (1) Image collection and (2) Image annotation. In order to develop a diverse and rich dataset of construction equipment images, many construction sites have been visited, and thousands of construction site images have been collected. The equipment in each image is then annotated to generate the ground truth for the purpose of evaluating the construction equipment recognition performances of existing object recognition methods.

## 4.1 IMAGE COLLECTION

Multiple construction sites are selected as the sources of image collection to develop the construction equipment recognition dataset. Around 2,000 images have been collected from more than 25 construction sites; Figure 4-1 represents the manifestation of image collection from construction sites. In this figure, a Nikon D40 digital SLR camera (Nikon Corporation, Tokyo, Japan) is used to collect the images. After the images are collected from the construction sites, the next step is to assemble them in the dataset. The CERD dataset is formed by organizing the images in an order according to the image collection dates. The images are then specified with such labels that are composed of the name of the dataset along with six digit consecutive numbers, such as CERD_000001, CERD_000002, CERD_000003 etc. All the images in the dataset are stored in the JPEG format, which is the most common format for representing photographic images and permits to select a suitable tradeoff between storage size and image quality.

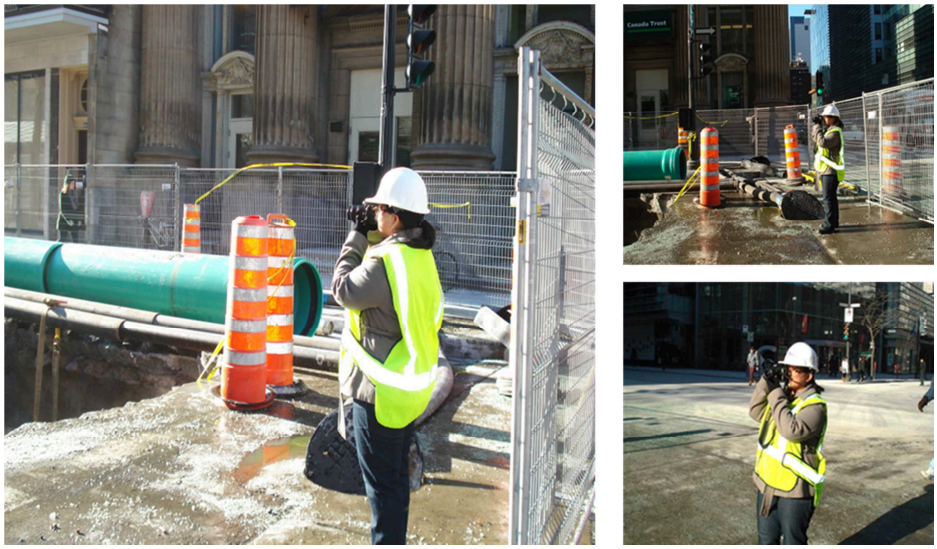


Figure 4-1: Image collection from construction sites

The EOI compiled in the image dataset covers a total of 5 classes of construction equipment, such as excavator, loader, tractor, compactor and backhoe loader, which represent 3 main categories: (1) excavating and lifting (excavator, backhoe loader), (2) loading and hauling (loader, tractor), and (3) compacting and finishing (compactor). For each class of the equipment, hundreds of images were collected in order to ensure a wide range of diversity in terms of size, shape, pose, camera viewpoint, illumination variation, and multiple instances of equipment contained in the same image. Examples of the collected images are illustrated in Figure 4-2.

In order to obtain the images under realistic site characteristics, i.e. dirty, disorderly and cluttered, the images of construction equipment are captured at real construction sites. The images of the dataset offer wide range of variation in different aspects. For example, the images include construction equipment from different manufacturers, e.g. Caterpillar, Volvo, Deere, Komatsu, Hitachi, Case, Kobelco, Kubota etc. Figure 4-3 illustrates the images of the excavator, loader, tractor, compactor and backhoe loader


Figure 4-2: Examples of collected images from the Construction Equipment Recognition Dataset (CERD)

from different manufacturers. Again, within the equipment from the same manufacturer, different models of the equipment are considered in order to obtain manifold sizes and shapes of the corresponding equipment class, such as the images of excavators from the dataset covers the models 336E and 349E for large size; 320E and 324D for medium size; 307D and 314C for small size etc. from the manufacturing company Caterpillar. Examples of different models are shown in Figure 4-4.


Excavator: (a) Hitachi, (b) Deere, (c) Caterpillar, (d) Volvo (from left)

Loader: (a) Volvo, (b) Deere, (c) Caterpillar, (d) Kubota (from left)


Compactor: (a) Bomag, (b) Caterpillar, (c) Dynapac, (d) Volvo (from left)

Tractor: (a) Caterpillar, (b) Hitachi, (c) Deere, (d) Volvo (from left)


Backhoe loader: (a) Deere, (b) Case, (c) Caterpillar, (d) Volvo (from left)

Figure 4-3: Images of Excavator, Loader, Compactor, Tractor, and Backhoe loader from different manufacturers

Large excavator: (a) CAT 336E, (b) CAT 349E, (c) CAT 336E (from left)

Medium excavator: (a) CAT 324D, (b) CAT 320E, (c) CAT 324D (from left)

Small excavator: (a) CAT 307D, (b) CAT 314C, (c) CAT 307D (from left)

Figure 4-4: Images of large, medium and small sized models of excavator from the same manufacturer (Caterpillar)

In addition to the different manufacturers and models, the EOI in the images are also captured in different states, i.e. idle or in operation. Again, the equipment, which is in operation, experiences wide range of pose variations since construction equipment are commonly consisted of multiple articulated and deformable components. Considering the fact that construction equipment typically undergoes drastic pose variation during operation, special attention was placed on capturing the images when they are working.

(a)

(b)

(c)

(d)

(e)

Figure 4-5: Examples of pose and viewpoint variations for (a) Excavator, (b) Loader, (c) Compactor, (d) Tractor, and (e) Backhoe loader

The equipment images are also captured in such a way that the camera viewpoint and orientation of the equipment (towards the camera) change gradually. For example, the images of an EOI are taken from different directions, i.e. front, rear, left, right and all the corners. Figure 4-5 demonstrates some examples of the collected images, where different classes of equipment are captured with pose and viewpoint variations.

During the image collection process, typical characteristics of construction sites have been considered. These include the facts that multiple pieces of the equipment working together; the equipment in the image is partially occluded by another piece of equipment, debris and/or other construction operational resources (i.e. materials, workers etc.). Moreover, different environmental lighting conditions are also reflected in the collected images; i.e. different periods of daytime (morning, noon, afternoon etc.), different sky conditions (sunny or cloudy etc.). Figure 4-6 exhibits some examples of the collected images containing multiple instances of construction equipment within the same image.


Figure 4-6: Examples of pose and viewpoint variations for (a) Excavator, (b) Loader, (c) Compactor, (d) Tractor, and (e) Backhoe loader

Examples of partial occlusions and illumination variations are illustrated in Figure 4-7 and Figure 4-8 respectively.

Figure 4-7: Examples of partially occluded equipment from the construction equipment recognition dataset (CERD)


(a) bright and sunny

(b) dull and cloudy

Figure 4-8: Examples of images with various environmental lighting conditions

The factors that are considered with special attention during the image collection process, are summarized as follows: (1) equipment from different manufactures (Caterpillar, Volvo, Deere, Komatsu, Hitachi, Case, etc.); (2) different models and sizes within the same class of equipment (large, medium and small); (3) variations in equipment states (idle or working); (4) diversity in equipment poses under working

conditions; (5) changes in orientations and camera viewpoints (front, rear, left, right, etc.); and (6) various environmental illumination conditions (different period of day time, different sky conditions); (7) partial occlusion by debris or other construction resources; (8) multiple pieces of equipment working together etc.

## 4.2 IMAGE ANNOTATION

After the collection and compilation of images in the dataset, the annotations of the EOI are performed. To annotate the construction equipment, an annotation tool has been developed based on the work of Korˇc and Schneider (2007) in MATLAB environment. Figure 4-9 represents the in-house built annotation tool, which is specifically designed to annotate the EOI and its different parts. A detailed description of the annotation process for construction equipment is provided in Appendix A.

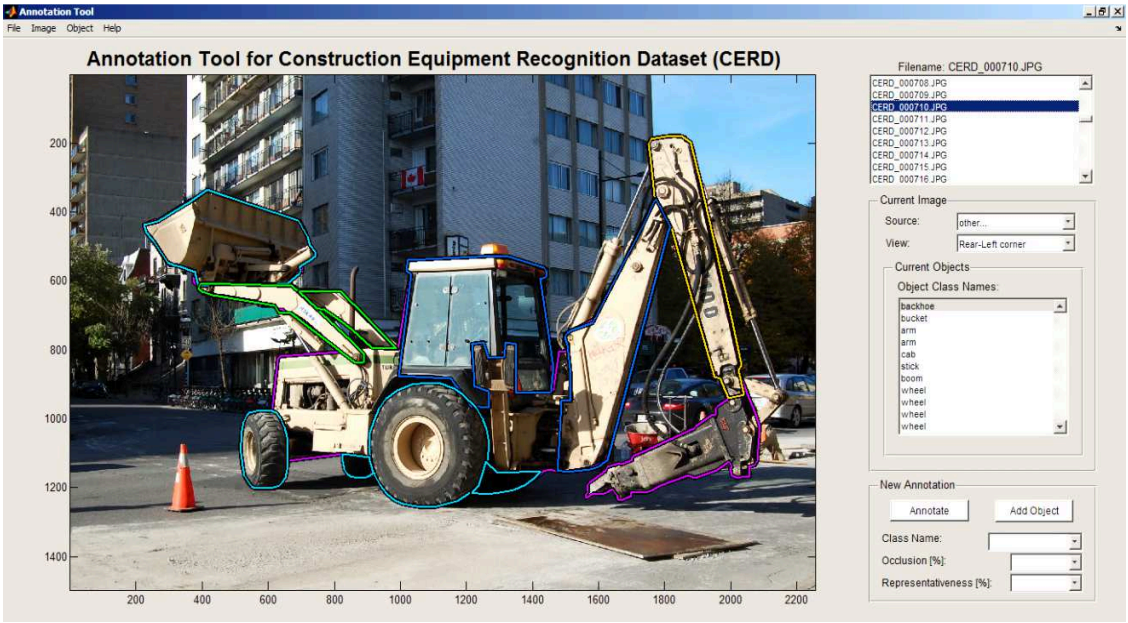



Figure 4-9: Annotation tool for Construction Equipment Recognition Dataset (CERD)

The annotations could provide the answers to the questions like which image from the dataset is being annotated (image name), and what is the resolution (width and height) of the image. Moreover, the information such as the image source, orientation of the equipment towards the camera, the equipment class, and the degree of occlusion and representativeness are provided. The occlusion means the percentage of the equipment that has been visually obstructed by other objects, whereas the representativeness indicates the percentage of the equipment that has not been truncated. For example, 100% representativeness means the entire equipment is visible within the frame of an image. Again, when 40% of the equipment is truncated (i.e. 60% of the equipment is visible in the image), the representativeness is then specified as 60%.

The current annotation tool also provides an option to include the identification of the components for each EOI and establish the relationship of these components with the source equipment by specifying their respective ID values. For instance, a typical excavator is composed of a bucket, stick, boom, cab and tracks (as shown in Figure 4-10). First, the entire equipment is bounded with a polygon and given an ID, 1. Then, its corresponding components are annotated by drawing boundary polygons for each of them individually. These annotations are then assigned the IDs in a hierarchic order, such as 1.1, 1.2, 1.3 etc., which indicate all the components of the excavator (EOI). This way, the IDs for the components correspond to the ID of the source equipment. Hence, the annotation files include the information about the source image and the contained equipment, such as the name of the source image, location, image resolution (height and width), camera viewpoint or orientation, equipment type, identification of the annotated equipment and their corresponding parts, degree of occlusions and representativeness, and the labels of the corresponding equipment components. A

typical annotation file along with the source image is shown in Figure 4-10, where the object IDs of the equipment and the components are highlighted in red circles.

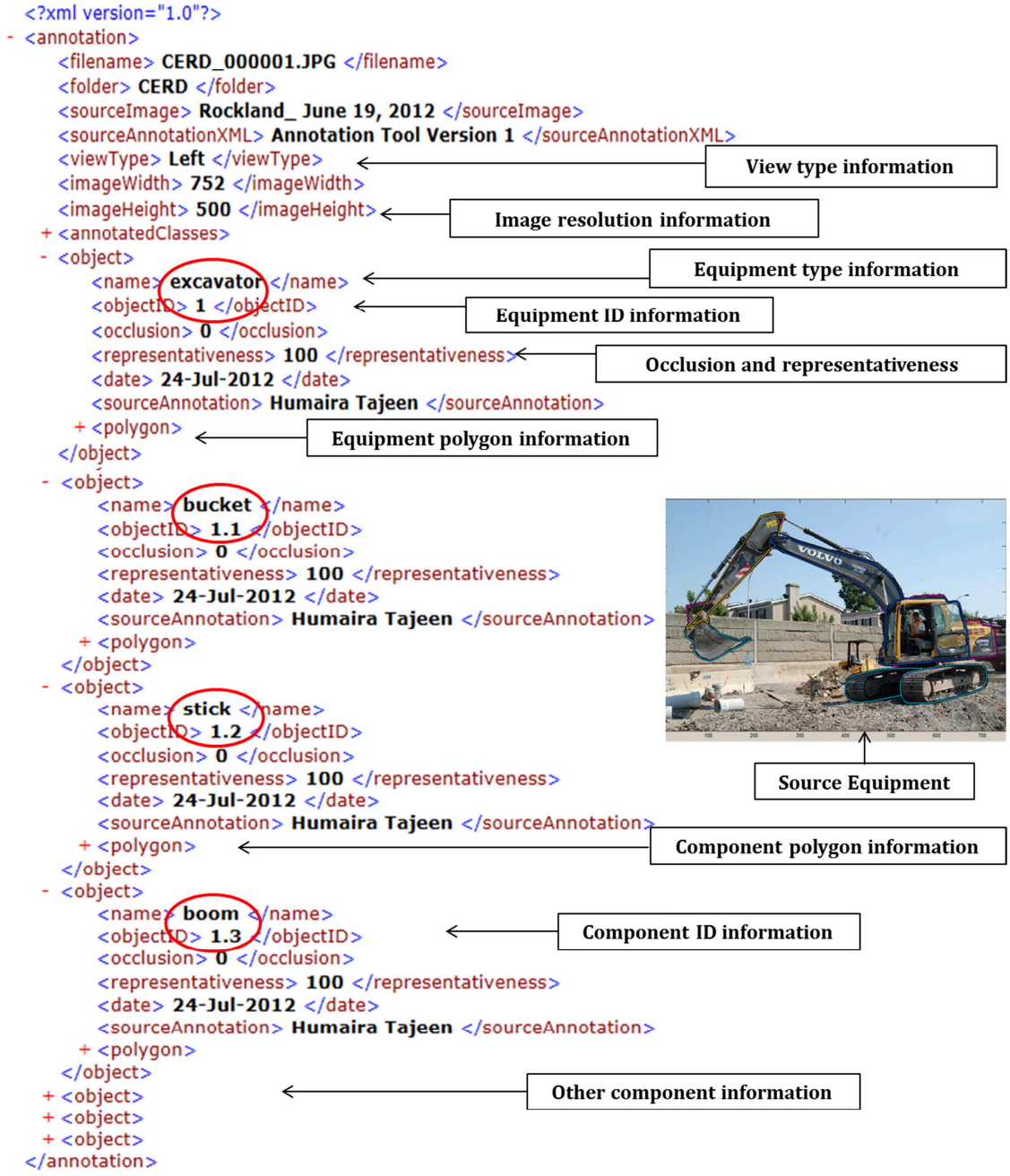



Figure 4-10: Example of an annotation file showing all the annotation information about the source image and the contained equipment


(a) excavator


(b) loader


(c) tractor

(d) compactor


(e) backhoe loader


(f) multiple excavators

Figure 4-11: Annotations of different construction equipment and corresponding parts with polygons

As mentioned earlier, boundary polygons are drawn along the edge of the entire equipment and its components to complete the annotation. A total of 2000 images have been annotated which include all the 5 types of equipment considered in this work.

Different classes of construction equipment are usually composed of different types of equipment parts – the excavator comprises a bucket, stick, boom, cab and tracks; a loader consists of a bucket, arms, cab and wheels; a compactor includes a front compactor, cab and wheels; a tractor contains a blade, cab and tracks; and a backhoe loader comprises bucket, arm, stick, boom, cab, wheels and stabilizer leg. The annotations for the excavator, loader, tractor, compactor, backhoe loader and multiple excavators along with their parts are illustrated in Figure 4-11 a, b, c, d, e and f, respectively.

When all the collected images of the construction equipment are annotated, the dataset is arranged into two folders. One folder contains all the image files in the dataset and the other one stores their corresponding annotation files, in XML format. Each annotation file in the annotation folder reflects its corresponding image file in the image folder, and vice versa. The relationship between an image file and its annotation file is indicated by their file names (Tajeen and Zhu, 2013). An annotation file bears the same name as its image file. For example, when there is an image file 'CERD_000001' in the image folder, there is a corresponding annotation file 'CERD_000001' in the annotation folder, provided that the image is annotated through the annotation tool (Tajeen and Zhu, 2013). The XML files generated by the annotation tool provide the ground truth information that can be used to evaluate the existing object recognition methods for the recognition of construction equipment.

# CHAPTER 5

# CONSTRUCTION EQUIPMENT RECOGNITION TESTS

This chapter is primarily focused on the execution of the recognition tests in order to evaluate the effectiveness of the dataset. Moreover, the construction equipment recognition performances of existing object recognition methods are tested using the dataset developed in the current research. Two object recognition methods have been selected and evaluated for the recognition of construction equipment with the dataset developed in this work. The methods and their implementation, the execution of recognition tests and their results, and the hardware and software configuration of the computer used for the tests are elaborately discussed in this chapter.

## 5.1 SELECTION OF OBJECT RECOGNITION METHODS

The CERD dataset, developed in this work, is used to evaluate the construction equipment recognition performances of the following two object recognition methods: (1) discriminatively trained deformable part-based model method developed by Felzenszwalb et al. (2010) and (2) a simple object detector with boosting method developed by Torralba et al. (2004). The first method has been built upon the discriminatively trained deformable part models. In this method, object models are trained from the training images and represented by mixtures of deformable part models. When the models are created, any given image can be tested for recognition using the models. On the other hand, the second method is developed relying on the simple object detector with boosting. Initially, features from the training images are

precomputed to train the detector, which is then used to recognize construction equipment from the search images.

The method proposed by Felzenszwalb et al. (2010) is commonly used for detecting and localizing objects in images. This method was successfully implemented and achieved state-of-the-art results for recognizing objects from the PASCAL and INRIA person datasets (Felzenszwalb et al., 2010), and the work was awarded the PASCAL VOC "Lifetime Achievement" Prize in 2010. The method proposed by Torralba et al. (2004) is another common object recognition method, which demonstrated efficient results on recognizing objects from the MIT-CSAIL dataset. The work was recognized and received the "Best Short Course Prize" at ICCV 2005 (Fei-Fei et al. 2005). Since both methods have shown promising performances in recognizing the general objects in natural scenes, they have been selected as potential candidates to evaluate the construction equipment recognition performance using the dataset developed in this research.

## 5.2 WORKING PRINCIPLES OF THE SELECTED METHODS

The method proposed by Felzenszwalb et al. (2010) relies on a discriminatively trained, multi-scale, deformable part model for object recognition. This method implies new approaches for discriminative training. The generalization of support vector machines (SVM) is defined to learn a model, which is called as latent SVM (LSVM). A margin-sensitive approach for data mining hard negative examples is combined with LSVM (Felzenszwalb et al., 2010). In this method, a histogram of oriented gradients (HOG) feature pyramid is constructed by computing HOG features from the training images as proposed by Dalal and Triggs (2005). The HOG features are captured at two different

scales. Coarse features are captured by a rigid template covering an entire detection window, whereas finer features are captured by part templates that can be moved with respect to the detection window (Felzenszwalb et al., 2010). In the feature pyramid, the coarse gradients are arranged at the top level and the finer gradients are stored at the bottom level of the feature pyramid (Felzenszwalb et al., 2010).

The method proposed by Torralba et al. (2004) is built on boosting technique for learning. In the boosting technique, several weak classifiers are combined into a final strong classifier. The classifier performs simple discrimination tasks as it uses stumps as weak classifiers, i.e. only lines parallel to the axis. However, stumps are frequently used in object recognition since they can select features efficiently (Torralba et al., 2004). This method implies new approaches for discriminative training. The algorithm is developed on a version of boosting called "gentleboost" since it is simple to implement and numerically robust (Torralba et al., 2004). In this method, a vocabulary of patches is first constructed which is then used to compute the features. Each feature is composed of a template i.e. image patch and a binary spatial mask indicating the region of the image in which the response will be averaged (Torralba et al., 2004).

### 5.3 EVALUATION OF DISCRIMINATIVELY TRAINED PART-BASED MODEL METHOD

The method developed by Felzenszwalb et al. requires the images of equipment and the bounding boxes that indicate the position of the equipment to create the recognition models through supervised training. In order to obtain the bounding box interface, slight conversions to the annotation files have to be made. The construction equipment recognition performance can be tested by using the recognition models generated by

the method for each type of equipment contained in the dataset. The computer used for the experiments (i.e. model training and recognition testing) is Dell Latitude E5430 with Intel® Core-i7-3520M CPU @2.90 GHz and 8.00GB memory, where the operating system was professional edition of 64 bits Windows 7. In addition, MATLAB R2012b was used to run the code for both model training and recognition testing phases.

### 5.3.1 DATASET CONVERSION

The annotation files, created in the form of boundary polygons, are required to be converted in order to meet the input requirements of the method developed by Felzenszwalb et al. (2010). Therefore, original XML files created in the dataset have been modified to meet the input requirements. The main idea of the conversion is to produce the new annotation files that will be legible by the method, and this is performed by retrieving the annotation information contained in the original files (Tajeen and Zhu, 2013). In particular, for each new image annotation file, the information about the image (e.g. file name, image width, image height, etc.), equipment type and view type is directly retrieved from the original annotation file of the dataset and transferred to the new file.

In order to generate the bounding box of the equipment, the information of the polygons is first extracted, and the coordinates of the polygon points are compared with each other. From this comparison, the maximum and minimum values of x and y coordinates are obtained from the polygon points. The procedure is schematically shown in Appendix B. Thus, the top-left and bottom-right corners of the bounding box are determined, and the bounding box is created. Separate bounding boxes are formed

for each of the equipment contained in an image. An example of the annotation information conversion results is illustrated in Figure 5-1. The left part of the image shows the annotation information in XML format with the bounding box interface, which is produced on the basis of the annotation information contained in the original annotation file with the polygon interface. The right part of Figure 5-1 shows the bounding box for the equipment in the image.

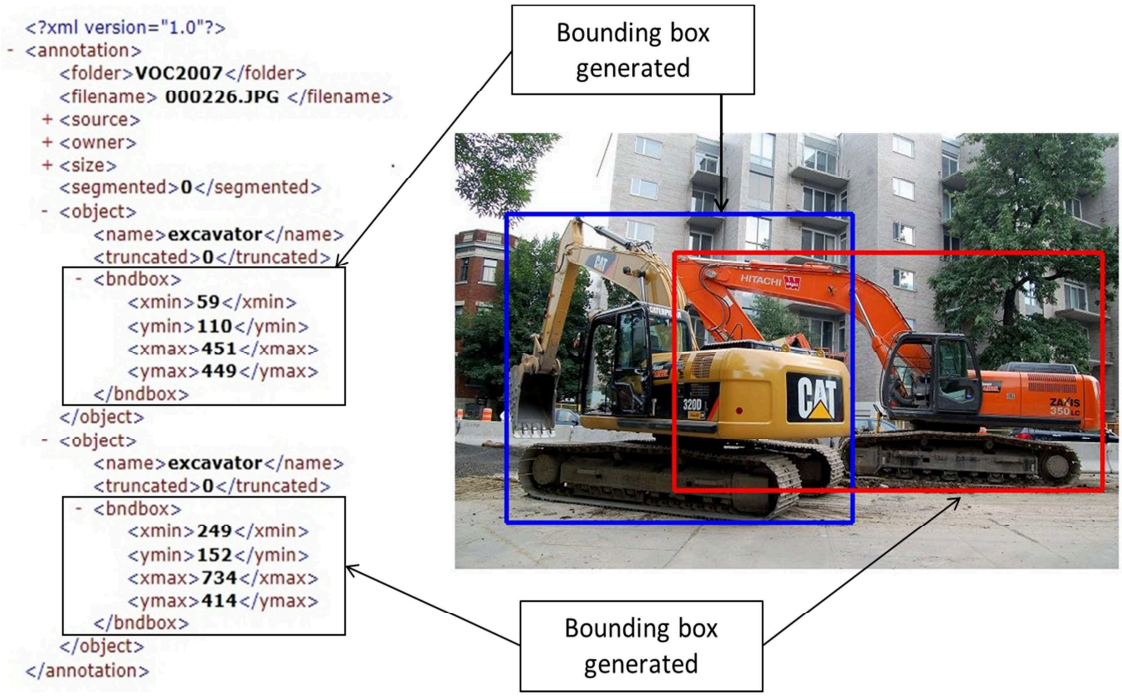
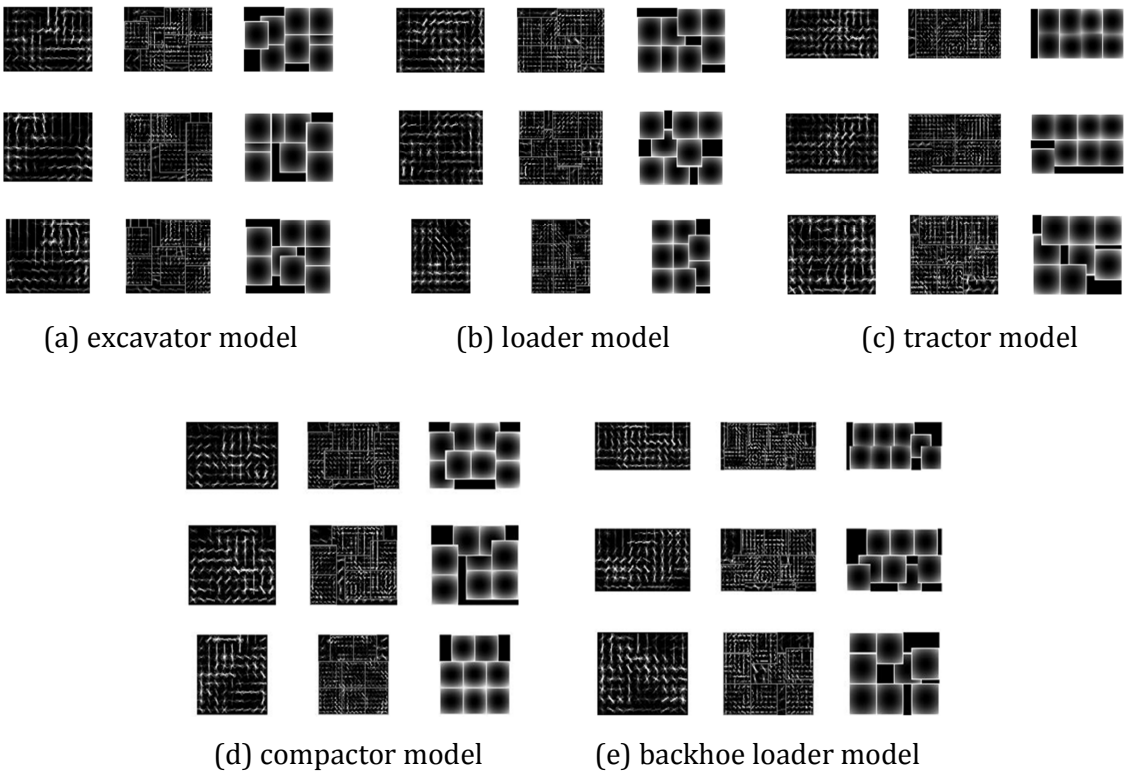



Figure 5-1: Conversion of annotation information

### 5.3.2 RECOGNITION MODELS TRAINING

After the annotation files are converted to the bounding box interface, the dataset can be used to train the recognition models for construction equipment. The method developed by Felzenszwalb et al. is a complete learning-based system, which trains

object models using a discriminative method (Felzenszwalb et al., 2010). The recognition models are separately trained for different classes of equipment such as excavator, loader, tractor etc. A total of 800 images were used for each class of equipment for the purpose of training and testing. Among these, 300 images contained positive instances of the EOI. The rest 500 images included negative instances, which means they contained other objects except the EOI to be trained and tested. Among the 300 images with positive instances, the models were trained by using 200 images and then the recognition test was performed using the rest 100 images as search image. Figure 5-2 shows the examples of the recognition models for all types of equipment, which are trained by the method using the dataset developed in this research.



(a) excavator model (b) loader model (c) tractor model

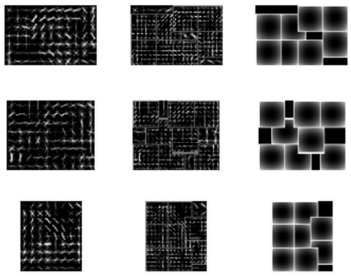
(d) compactor model (e) backhoe loader model

Figure 5-2: Recognition models trained with the dataset, for different equipment class – (a) excavator, (b) loader, (c) tractor, (d) compactor, and (e) backhoe loader

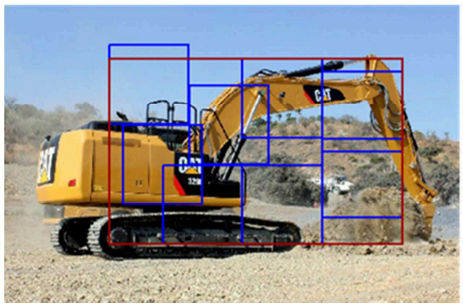
### 5.3.3 RECOGNITION TESTS

When the recognition models are generated, they can be used to recognize the construction equipment from any given test image. Figure 5-3 exhibits the steps involved in the recognition process of an excavator, using the recognition model trained by the method of Felzenszwalb et al. (2010). In the first step, the method reads the input image; the model is visualized in the second step. In the third step, the detection is performed by comparing the visual features of the model and the object in the input image. When satisfactory level of features are matched, the object in the image is detected by displaying multiple detection windows covering the equipment (blue boxes in Figure 5-3c). In the final step of the recognition process, a bounding box is generated to cover the entire equipment (red box in Figure 5-3d).

(a) Input image



(b) Model visualization

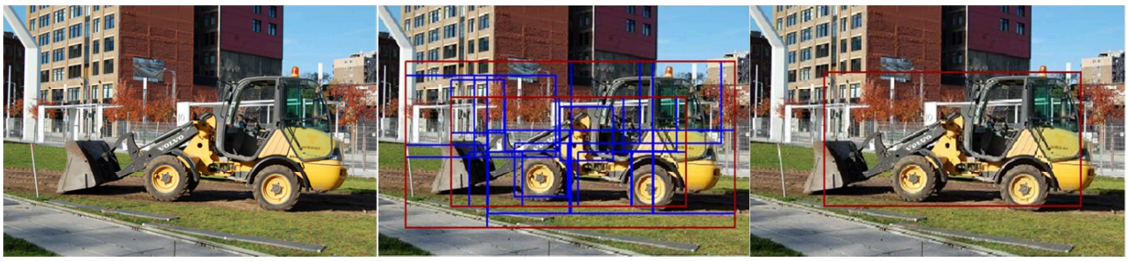


(c) Detections

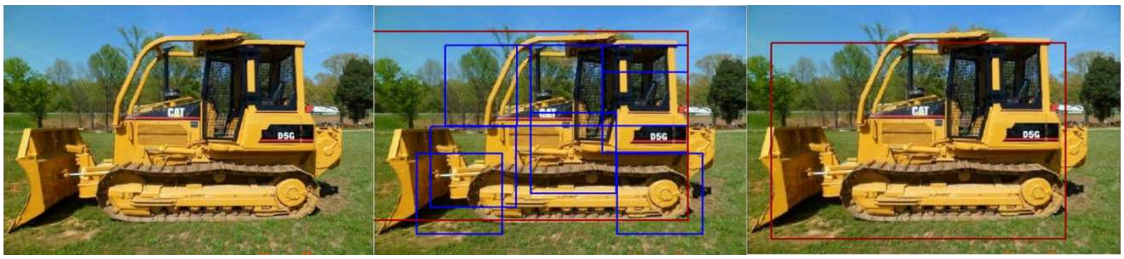
(d) Bounding box

Figure 5-3: Recognition test of an excavator by the method of Felzenszwalb et al. (2010), showing different steps involved in the recognition process

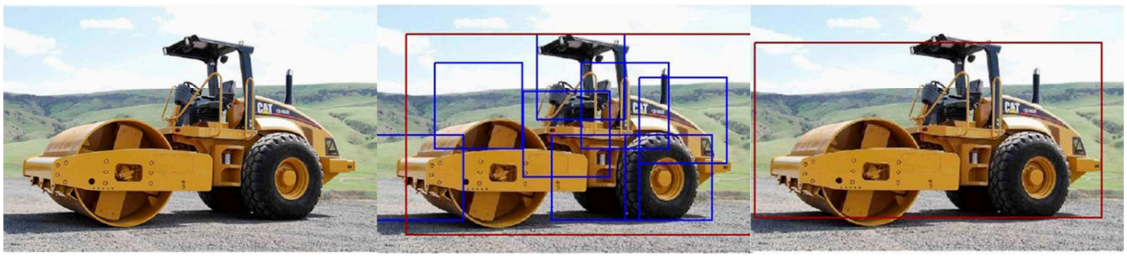
Figure 5-4 shows the examples of the recognition results for other classes of equipment, generated by the discriminatively trained part-based model method. Figure 5-4 (a), (b), (c), and (d) depict the recognized equipment, such as loader, tractor, compactor and backhoe loader, respectively. Nevertheless, the recognition tests are performed by the trained models based on the construction equipment dataset developed in this work.



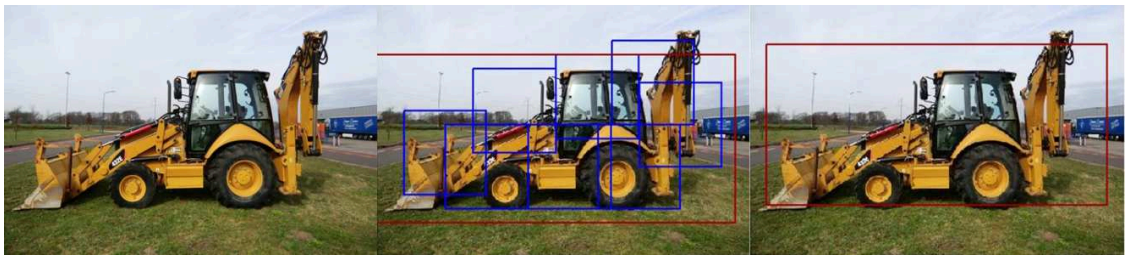
(a) Loader recognition test



(b) Tractor recognition test



(c) Compactor recognition test



(d) Backhoe loader recognition test

Figure 5-4: Recognition tests by the method of Felzenszwalb et al. (2010)

## 5.4 EVALUATION OF SIMPLE OBJECT DETECTOR WITH BOOSTING METHOD

The method proposed by Torralba et al. (2004) requires the input of a set of training images and the bounding polygons, showing the positions of the equipment in the images. A vocabulary of patches is first created, which is then used to compute features from the training images. Thus, a detector is trained by the gentle boosting method. The detector is then used to recognize construction equipment from the query images with ground truth annotations of the EOI. The computer configuration and operating system used for the experiments (i.e. detector training and recognition testing) was the same as it was used for testing the method of Felzenswalb et al. (2010). The code for this method was run in MATLAB R2012b, for the purpose of both training and testing.

### 5.4.1 DETECTORS TRAINING

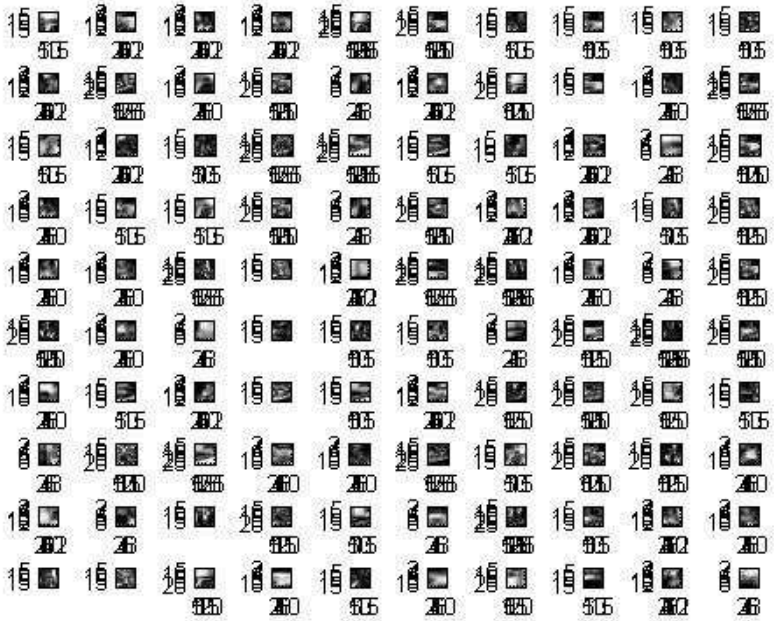
In the simple object detector with boosting method developed by Torralba et al. (2004), the annotation files from the dataset can be used directly with the boundary polygon interface. The images and annotations are used to create a detector, which is later used by the query tools to test an image. This method employs the LabelMe toolbox with its numerous utility functions to train the detectors and to test the search images. Initially, the method reads the images and their corresponding annotation files to create the training and test database as shown in Figure 5-5. When the database is created, the training process begins with the formation of a dictionary of filtered patches that are extracted from the target EOI (Figure 5-6a). The number of images that is used to create the dictionary is specified by the user.

Figure 5-5: Training and test database created by the simple object detector with boosting method with the images of construction equipment

A set of 300 images were used for the purpose of training and testing for each class of equipment. Among these, the detectors were trained by using 200 images and then the recognition tests were performed using the rest 100 images as search image. All the images contained positive instances of the particular EOI. The features of the target equipment from all the training images are precomputed, and the feature outputs are stored at the center of the equipment in an image (Figure 5-6b). Moreover, a number of negative samples are extracted from scattered background locations of the training images, where the EOI is typically located in the foreground (Torralba et al. 2004). Thus, the detector for the target equipment class is trained, which acts as the strong classifier

during the recognition process. Besides these strong classifiers, a number of weak detectors are also trained by the method (Torralba et al., 2004). The detectors are trained separately for each class of construction equipment.

(a)



(b)

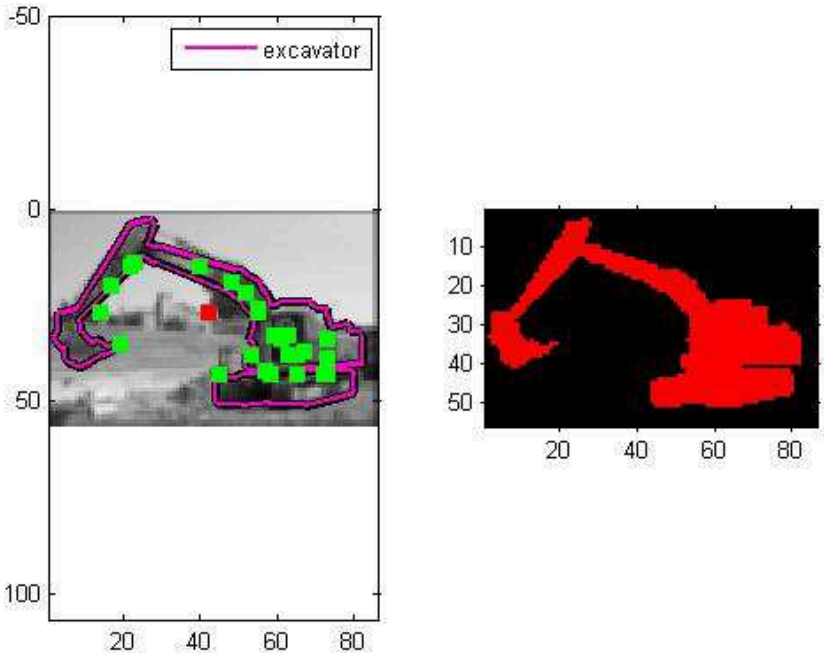
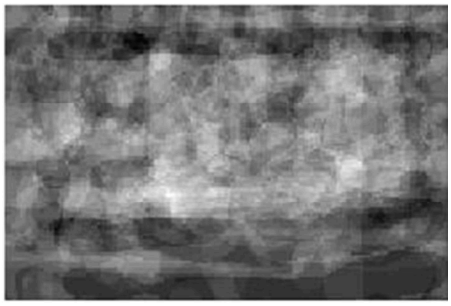



Figure 5-6: (a) Dictionary of filtered patches created from the target EOI, (b) Precomputed features stored at the center of the EOI

### 5.4.2 RECOGNITION TESTS

When the detectors are trained for each class of construction equipment, the recognition tests can be performed by using these detectors. In this method, both the strong classifier and weak detectors are used for recognition. The strong classifier is used to recognize the EOI from an image. When the EOI is recognized by the detector, i.e. strong classifier, the bounding boxes and scores are obtained as output. The weak detectors are further used to confirm the presence of the EOI by matching templates from the detector and the EOI in the test image. Figure 5-7 exhibits different steps for the recognition of an excavator, using the detector trained by the method of Torralba et al. (2004).

(a) Input image with ground truth



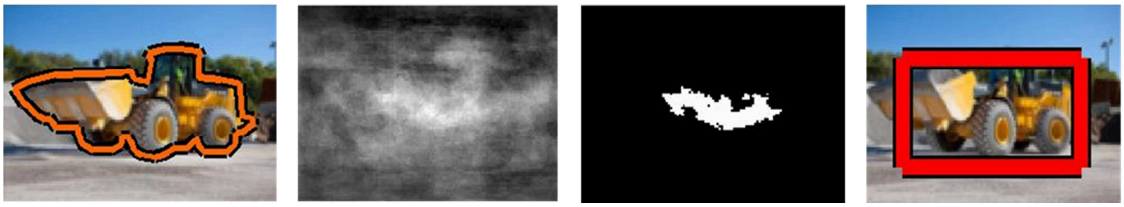
(b) Boosting margin

(c) Thresholded output

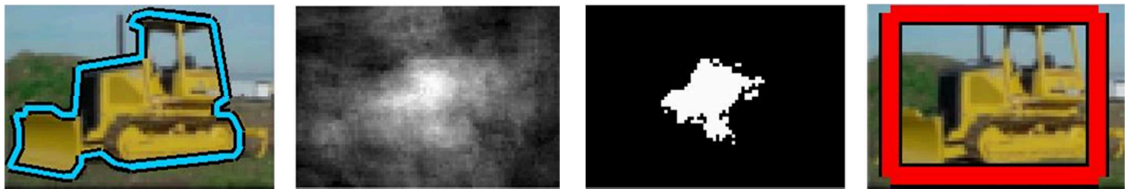
(d) Detector output
Targets=2, correct=2, false alarm=1

Figure 5-7: Recognition test of an excavator by the method of Torralba et al. (2004), showing different steps involved in the recognition process

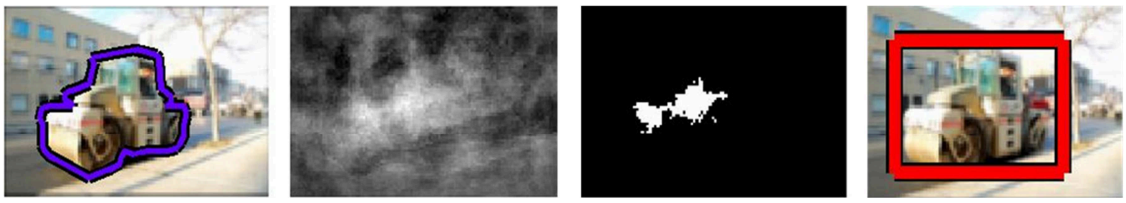
In the first step, the method reads the input image of the EOI with the ground truth annotation from the database (Figure 5-7a). Then it employs the boosting margin to recognize the EOI, using the detector obtained from the training phase (Figure 5-7b). In the third step, the thresholded output for the recognition is displayed (Figure 5-7c). The recognition is performed by comparing the features of the detector and the EOI in the input image. When the number of matched features reaches a satisfactory level defined by the method, the EOI in the image is recognized. The final step provides the detector output, which includes the number of the target EOI in the test image, correctly recognized EOI (red boxes in Figure 5-8) and also the false detections.



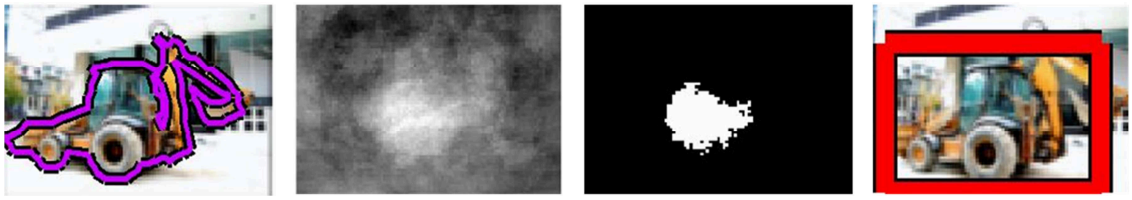
(a) Loader recognition



(b) Tractor recognition



(c) Compactor recognition



(d) Backhoe loader recognition

Figure 5-8: Recognition tests by the method of Torralba et al. (2004)

Figure 5-8 (a), (b), (c), and (d) show the examples of the recognition tests for loader, tractor, compactor and backhoe loader respectively, using the detectors trained by the gentle boosting method. More examples of recognized construction equipment are presented in Figure 5-9. Figure 5-9 (a) and (b) represent the recognition results for different types of construction equipment, achieved by applying the methods of Felzenswalb et al. (2010) and Torralba et al. (2004), respectively.

(a) Recognition of construction equipment by the method of Felzenswalb et al. (2010)

(b) Recognition of construction equipment by the method of Torralba et al. (2004)

Figure 5-9: Recognition results for different types of construction equipment

# CHAPTER 6

# RESULTS AND DISCUSSION

This chapter concentrates on the comparison of the performances of the two methods, the methods developed by Felzenszwalb et al. (2010) and Torralba et al. (2004), for the recognition of construction equipment from images. The performances are measured on the basis of the recognition tests as discussed in the previous chapter. In order to compare the performances of both the methods, a set of 300 images were used for the purpose of recognition – training and testing – for each class of equipment (discussed in chapter 5). The recognition tests were performed separately for each equipment class by using their respective recognition models/detectors. Based on the recognition results, the common performance metrics (stated in chapter 2) – correctness, robustness and speed – have been used to evaluate the performances of the methods on the recognition of construction equipment. The aforementioned performance metrics are discussed in details in the following sections.

## 6.1 CORRECTNESS

The correctness of the recognition methods is measured by calculating the values of precision and recall, average precision, accuracy, and $F_1$ score using the equations summarized in Table 2-1. The precision-recall (P/R) curves of the methods are additionally plotted to compare their performances. The recognition tests for the methods were performed at different threshold levels. During the recognition process, multiple precision and recall are obtained for different images. Moreover, the precision and recall differ considerably with the changes in threshold levels. Based on the

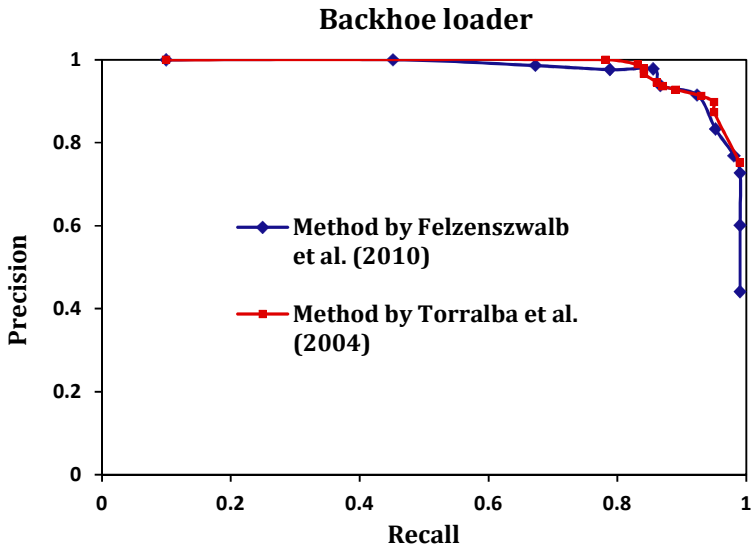
recognition results, the numbers of true positive, false positive and false negative for each of the query images are recorded manually to calculate the precision and recall with various threshold values. Table 6-1 shows one part of the precision-recall results, considering 20 images chosen from the 100 test images, for the recognition of compactor using the method developed by Felzenszwalb et al. (2010).

Table 6-1: Calculation of precision and recall for construction equipment recognition

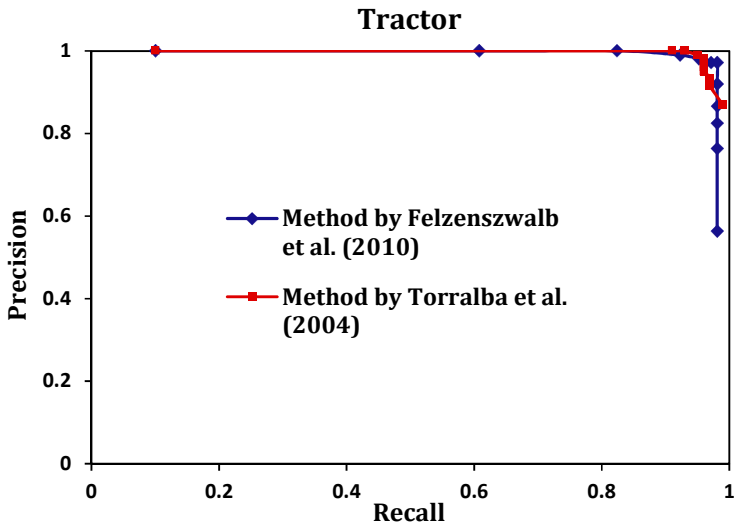
| No. | TP | FP | FN | Precision % | Recall % |
|---|---|---|---|---|---|
| | | | | TP/(TP+FP) | TP/(TP+FN) |
| 1 | 1 | 2 | 0 | 33.33 | 100.00 |
| 2 | 1 | 0 | 0 | 100.00 | 100.00 |
| 3 | 1 | 0 | 0 | 100.00 | 100.00 |
| 4 | 1 | 0 | 0 | 100.00 | 100.00 |
| 5 | 1 | 1 | 0 | 50.00 | 100.00 |
| 6 | 1 | 1 | 0 | 50.00 | 100.00 |
| 7 | 1 | 1 | 0 | 50.00 | 100.00 |
| 8 | 1 | 1 | 0 | 50.00 | 100.00 |
| 9 | 1 | 1 | 0 | 50.00 | 100.00 |
| 10 | 1 | 0 | 0 | 100.00 | 100.00 |
| 11 | 1 | 0 | 0 | 100.00 | 100.00 |
| 12 | 1 | 0 | 0 | 100.00 | 100.00 |
| 13 | 1 | 0 | 0 | 100.00 | 100.00 |
| 14 | 2 | 1 | 0 | 66.67 | 100.00 |
| 15 | 1 | 0 | 1 | 100.00 | 50.00 |
| 16 | 1 | 0 | 0 | 100.00 | 100.00 |
| 17 | 1 | 0 | 0 | 100.00 | 100.00 |
| 18 | 1 | 0 | 0 | 100.00 | 100.00 |
| 19 | 1 | 0 | 0 | 100.00 | 100.00 |
| 20 | 1 | 1 | 0 | 50.00 | 100.00 |
| **SUM** | **21** | **9** | **1** | **70.00** | **95.45** |

After the calculation of the precision-recall values at different threshold settings, the P/R curves can be obtained by plotting these values. The performance of a recognition method can be considered higher for the larger area under the plotted P/R curve. This way, the comparison of the recognition performances can be obtained from the P/R curves. Figure 6-1 shows the P/R curves of both the methods for the recognition of backhoe loader, tractor, excavator, loader and compactor.

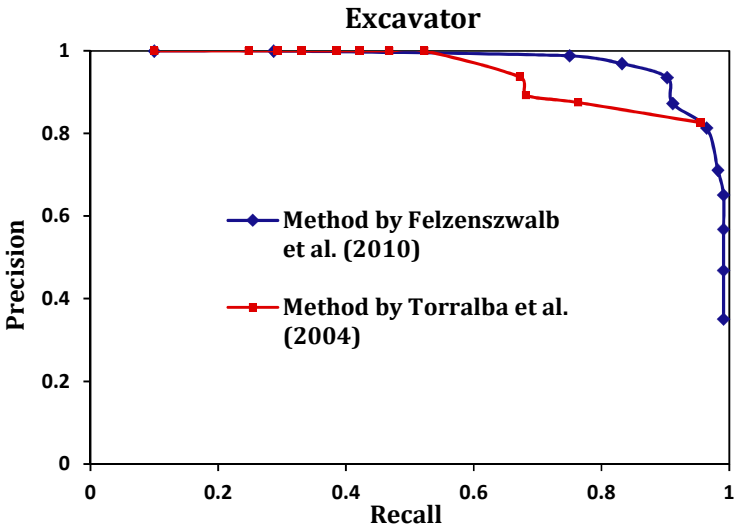
(a)


Backhoe loader
Precision
Recall
Method by Felzenszwalb et al. (2010)
Method by Torralba et al. (2004)
0
0.2
0.4
0.6
0.8
1

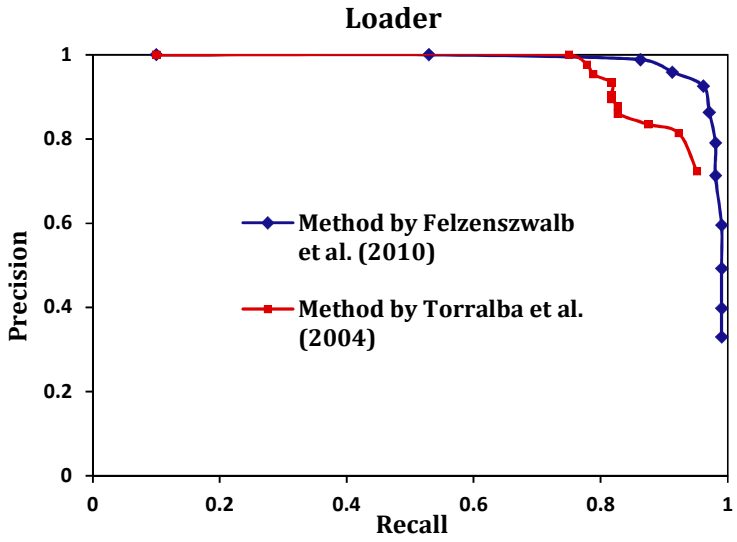
(b)


Tractor
Precision
Recall
Method by Felzenszwalb et al. (2010)
Method by Torralba et al. (2004)
0
0.2
0.4
0.6
0.8
1

(c)


Excavator
Precision
Recall
Method by Felzenszwalb et al. (2010)
Method by Torralba et al. (2004)
0
0.2
0.4
0.6
0.8
1

(d)




(e)

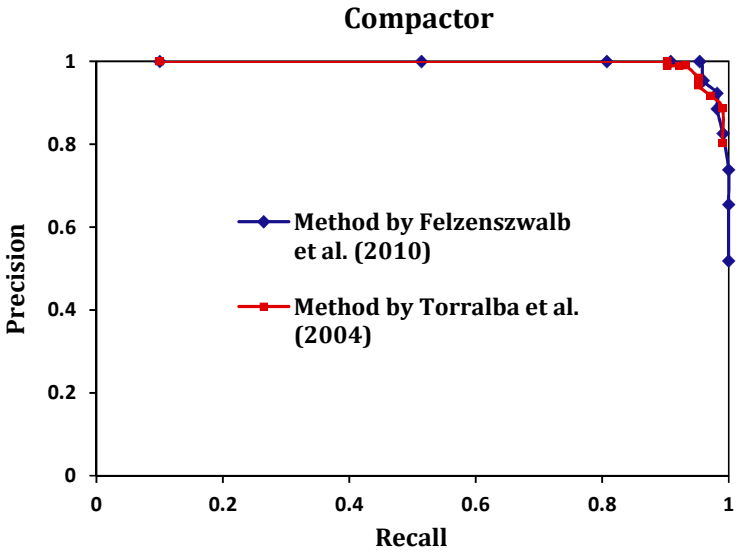



Figure 6-1: P/R curves for the recognition of (a) backhoe loader, (b) tractor, (c) excavator, (d) loader, and (e) compactor

The average precision (AP) is also obtained from the P/R curves. It is calculated as the sum of the precision multiplied by the change in recall at different threshold settings (Zhu, 2004). An example of the calculation procedure of AP is shown in Appendix C, where the APs for both the methods are calculated from the P/R curves of loader

recognition. However, the results for each class of construction equipment have been shown and compared in Figure 6-2. It can be seen that the method of Felzenszwalb et al. (2010) achieved 0.9637, 0.9785, 0.9594, 0.9738 and 0.994 for the recognition of backhoe loader, tractor, excavator, loader and compactor respectively. The average value was 0.9738 for all equipment classes. On the orther hand, , the method of Torralba et al. (2004) attained 0.9707, 0.9863, 0.9007, 0.9227 and 0.9853 for the recognition of backhoe loader, tractor, excavator, loader and compactor respectively, with the average value of 0.9531.

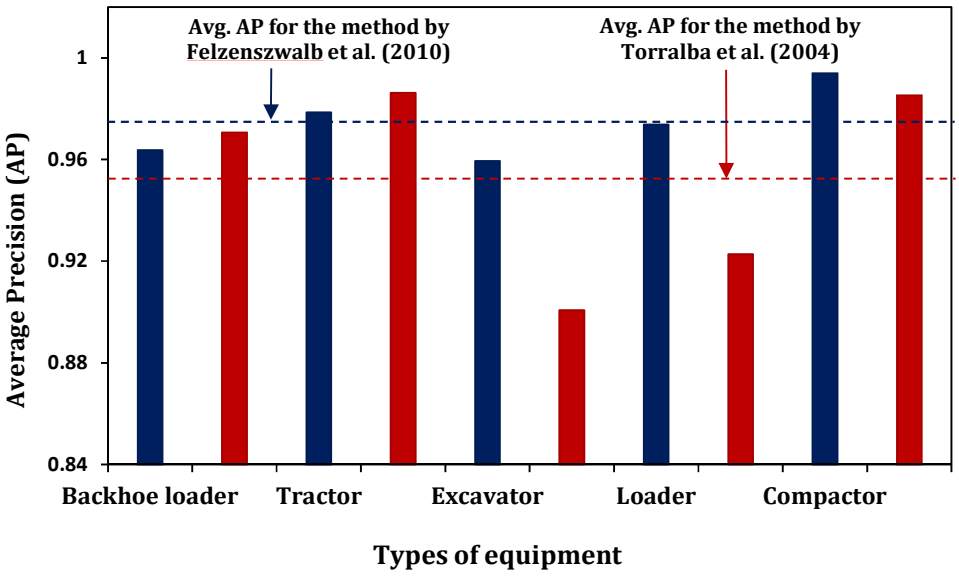



Figure 6-2: Comparison of average precision (AP)

The accuracy and $F_1$ score of the two methods are calculated for all equipment classes to obtain the average values of accuracy and $F_1$ score. Figure 6-3 illustrates the comparison of these values for different equipment classes. It is found that for the recognition of backhoe loader, tractor, excavator, loader and compactor, the method of Felzenszwalb et al. (2010) achieved the accuracy of 0.8135, 0.9175, 0.7589, 0.7617 and

0.8929; and the method of Torralba et al. (2004) accomplished the accuracy of 0.8165, 0.9123, 0.552, 0.7391 and 0.8827. The average accuracy for all the equipment classes is 0.8289 and 0.7805 for the methods of Felzenszwalb et al. (2010) and Torralba et al. (2004), respectively. The average values of $F_1$ score demonstrate that the method of Felzenszwalb et al. (2010) attained 0.8968, 0.9566, 0.8609, 0.8605 and 0.9428 for the recognition of backhoe loader, tractor, excavator, loader and compactor respectively, with the average of 0.9035 for all the equipment classes. On the other hand, the method of Torralba et al. (2004) reached the $F_1$ score of 0.8985, 0.9539, 0.6475, 0.8498 and 0.937 for the recognition of backhoe loader, tractor, excavator, loader and compactor, respectively. The average $F_1$ score for all classes of construction equipment was 0.8573.

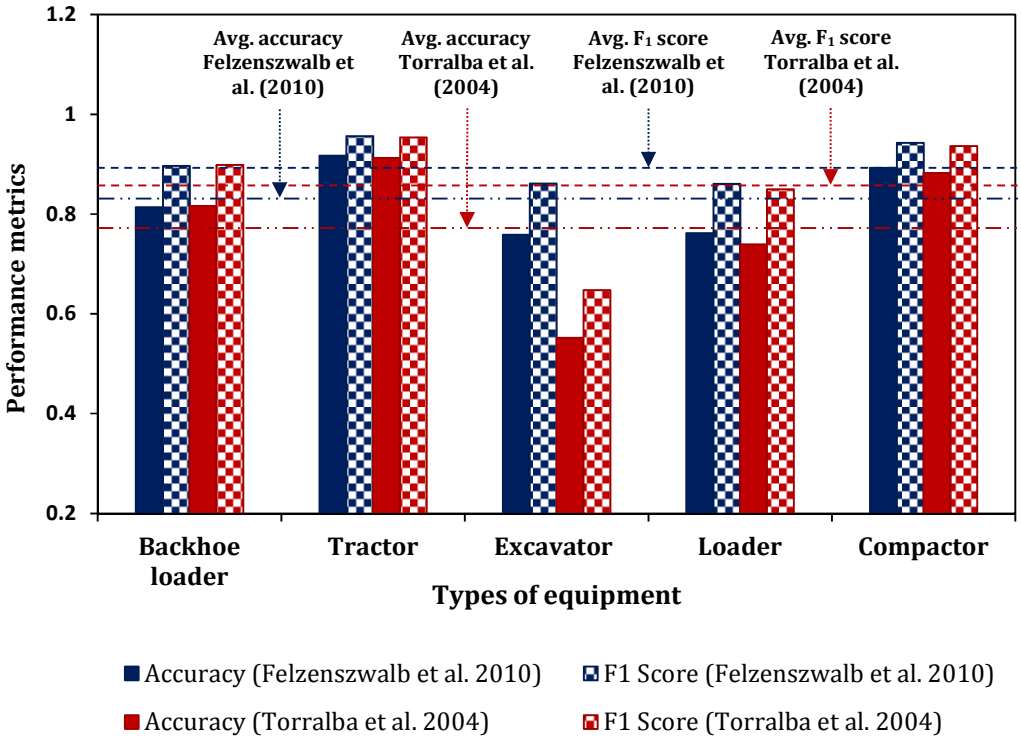
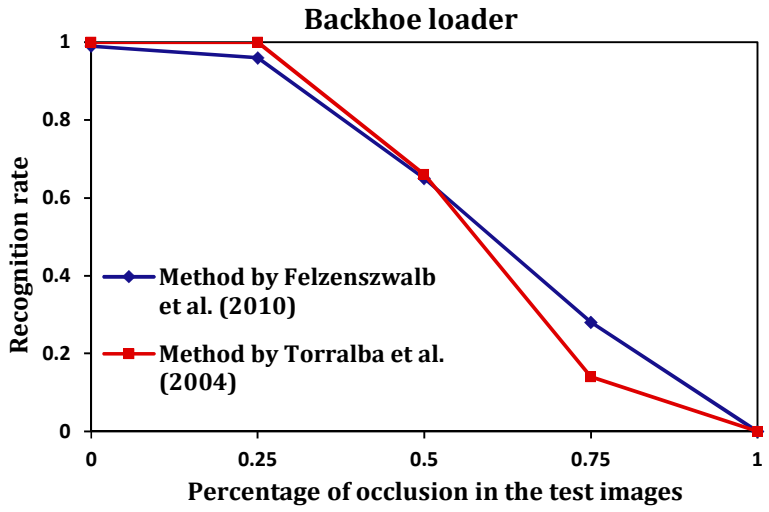



Figure 6-3: Comparison of accuracy and F1 score

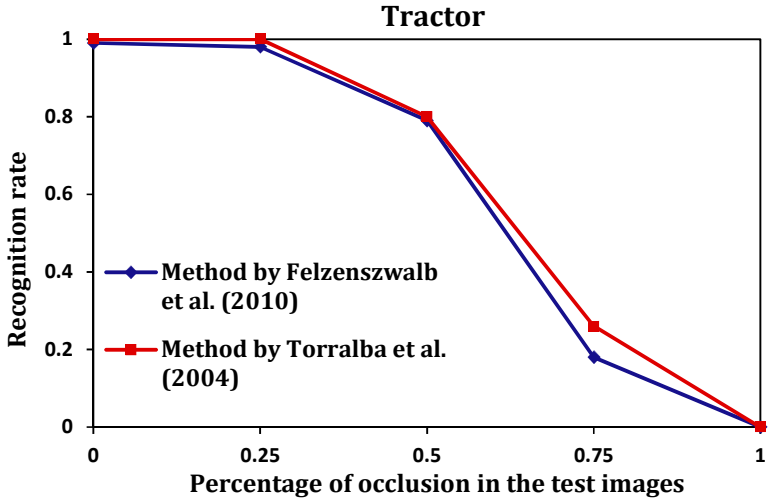
## 6.2 ROBUSTNESS

In order to measure the robustness of the object recognition methods against occlusions, the testing procedure mentioned by Caputo (2004) is adopted here. Specifically, the images in the dataset are first sorted based on the level of occlusions of the EOI indicated in their annotation files. The levels of the occlusions in these images vary from 0% to 100%. Then the recognition tests are performed and the recognition rates are calculated at each level of the occlusions. This way, the recognition rate could be represented as a function of the level of the occlusions in the test images.

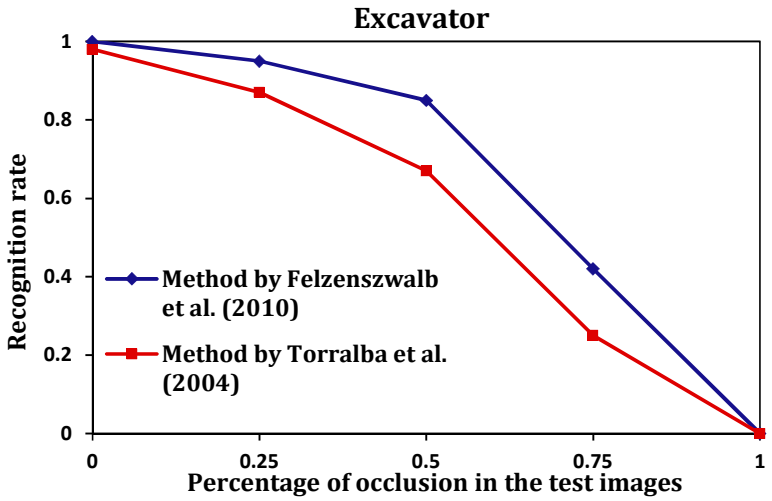
(a)



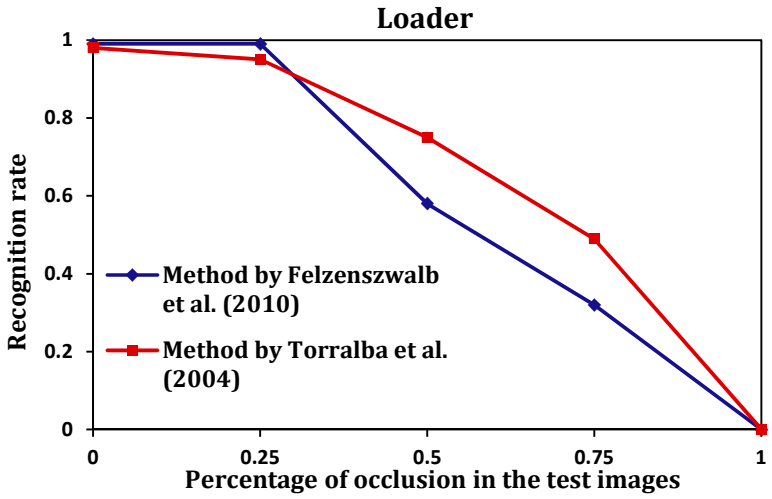

(b)



(c)




(d)




(e)

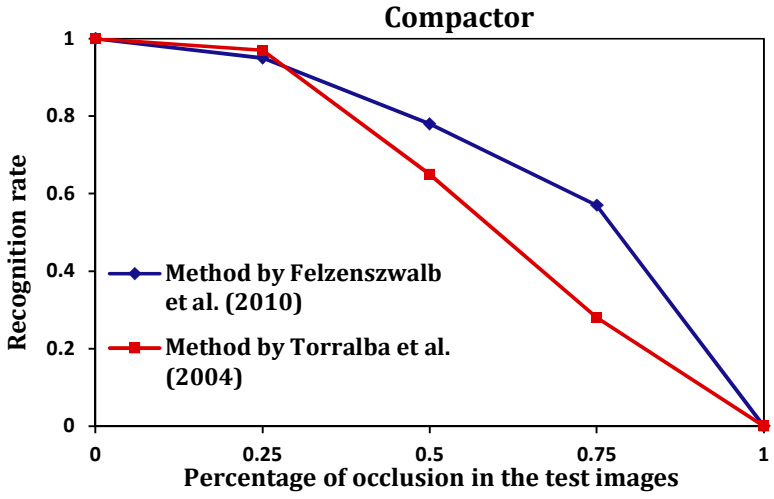



Figure 6-4: Comparison of robustness against occlusions

Figure 6-4 shows the changes of the recognition rate at different levels of the occlusions for different equipment classes. The recognition tests were performed to evaluate the recognition rates – recall – with a sub set of images from the dataset that contained an increasing level of occluded views, such as 25%, 50%, 75% etc. From these results, it is evident that the method of Felzenszwalb et al. (2010) is robust for the recognition of backhoe loader, excavator and compactor. Specifically, the aforementioned method performed significantly well for the recognition of excavator with different levels of occlusions. However, the method of Torralba et al. (2004) showed higher performance for the recognition of tractor and loader with partially occluded views.

### 6.3 SPEED

In order to compare the performances of the methods regarding recognition speed, the times taken by both the methods for the recognition of construction equipment at different threshold settings are recorded and the average is obtained. The method, which needs the less computation time, has the higher recognition speed. For the comparison purpose, the test images were all fixed at the resolution of 375 x 250 pixels. Based on the results, it is found that the method of Felzenszwalb et al. (2010) required 18.86 seconds, 16.38 seconds, 20.64 seconds, 19.07 seconds and 18.05 seconds for the recognition of backhoe loader, tractor, excavator, loader and compactor, respectively. The average computation time was 18.6 seconds. In contrast, the method of Torralba et al. (2004) required 1.75 seconds, 1.45 seconds, 1.71 seconds, 1.98 seconds and 1.66 seconds for the recognition of the aforementioned equipment classes respectively, with the average of 1.71 seconds. Figure 6–5 depicts the comparison of the computation time required for construction equipment recognition with one standard error.

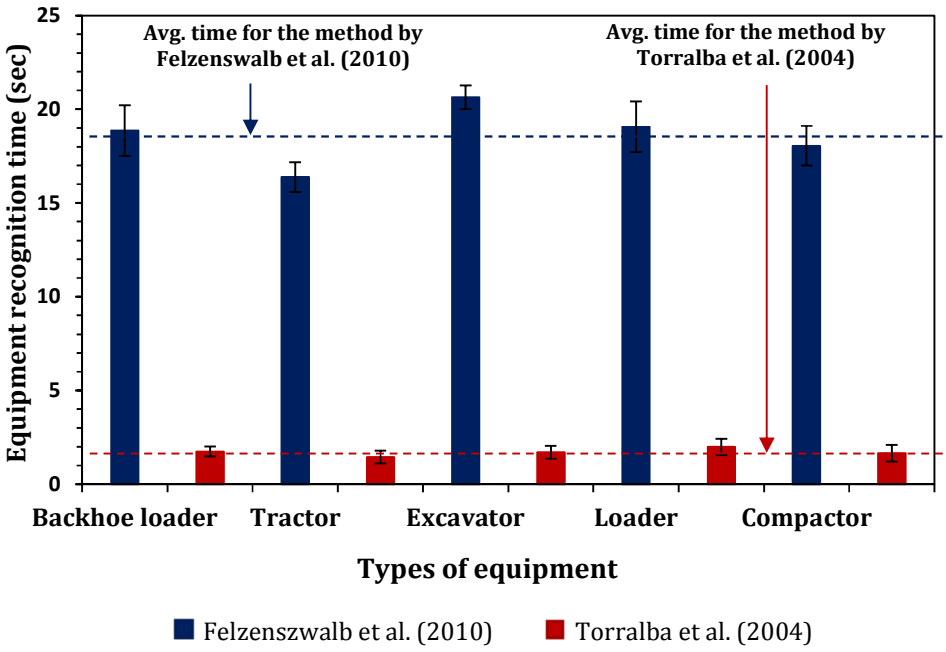
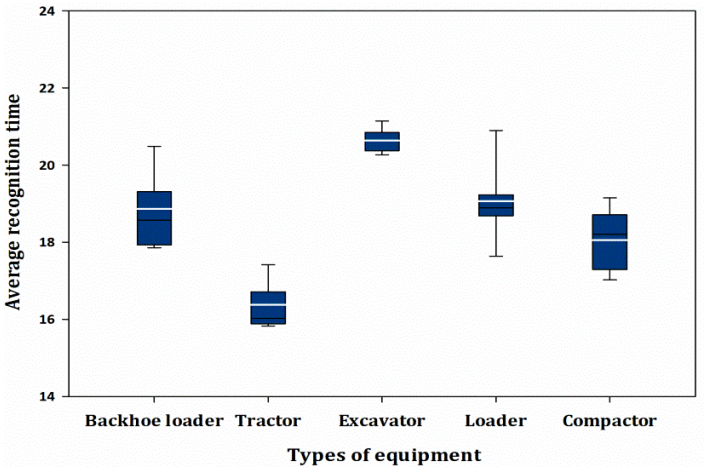



Figure 6-5: Comparison of computation time required for construction equipment recognition with one standard error

The computation times were measured for each class of construction equipment with different thresholds to obtain statistically significant results. Figure 6–6 shows the box and whisker plots of data from the time requirements for recognizing backhoe loader, tractor, excavator, loader and compactor. Figure 6–6 (a) and (b) represent the results of the methods by Felzenszwalb et al. (2010) and Torralba et al. (2004), respectively. The upper and lower boundaries of the box indicate upper (75th percentile) and lower (25$^{th}$ percentile) quartile, whereas the internal black line indicates the median data (computation time), and the thick white line represents the mean value (computation time). The lines extending vertically from the boxes, known as whiskers, illustrate the variability outside the upper and lower quartiles. The size of the box and the spacing between the different parts of it indicate the dispersion/spread of the measured data.

(a)




(b)

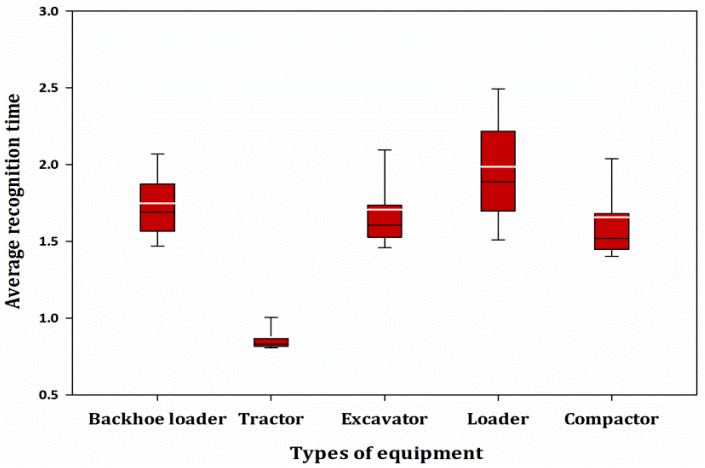



Figure 6-6: Measurement of computation time for construction equipment recognition: (a) method by Felzenszwalb et al. (2010), (b) method by Torralba et al. (2004)

## 6.4 DISCUSSION

The P/R curves of both the methods for all 5 classes of construction equipment show that they have common characteristics with respect to the variations in precision and recall for varying thresholds. The precision increases with the decrease of recall values. According to the P/R curves for the recognition of backhoe loader, tractor and compactor (Figure 6-1 a, b and e), it could be seen that the recognition performances of

both methods are similar. However, the P/R curves for the recognition of excavator and loader (Figure 6-1 c and d) show the performance discrepancies between these two methods. The method proposed by Felzenszwalb et al. (2010) performs better than the method proposed by Torralba et al. (2004) for the recognition of excavator and loader. One potential reason lies in the fact that the excavators and loaders may produce relatively drastic pose variations, when they are in operation. Figure 6–7 shows an example of the changing behavior of precision and recall values with the increase in threshold level.

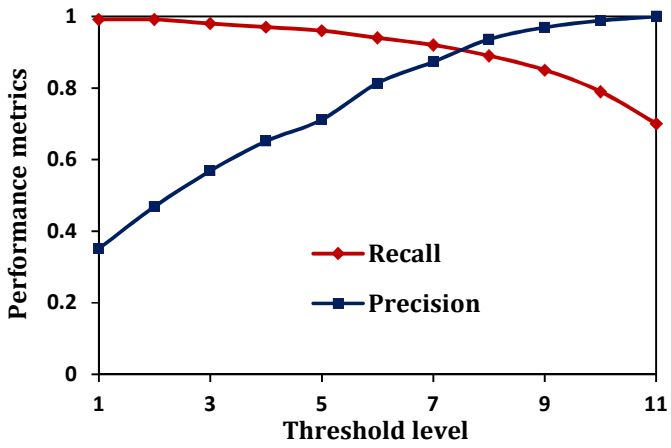



Figure 6-7: Changes in precision and recall with the change of threshold

Based on the APs calculated for both methods, it is observed that the method proposed by Torralba et al. (2004) scored higher for the recogniton of the backhoe loader and tractor (Figure 6-2). This indicates that the method could recognize the backhoe loader and tractor more precisely across all recall values in the P/R curves. In contrast, the method proposed by Felzenszwalb et al. (2010) achieved higher AP for the recogniton of the rest of the equipment classes (Figure 6-2), which signifies that it could recognize the excavator, loader and compactor more precisely along with higher recall rates.

The average values of accuracy and $F_1$ score demonstrate that both the methods performed in an identical manner for the recognition of backhoe loader and tractor, whereas the method proposed by Felzenszwalb et al. (2010) showed remarkably higher performance for the recognition of excavator (Figure 6-3). In addition, the performance of this method was slightly high for the recognition of loader and compactor. This indicates that the recognition rate, i.e. recall, of the aforementioned method was higher for excavator, loader and compactor and the performance of the method usually does not degrade rapidly for drastic pose variations of the EOI in the test images. Moreover, as accuracy and $F_1$ score are the functions of both precision and recall, the higher accuracy and $F_1$ score mean a better trade-off between the precision and recall values.

From the robustness performance test, both the methods show common characteristics, such as the recognition rate drops down with the increase in the level of occlusions. Based on the results, it is found that the method proposed by Felzenszwalb et al. (2010) is more robust than the method proposed by Torralba et al. (2004) for construction equipment recognition, in general. The higher robustness of the method of Felzenszwalb et al. (2010) is partially due to the fact that this method initially reognizes certain equipment parts. After the succesful recognition of the equipment parts, the presence of the whole equipment is determined. Therefore, the recognition rate of the method proposed by Felzenszwalb et al. (2010) typically does not drop as fast as the method proposed by Torralba et al. (2004) with the increase in occlusion level.

From the performance of the recognition methods, three general remarks can be made. First, the recognition performance of a method largely depends on the similarity between the test and training images. When the images are similar, the recognition

methods perform well and obtain more stable results. For example, if the model of the excavator is trained with such images, where the left view of the EOI is only used, then the model will be able to recognize the excavator in the test images only when it is viewed from the left. Similarly, if the training image set includes images, where the EOI appear with partial occlusions, it is highly probable that the methods will be able to recognize the EOI even with the partially occluded views in the test images. Second, the results depend on the threshold level that is used for the recognition test. The recognition rate drops rapidly with the increase in threshold level, specifically for recognizing the equipment with occlusions and difficult poses. Finally, it can be stated that the method proposed by Felzenszwalb et al. (2010) shows invariance towards considerable variations in poses and camera viewpoints, which makes it more robust towards recognizing equipment from construction site images. However, the method proposed by Torralba et al. (2004) is much faster and needs less time for the recognition of construction equipment.

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORK

## 7.1 OVERVIEW

Automation has turned into a burning issue in the construction industry, which has been transformed as one of the leading industrial sectors in many countries. It is important to promote the automation in the construction industry to speed up the construction process, reduce the project costs, etc. The ultimate goal of automation in construction is to enable the engineers and/or managers to accomplish the construction tasks in an automated way. In order to achieve this goal, one fundamental step is to automatically recognize various operational resources, including construction equipment, at construction sites. Since the construction site images can provide important management information of an ongoing project, the automatic recognition of construction resources from the site images could significantly help to achieve the automated monitoring and control tasks. Thus, the monitoring and control of construction site operations could be performed remotely, which is more dynamic, fast and efficient.

In the field of computer vision, many object recognition methods have been developed by the researchers. In order to evaluate the effectiveness of these methods, many datasets have also been created. However, none of these datasets include construction equipment images. In addition, there are several other issues restricting the use of the prevailing datasets to evaluate the existing methods for the recognition of construction

equipment. These issues include limited variability of the images in the datasets; stereotypical pose of an object, often centered, in an image (Ponce et al., 2006). Also, most images in the datasets have little or no occlusion and background clutter (Griffin et al., 2007). Since there is no dataset available to evaluate the performance of existing methods for the recognition of construction operational resources, it is unpredictable whether or not these methods could be used for on-site construction equipment recognition.

In order to address this issue, the current research proposes to create a standardized dataset of construction site images. The focus has been placed on capturing the images of construction equipment under realistic site conditions, such as multiple pieces of equipment working together with illumination variations and partial occlusions by debris, materials and other equipment. Thousands of images for 5 classes of construction equipment – backhoe loader, tractor, excavator, loader and compactor – have been collected, annotated and compiled. The images in the dataset offer a wide range of varieties, such as equipment from different manufacturers along with different sizes, poses, viewpoints and degrees of occlusions (discussed in chapter 4). The dataset developed in this work could be used to evaluate the construction equipment recognition performance of existing object recognition methods.

Two object recognition methods have been tested with the dataset developed in this research. Their performances are evaluated on the basis of three metrics, i.e. correctness, robustness and speed (discussed in chapter 4). The evaluation results show that neither of them absolutely outperforms the other, when they are used to recognize backhoe loader, tractor, and compactor. However, the method proposed by

Felzenszwalb et al. (2010) performs better for the recognition of excavators and loaders. In addition, the aforementioned method is more robust to occlusions, whereas the method proposed by Torralba et al. (2004) shows much faster recognition rate. The test results have revealed the potentials of the current dataset to evaluate the performance of existing object recognition methods for the recognition of construction equipment in a standard, unbiased, and extensive way.

## 7.2 CONTRIBUTIONS TO THE KNOWLEDGE

The main contribution of this research, to the body of knowledge in construction engineering and management, is – the accomplishment of successful equipment recognition from construction site images. The novelty of this work lies in its initiative approach to investigate the effectiveness of the existing object recognition methods for the recognition of construction equipment, which is considered as a challenging task, specifically for the fact that construction sites are typically characterized as being dirty, disorderly and cluttered with tools, materials and debris (Tajeen and Zhu, 2013). Moreover, the construction objects in the site images are often captured with partial occlusions, which makes the recognition task even more difficult, and challenging (Tajeen and Zhu, 2013). Considering these facts, this research focuses on evaluating the performance of existing object recognition methods for the recognition of construction equipment from on-site images. In doing so, the existing recognition methods are adapted for recognizing construction equipment in a comprehensive and methodical way. The main contributions that are pursued within the scope of this research are highlighted as follows:

- Dataset Development

A standardized dataset of construction equipment images is developed in this research. More than 25 construction sites were visited, and around 2,000 images have been collected, which cover a total of 5 classes of construction equipment, such as excavator, loader, tractor, compactor and backhoe loader. The images of construction equipment are captured at real construction sites to represent realistic site conditions, i.e. dirty, disorderly and cluttered. The equipment in each image is then annotated to generate the ground truth for the purpose of evaluating the construction equipment recognition performances of existing object recognition methods.

- Annotation Tool Customization

The annotations of the construction equipment are performed by using an annotation tool that has been customized based on the work of Korˇc and Schneider (2007). The novelty of the new annotation tool, which is specifically designed to annotate construction equipment, is that it provides an option of establishing the relationship between the equipment and its different parts. The relationship is established by specifying the object ID for the equipment and its parts in a hierarchical order. For example, when the equipment is identified with object ID 1, the IDs for its parts are specified as 1.1, 1.2, 1.3 etc.

- Exploring the Effectiveness of the Dataset

The dataset developed in this research is used to evaluate the construction equipment recognition performance of existing object recognition methods. So far, two common object recognition methods have been tested with the developed dataset –

discriminatively trained part-based model method (Felzenszwalb et al., 2010) and simple object detector with boosting method (Torralba et al., 2004). The images and annotations of the dataset are used, through slight modifications, as the ground truth for the evaluation of these methods. The evaluation results show that the dataset could provide an unbiased foundation to evaluate the existing recognition methods.

- Performance Comparison of Existing Recognition Methods

The performances of the methods are compared methodically and elaborately on the basis of the common performance metrics, such as correctness, robustness and computation speed. An analysis on the recognition performances of both the methods is subsequently made in this thesis. Based on the results, it is found that the performances of both methods are nearly identical regarding correctness (e.g. precision and recall), for the equipment classes – backhoe loader, tractor and compactor, in particular. However, the method of Felzenszwalb et al. (2010) performed more robustly against partial occlusions and pose variations, whereas the method of Torralba et al. (2004) is computationally favorable as it needs less time for construction equipment recognition.

## 7.3 FUTURE RESEARCH DIRECTION

The automated recognition of construction equipment accomplished through this research work provides an insight towards achieving construction site monitoring and control tasks in an automated and remote way. The future path of this research will focus on the enrichment of the dataset by including the images of other classes of construction equipment, such as scraper, grader, clamshell, crane, truck etc. In addition,

more object recognition methods will be tested to evaluate their construction equipment recognition performance, using the proposed dataset. The process followed in this thesis allows the recognition of other construction resources, i.e. workers, materials etc. Moreover, the approach can be employed in detailed applications related to construction productivity analysis, safety management and quality control. The following are a few areas, in which the current research can be further extended or applied:

- Equipment Action Recognition

The automated recognition of construction equipment from images or video frames can facilitate the automated equipment action recognition from site videos. It can help to achieve the automated progress monitoring and productivity analysis of construction work, and subsequently minimize the equipment idle time (Gong and Caldas, 2010). Hence, the future work of this research could include the real-time tracking of construction equipment in order to transform the traditional way of productivity analysis, which is manual, slow, labor intensive, and error-prone (Heydarian and Golparvar-Fard, 2012).

- Recognition of Construction Materials

The successful accomplishment of automated construction equipment recognition can be perceived as the introductory step to automate the recognition of other construction resources. The recognition of construction materials is required to guarantee proper handling, storage and availability throughout the construction work (Song, 2005). Since construction materials bear large portion of the total construction costs, the remote and

automated tracking and monitoring of construction materials can turn out to be very beneficial. Eventually, it enables the construction engineers/managers to take any rapid and corrective decision for the improvement of a construction project.

- Recognition of Construction Workers

The future direction of this research also includes automated construction workers recognition, which can be used to track the location of on-site workforce, observe their performances, and improve communications and safety in the construction sites (Chi and Caldas, 2011). Moreover, it can promote the construction safety management and visualization approach by detecting any safety violation conducted by the workers. For example, the automated recognition of workers can help to determine the use of safety equipment at construction sites, such as hard hats (Shrestha et al., 2012). As a whole, the procedure applied in the present research to recognize construction equipment (i.e. implementing different recognition methods and evaluating their performances) can provide a framework for achieving the automated recognition of other types of construction operational resources including materials, workers, and safety equipment.

**REFERENCES**


Agarwal, S., Awan, A., and Roth, D. (2004) "Learning to Detect Objects in Images via a Sparse, Part-based Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11): 1475-1490.

Azar, E. R., and McCabe, B. (2012) "Vision-based Recognition of Dirt Loading Cycles in Construction Sites", Construction Research Congress, Purdue University, West Lafayette, IN, USA, 1042-1051.

Ballard, D. H. (1981) "Generalizing the Hough Transform to Detect Arbitrary Shapes", Pattern Recognition, 13(2): 111-122.

Bay, H., Tuytelaars, T., and Gool, L. V. (2006) "Surf: Speeded Up Robust Features", 9th European Conference on Computer Vision, Graz, Austria, 3951:404-417.

Bohn, J., and Teizer, J. (2010) "Benefits and Barriers of Construction Project Monitoring using High-Resolution Automated Cameras", Journal of Construction Engineering and Management, 136(6): 632-640.

Borgefors, G. (1988) "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm", Transactions on Pattern Analysis and Machine Intelligence, IEEE, 10(6): 849-865.

Brilakis, I. K., Soibelman, L., and Shinagawa, Y. (2006) "Construction Site Image Retrieval based on Material Cluster Recognition", Journal of Advanced Engineering Informatics, 20(4): 443–452.

Caputo, B. (2004) "A New Kernel Method for Object Recognition: Spin Glass Markov Random Fields", PhD thesis, Stockholm, *http://www.nada.kth.se/~caputo* (Aug. 15, 2013).

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

Chinchor, N. (1992) “MUC-4 Evaluation Metrics”, In Proceedings of the 4th Message Understanding Conference, pp. 22–29. h*ttp://www.aclweb.org/anthology-new/M/M92/M92-1002.pdf* (Aug. 15, 2013).

Chutter, S. (2012) “Construction Industry”, *http://www.thecanadianencyclopedia.com/articles/construction-industry* (Aug 15, 2013).

Dalal, N., and Triggs, B. (2005) “Histograms of Oriented Gradients for Human Detection”, IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 1: 886-893.

Davidson, I. N., and Skibniewski, M. J. (1995) “Simulation of Automated Data Collection in Buildings”, Journal of computing in civil engineering, 9(1): 9-20.

Demsetz, L. (1990) “Automated Construction?” Construction Dimensions.

Dorkó, G., and Schmid, C. (2005) “Object Class Recognition using Discriminative Local Features”, Technical Report RR-5497, INRA-Phõne-Alpes, Feburay, 2005.

Elatter, S.M.S. (2008) “Automation and Robotics in construction: Opportunities and Challenges”, Emirates Journal for Engineering Research, 2008, pp. 21-26.

Everingham, M., Zisserman, A., Williams, C., Van Gool, L., Allan, M., Bishop, C., Chapelle, O., Dalal, N., Deselaers, T., Dorko, G., Duffner, S., Eichhorn, J., Farquhar, J., Fritz, M., Garcia, C., Griffiths, T., Jurie, F., Keysers, D., Koskela, M., Laaksonen, J., Larlus, D., Leibe, B., Meng, H., Ney, H., Schiele, B., Schmid, C., Seemann, E., Shawe-Taylor, J., Storkey, A., Szedmak, S., Triggs, B., Ulusoy, I., Viitaniemi, V., Zhang, J. (2006) “The

2005 PASCAL Visual Object Classes Challenge", In Selected Proceedings of the First PASCAL Challenges Workshop.

Everingham, M., Gool, L. V., Williams, C., Winn, J., and Zisserman, A. (2010) "The PASCAL Visual Object Classes Challenge Workshop 2010", 11th European Conference on Computer Vision, ECCV 2010, Crete, Greece.

Fei-Fei, L., Fergus, R., and Torralba, A. (2005) "Recognizing and Learning Object Categories", *http://people.csail.mit.edu/torralba/shortCourseRLOC/* (Aug. 15, 2013).

Fei-Fei, L., Fergus, R., and Perona, P. (2006) "One-Shot Learning of Object Categories", IEEE Transactions on Pattern Recognition and Machine Intelligence, *http://www.vision.caltech.edu/Image_Datasets/Caltech101/* (Aug. 15, 2013).

Felzenszwalb, P. F., Girshick, R. B., McAllester, D. and Ramanan, D. (2010) "Object Detection with Discriminatively Trained Part-Based Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9): 1627-1645.

Fergus, R., Perona, P. and Zisserman, A. (2003) "Object Class Recognition by Unsupervised Scale-Invariant Learning", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, 2: 264--271.

Freidman, J., Bentley, J. & Finkel, A. (1977), "An Algorithm for Finding Best Matches in Logarithmic Expected Time", ACM Transactions on Mathematical Software 3(3): 209–226.

Georghiades, A.S., Belhumeur, P.N. and Kriegman, D.J. (2001) "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6): 643-660.

Gong, J., Caldas, C. H. and Gordon, C. (2011) “Learning and Classifying Actions of Construction Workers and Equipment using Bag-of-Video-Feature-Words and Bayesian Network Models”, Advance Engineering Informatics, 25(4): 771-782.

Gong, J., and Caldas, C. H. (2010) “Computer Vision-Based Video Interpretation Model for Automated Productivity Analysis of Construction Operations”, Journal of Computing in Civil Engineering, 24(3): 252-263.

Grauman K. and Leibe B. (2011) “Visual Object Recognition”, Synthesis Lectures on Artificial Intelligence and Machine Learning, 5(2): 1-181.

Griffin, G., Holub, A., and Perona, P. (2007) “Caltech-256 Object Category Dataset”, http://www.vision.caltech.edu/Image_Datasets/Caltech256/ (Aug. 15, 2013).

Heydarian, A., Golparvar-Fard, M., and Niebles, J. C. (2012) “Automated Visual Recognition of Construction Equipment Actions using Spatio-temporal Features and Multiple Binary Support Vector Machines.” Proc.of Const. Research Congress.

Historica-Dominion (2012) “Construction Industry”, *www.thecanadianencyclopedia.com*

JCGM (2008) “International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM)”, *http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf*.

Kasim, N. B., Liwan S. R., Shamsuddin A., Zainal R. and Kamaruddin N. C. (2012) “Improving On-site Materials Tracking for Inventory Management in Construction Projects”, Proceedings of International Conference of Technology Management, Business and Entrepreneurship, 447-452.

Katz, M. (2010) “Simple XML Node Creation”, MATLAB Central, http://blogs.mathworks.com/community/2010/09/13/simple-xml-node-creation/ © The MathWorks Inc. 1994-2013 (Aug. 15, 2013).

Korč, F. and Schneider, D. (2007) “Annotation Tool”, Technical report TR-IGG-P-2007-01, Department of Photogrammetry, University of Bonn. *http://www.ipb.uni-bonn.de/html_pages_software/annotation-tool/* (Aug. 15, 2013).

Lamdan, Y. and Wolfson, H. J. (1988) “Geometric Hashing: A General and Efficient Model-based Recognition Scheme”, 2nd International Conference on Computer Vision, Tampa, Florida, USA, 238-249.

Leibe, B., Leonardis, A. and Schiele, B. (2004) “Combined Object Categorization and Segmentation with an Implicit Shape Model”, In ECCV 2004 Workshop on Statistical Learning in Computer Vision, Prague, Czech Republic.

Lowe, D. G. (1999) “Object Recognition from Local Scale-Invariant Features”, 7th International Conference on Computer Vision, Corfu, Greece, 2: 1150-1157.

Matas, J. and Obdrzalek, S. (2004) “Object Recognition Methods Based on Transformation Covariant Features”, 12th European Signal Processing Conference (EUSIPCO 2004), Vienna, Austria.

McCann, S. (2011) “It’s a Bird… It’s a Plane… It depends on your Classifier’s Threshold”, sanchom.wordpress.com*/2011/09/01/precision-recall/*.

Murase, H. and Nayar, S. K. (1995) “Visual Learning and Recognition of 3-D Objects from Appearance”, International Journal of Computer Vision, 14: 5-24.

Navon, R. and Berkovich, O. (2006) “An Automated Model for Materials Management and Control”, Journal of Construction Management and Economics, 24: 635-646.

Nayar, S. K., Nene, S. A. and Murase, H. (1996) “Real-time 100 object recognition system”, IEEE Int’l Conference on Robotics and Automation, 3: 2321-2325.

Nitithamyong, P. and Skibniewski, M. (2004) "Web-based Construction Project Management Systems: How to Make Them Successful?", Automation in Construction, 13(4): 491-506.

Olsen D. L. and Delen D. (2008) "Advanced Data Mining Techniques", Springer Link, ISBN: 978-3-540-76916-3.

Park, M. and Brilakis, I. (2012) "Construction Worker Detection in Video Frames for Initializing Vision Trackers", Journal of Automation in Construction, 28: 15-25.

Poggio T. and Edelman S. (1990), "A network that learns to recognize three-dimensional objects," Nature, 343: 263-266.

Ponce, J., Berg, T., Everingham, M., Forsyth, D., Hebert, M., Lazebnik, S., Schmid C., Russell B. C., Torralba, A., Williams C. K. I., Zhang J. and Zisserman A. (2006) "Dataset Issues in Object Recognition", Toward Category-Level Object Recognition, Springer, pp. 29-48.

Pope, A. R. (1994) "Model-based Object Recognition - A Survey of Recent Research", Technical Report TR-94-04, University of British Columbia.

Powers, D. M. W. (2011) "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation", Journal of Machine Learning Technologies, 2(1): 37-63.

Rijsbergen C. J. V. (1979), "Information Retrieval", London: Butterworths. *http://www.dcs.gla.ac.uk/Keith/Preface.html* (Aug. 15, 2013).

Rucklidge, W. J. (1995) "Locating Objects using the Hausdorff Distance", 5th International Conference on Computer Vision, Boston, USA, pp. 457-464.

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008) "LabelMe: A Database and Web-based Tool for Image Annotation", International Journal of Computer Vision, 77(1): 157-173.

Sasaki Y. (2007) “The Truth of F-measure”, Teaching, Tutorial materials, School of Computer Science, University of Manchester, Version: 26th October, 2007.

Sch¨olkopf, B. and Smola, A. (2002) “Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond” MIT Press: Cambridge, MA.

Sheskin, D. (2004) “Handbook of Parametric and Nonparametric Statistical Procedures”, CRC Press, pp. 59.

Shrestha, P., Yfantis, E. and Shrestha, K. (2012) “Construction Safety Visualization”, Civil and Environmental Engineering Department, University of Nevada, Las Vegas.

Sim, T., Baker, S. and Bsat, M. (2002) “The CMU Pose, Illumination, and Expression (PIE) Database”, Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition.

Song, J. (2005) “Tracking the Location of Materials on Construction Projects”, University of Texas, PhD Thesis.

Steger, C. (2002) “Occlusion, Clutter, and Illumination Invariant Object Recognition”, In: International Archives of Photogrammetry and Remote Sensing, 34(3): 345-350.

Steger, C. (2001) “Similarity Measures for Occlusion, Clutter, and Illumination Invariant Object Recognition”, Pattern Recognition, 2191: 148-154.

Swain, M. J. and Ballard, D. H. (1991) “Color Indexing”, International Journal of Computer Vision, 7(1): 11-32.

Tajeen, H. and Zhu, Z. (2013) “Image Dataset Development for Construction Equipment Recognition”, Proceedings of 4th Construction Specialty Conference, CSCE, Montreal, Canada.

Tatum, C. B. (1989) “Design and Construction Automation: Competitive Advantages and Management Challenges”, Proceedings of 6th International Symposium on Automation and Robotics in Construction, Austin, Texas, pp. 332-339.

Taylor, J. R. (1999) “An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements”, Second edition, University Science Books, California.

Torralba, A., Murphy, K. P. and Freeman, W. T. (2004) “Sharing Features: Efficient Boosting Procedures for Multiclass Object Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 762- 769.

Ulrich, M. and Steger, C. (2002) “Performance Comparison of 2D Object Recognition Techniques”, In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34 (Part 3A), pp. 368–374.

Yang, M. H., Roth, D. and Ahuja, N. (2000) “Learning to Recognize 3D Objects with SNoW”, In ECCV 2000, pages 439–454.

Yang, M. H. (2009) “Object Recognition”, Liu L. and Ozsu M. T. (Ed.), Encyclopedia of Database Systems, 1936-1939. http://faculty.ucmerced.edu/mhyang/papers/object-recognition-chapter.pdf (Aug. 15, 2013).

Zhang, J., Marszalek, M., Lazebnik, S. and Schmid, C. (2007) “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study,” International Journal of Computer Vision, 73(2): 213-238.

Zhu, M. (2004) “Recall, Precision and Average Precision”, Working paper 2004-09, Department of Statistics and Actuarial Science, University of Waterloo.

Zou, J. and Kim, H. (2007) “Using Hue, Saturation, and Value Color Space for Hydraulic Excavator Idle Time Analysis”, Journal of Computing in Civil Engineering, ASCE, 21: 238-246.

**APPENDIX A**

# ANNOTATION OF CONSTRUCTION EQUIPMENT

The annotation process of construction equipment is performed with the annotation tool that has been customized based on the work of Korˇc and Schneider (2007). The annotation tool is a MATLAB-based tool for image annotation and coupled with LabelMe toolbox, which has been developed at MIT. In order to annotate the images, both the annotation tool and the LabelMe toolbox are first added to MATLAB search path, such as C:\AnnotationTool and C:\LabelMeToolbox, respectively. Afterwards, home folders for the images and annotations are created and the directories are specified such as C:\my-image-database\images and C:\my-image-database\annotations. Figure A-1 represents the customized annotation tool, which is tailored to annotate the construction equipment and its different parts. The redesigned Graphical User Interface (GUI) of the annotation tool is shown in Figure A-2. A detail working steps with the annotation tool is described in the following sections.

The annotation tool initially opens in a separate window, while working in the MATLAB environment. After that, the image dataset folder is opened by pressing the button 'Open Image Folder' from the 'File' menu. The names of all the images are displayed in the 'Filename' list box from which any image can be selected for annotation. The image source and view type are then specified from the dropdown menu of 'Current Image' panel. The view type options are designated as front, rear, left, right and the corners (Figure A-3a). The images are annotated by pressing the 'Annotate' button on the 'New Annotation' panel of the tool (Figure A-3b), which activates the annotation mode (Korˇc and Schneider, 2007). The bounding polygons are drawn by using the left mouse button, along the edge of the entire construction equipment. When the entire equipment is bounded by the polygon, it can be closed by pressing the right mouse button.

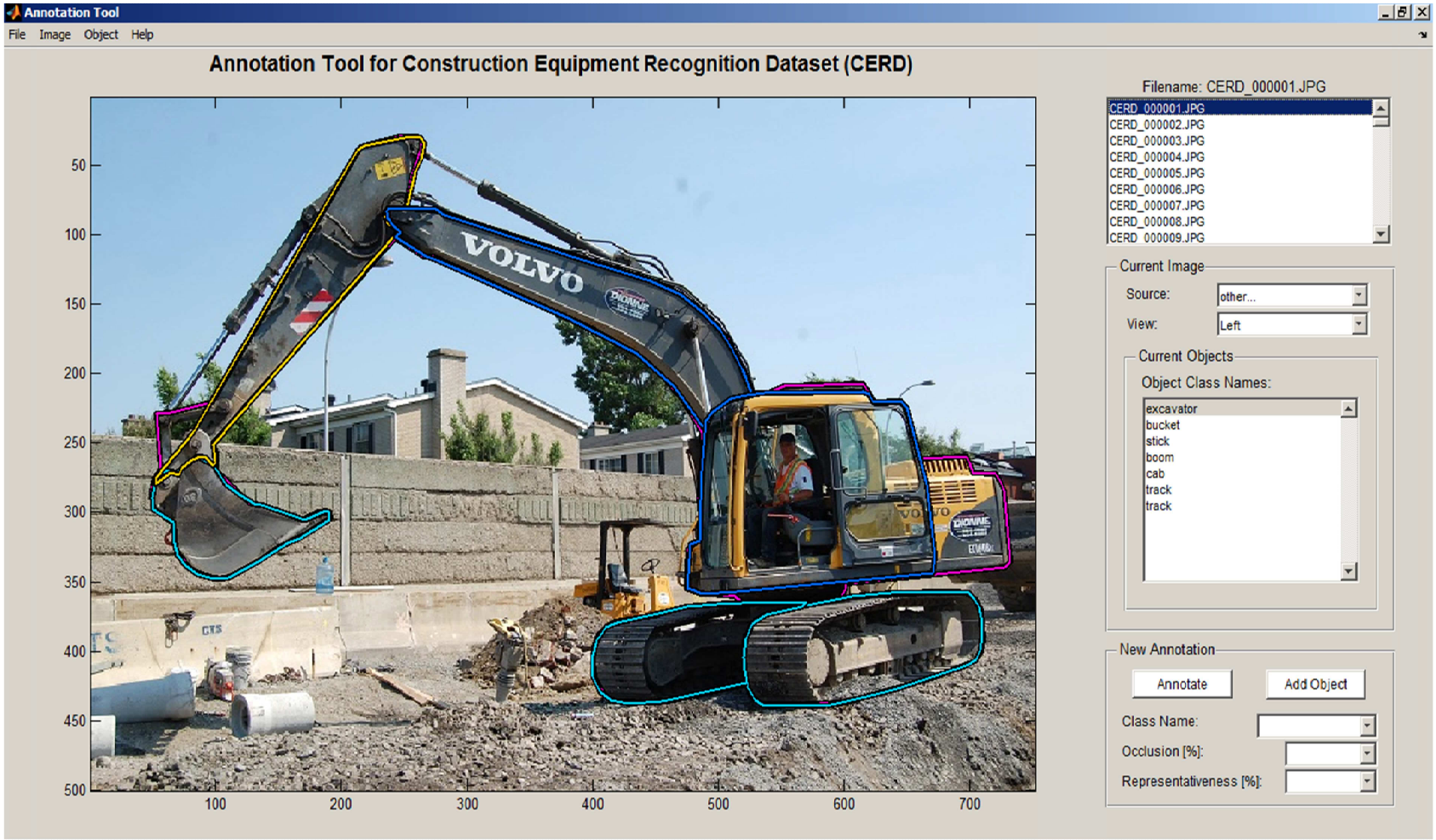



Figure A-1: The customized annotation tool to annotate construction equipment with different parts

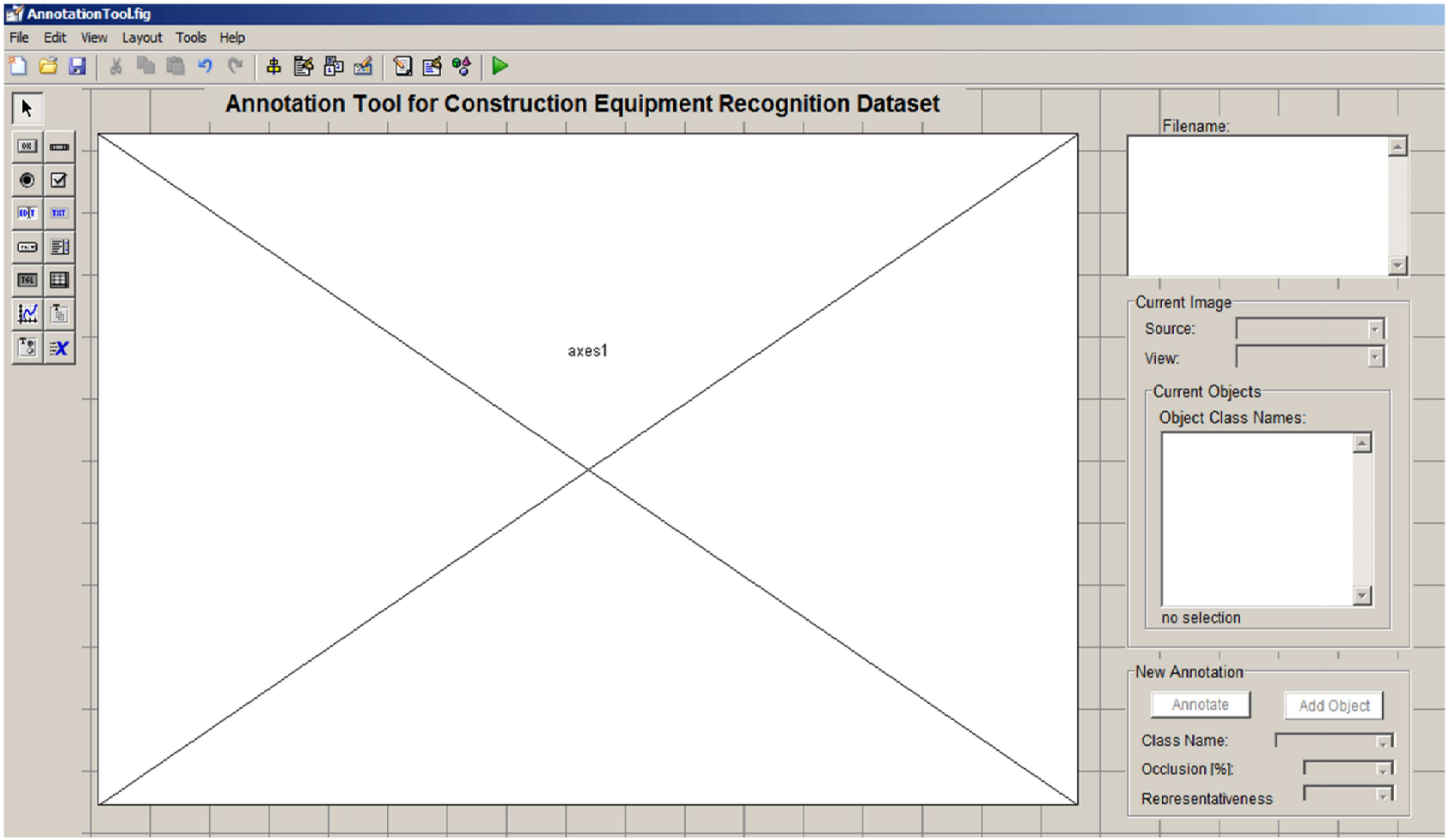
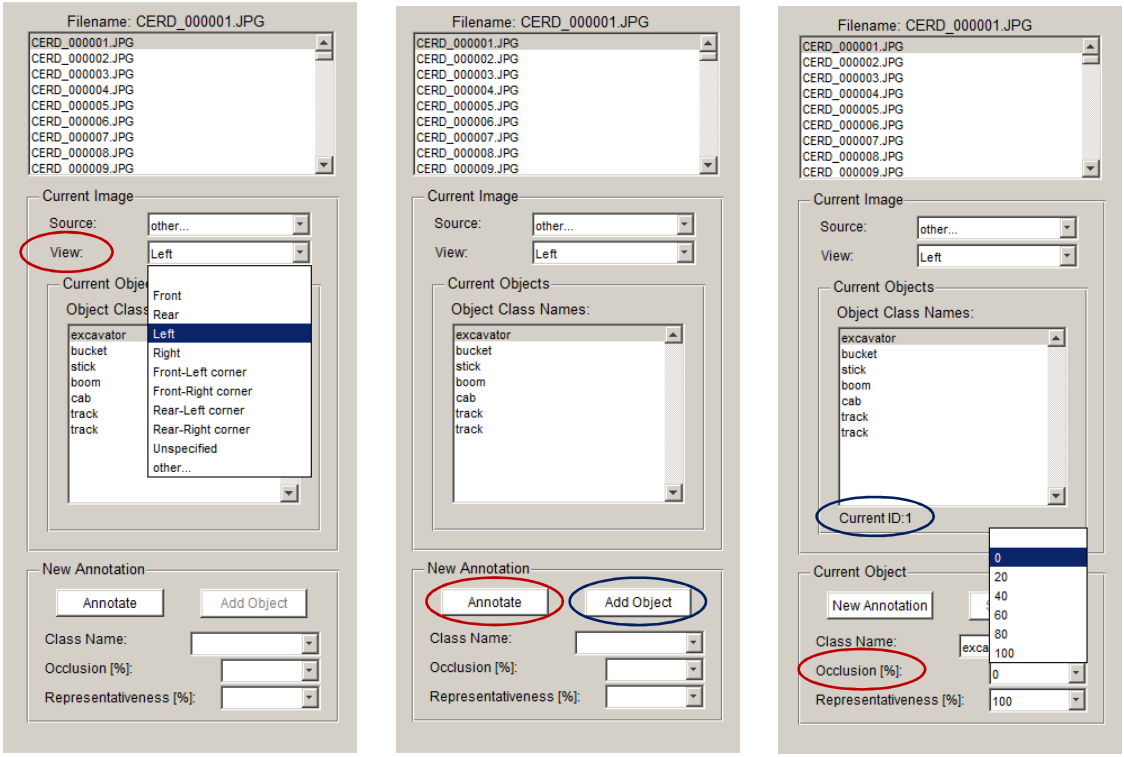


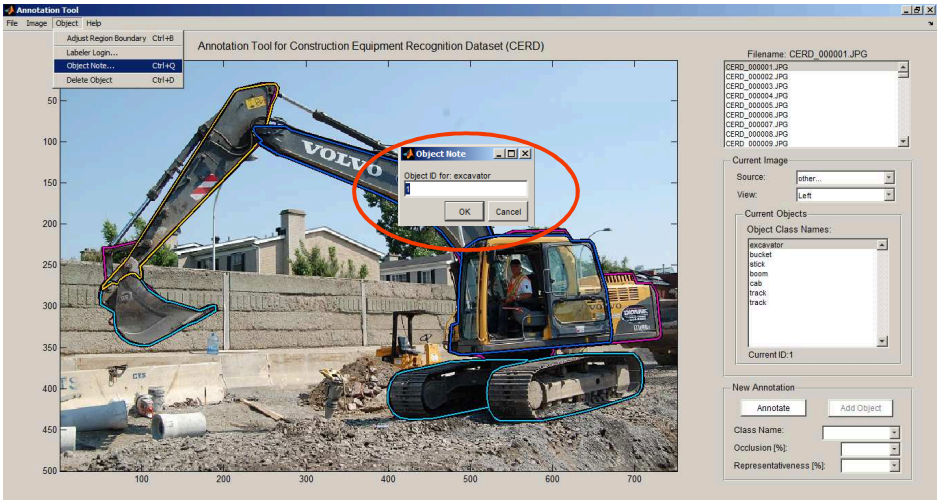

Figure A-2: The redesigned Graphical User Interface (GUI) of the annotation tool

The class of equipment is specified from the 'Object Class' pop up menu, which contains all the names of the equipment and their parts. If multiple equipment are present within an image, they can be annotated with individual polygon one after another. Then the degrees of occlusion and representativeness are approximately quantified, and indicated by choosing the right option from their respective pop up menu (Figure A-3c) (Korˇc and Schneider, 2007). After these information are specified, the annotation is then added to the 'Current Objects' list box by pressing the 'Add Object' button (shown in Figure A-3b). The identifications of the annotated objects are provided by unique identifiers, i.e. object IDs, which can be added by pressing the 'Object Note' button from the 'Object' menu bar as shown in Figure A-4(a).

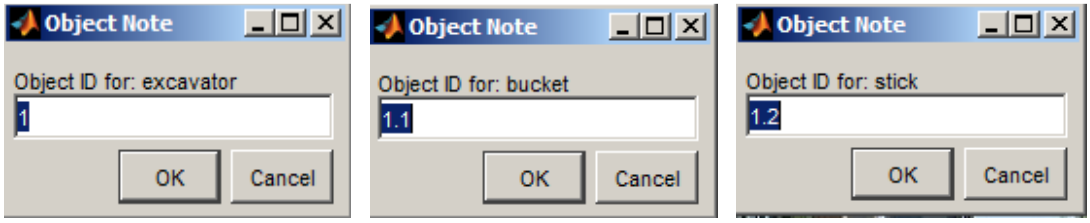



(a) (b) (c)

Figure A-3: (a) 'View type' in 'Current Image' panel, (b) 'Annotate' and 'Add Object' button in 'New Annotation' panel, (c) 'Occlusion' in 'Current Object' panel




(a)




(b) (c)

Figure A-4: (a) Object IDs added by 'Object Note' from the 'Object' menu bar, (b) ID for equipment (excavator) as 1, (c) ID for equipment parts (bucket and stick) as 1.1, 1.2 etc.

The IDs for equipment are indicated with sequential numbers such as 1, 2, 3 etc. (shown in Figure A-4b). Similarly, the annotations of the equipment parts are also performed by drawing polygons. However, the object IDs are specified in a hierarchical order, such as 1.1, 1.2, 1.3 etc. for the equipment parts (shown in Figure A-4c). If the IDs are specified for an equipment and its parts within an image, it is possible to display their respective IDs by selecting the name in the 'Current Objects' list box as shown in Figure A-3(c). After annotating an image, all the aforementioned annotation information is stored in an XML file with the same name as the annotated image. In this way, one can store all the XML files in the annotation home folder.

**APPENDIX B**

# CONVERSION OF ANNOTATION INFORMATION

As stated in chapter 5, the annotations of the dataset can be used, through slight conversions, as the ground truth to evaluate the performance of the recognition method developed by Felzenszwalb et al. (2010). In order to convert the annotation files of the dataset, the functions for reading, writing, and transforming XML files as mentioned by Katz (2010) is adopted here. For example, the function 'xmlread' is used to read the specified XML file, which returns a 'Document Object Model' (DOM) node representing the document. Additionally, the function 'xmlwrite' is used to write the 'Document Object Model' (DOM) node to the new file as specified in the filename. The syntaxes for the functions are as follows: DOMnode = xmlread (filename), and xmlwrite (filename, DOMnode). The 'xmlread' and 'xmlwrite' functions are used as shown below.

```
%% Read XML file
fileName='C:\my-image-database\annotations\CERD\CERD_000001.xml';
xDoc = xmlread (fileName);

%% Write new XML file
xmlFileName = 'D:\New XML\CERD\CERD_000001.xml';
xmlwrite (xmlFileName, docNode);
```

The XML files produce a hierarchical tree structure represented as a set of linked nodes, where a root node and sub-trees of child nodes construct the tree structure (Katz, 2010). In order to obtain the child node information from the XML files, the 'getChildNodes' function is used to return a node list of the children of the current Node. Moreover, the 'getElementsByTagName' method extracts the information contained within the specified name (Katz, 2010). This way, the required information from the

annotation files, such as file name, image size, equipment class, polygon co-ordinates are extracted. Then the coordinates of the polygon points are compared with each other in order to obtain the maximum and minimum values in x and y directions. After the determination of the maximum and minimum coordinates of the bounding polygon, a bounding box (rectangle in Figure 5-1) is plotted to surround the equipment. Hence, a new XML file is created, containing the information (x and y coordinates) of the newly drawn rectangle. The new XML file is, therefore, simple while having reduced file size. The procedure of creating new simplified XML file is schematically shown in Figure B-1. To create the new XML files, the 'getDocumentElement' function is initially used to create the root element node of the document (Katz, 2010). The function 'appendChild' is used to add a new child node in the hierarchical tree structure of the new XML file (Katz, 2010). A typical example of the extraction of 'polygon' information is presented below. In addition, the node creation for 'object', 'bndbox' (bounding box) and 'xmin' (minimum x coordinate) are shown.

```
%% Get the "polygon" node
objectInfo = annotationNode.getChildNodes;
polygonInfo = objectInfo.getChildNodes;
polygon = polygonInfo.getElementsByTagName('polygon').item(0).getTextContent;

%% New XML creation
docNode = com.mathworks.xml.XMLUtils.createDocument('annotation');
docRootNode = docNode.getDocumentElement;
%% Object node creation
object_node = docNode.createElement('object');
object_node.appendChild(bndbox_node);
docRootNode.appendChild(object_node);
%% Bounding box (bndbox) node creation
bndbox_node = docNode.createElement('bndbox');
xmin_node = docNode.createElement('xmin');
xmin_node.appendChild(docNode.createTextNode(min_eq_x));
bndbox_node.appendChild(xmin_node);
```

Figure B-1 shows the XML file conversion determining the maximum and minimum x and y coordinates of the polygon. Figure B-1(a) shows the original XML with polygon format, and Figure B-1(b) shows the converted XML file with the bounding box format.

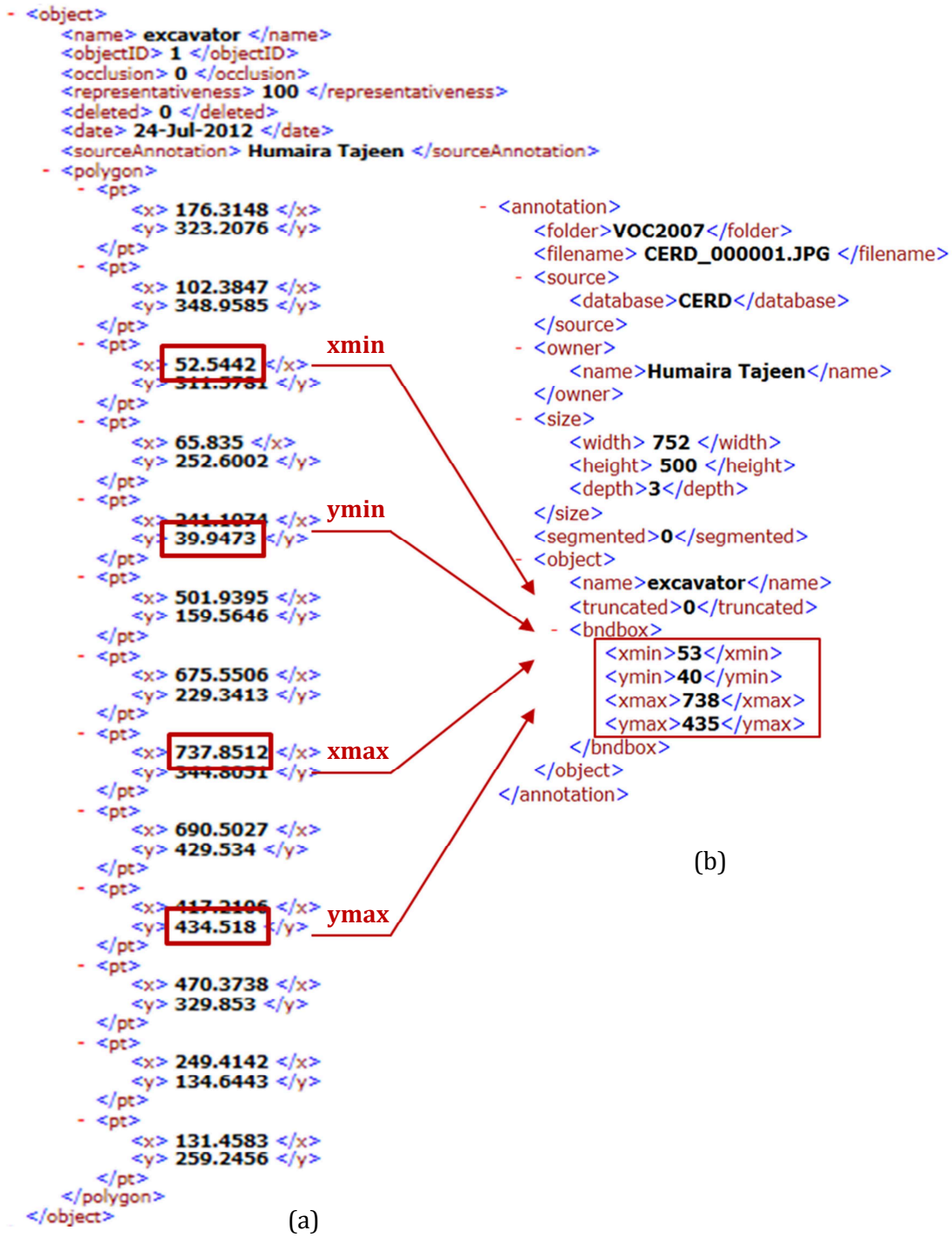



Figure B-1: (a) XML file with polygon format (b) XML file with bounding box format

**APPENDIX C**

## CALCULATION OF AVERAGE PRECISION

The average precision (AP) is obtained by calculating the sum of the precision multiplied by the change in recall, at each threshold value (Zhu, 2004). It can be expressed in the form of equation (C-1). The calculation procedure of AP for loader recognition is shown by determining the area under the P/R curves (Figure C-1).

$$AP = \sum_{i=1}^{k} P_i \left( R_i - R_{i-1} \right) \tag{C-1}$$

where, $i$ = threshold level and $R_{i-1} = 0$, if $i = k$

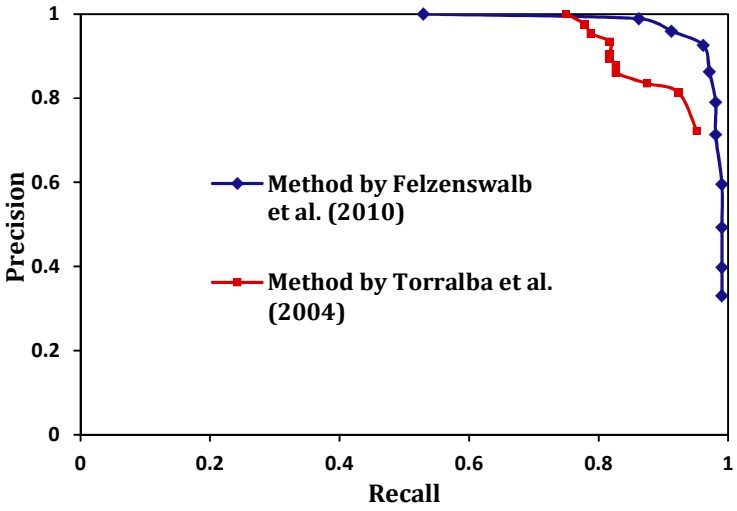



Figure C-1: P/R curves for the recognition of loader

AP (Method by Felzenszwalb et al. 2010)

= (1 x 0. 5294) + (0.9888 x 0.3333) + (0.9592 x 0.0499) + (0.9259 x 0.0489) + (0.8632 x 0.0097) + (0.7907 x 0.0096) + (0.7133 x 0) + (0.5954 x 0.0096) + (0.4928 x 0) + (0.3977 x 0) + (0.3301 x 0) = 0.9738.

AP (Method by Torralba et al. 2004)

= (1 x 0. 75) + (0.9759 x 0.0288) + (0.9535 x 0.0097) + (0.9341 x 0.0288) + (0.9043 x 0) + (0.8947 x 0) + (0.8776 x 0.0096) + (0.86 x 0) + (0.8349 x 0.0481) + (0.8136 x 0.0481) + (0.7226 x 0.0288) = 0.9227.

The APs of the methods are obtained as 0.9738 for the method by Felzenszwalb et al. (2010) and 0.9227 for the method by Torralba et al. (2004). It is calculated as the sum of precision multiplied by the change in recall at different thresholds, as stated in equation (C-1). From the example, it can be observed that the points at which the recall does not change do not contribute to this sum. These are the points in the graph, where the recall drops straight down vertically. Since AP is computed from the sum of the area under the curve, those vertical sections of the curve do not add any area (McCann, 2011). The APs for other equipment classes, i.e. backhoe loader, tractor, excavator and compactor, are also obtained using the same calculation procedure.