Estimating construction waste truck payload volume using monocular vision

Junjie Chen, Weisheng Lu*, Liang Yuan, Yijie Wu, and Fan Xue

Department of Real Estate and Construction, University of Hong Kong, Pokfulam Hong Kong, China

Abstract

Quantifying the amount of truck-loaded materials is a universal problem encountered in numerous industrial operation scenarios. Likewise, in construction waste management, inspectors at disposal facilities are often required to measure the amount of construction and demolition waste (CDW) loaded by incoming trucks to determine their admissibility. Due to the bulky and clutter nature of construction materials, accurate waste quantification without sacrificing the operability in field has always been a challenge. This study proposes a CDW volume estimation algorithm based on monocular vision, which can automatically quantify the amount of specific material components, e.g., rock, gravel, and wood, in waste mixtures from a single image. The algorithm achieved a relative error of 0.065 in estimating tuck bucket dimensions, and consumed 3.3 s in average to process per image. The results demonstrate the efficacy of the proposed algorithm in achieving material-level CDW volume estimation. The algorithm was applied to analyze 2,914 waste truckloads received by an off-site sorting facility in Hong Kong. It was observed that the facility entrance received around 800.0 m^3 CDW per day, of which about 10.8 m^3 were rejected. The non-inert wood/cardboard accounts for the highest proportion among all material types, implying many waste dumps accepted by the facility might have violated the admissibility criteria. The study contributes to the knowledge body by providing a novel, non-destructive approach to quantify CDW via monocular vision. It can also be extended to address the general problem of truck payload quantification in other scenarios such as road construction, warehouse inventory management, and logistics and supply chain management.

Keywords: Construction and demolition waste; Amount quantification; Volume estimation; Computer vision; Semantic segmentation; Photogrammetry.

* Corresponding author.
E-mail address: wilsonlu@hku.hk.

1. Introduction

In many industrial activities, it is a conventional need to measure the amount or properties of truck payloads. A typical example is the operation of expressway, which measures the weight of incoming vehicles to prevent overloaded trucks from causing road damages (Visco India, 2020). In infrastructure development, load volume of outgoing dump trucks needs to be tracked and monitored to estimate earthwork volumes and construction intensity (Loadscan Ltd., 2021). For special materials, it is also required to ensure their physical properties, e.g., temperature for asphalt concrete and moisture for earth-fill materials (Liu et al., 2013), have met relevant construction standards. In cross-region logistics, truck cargo needs to be inspected with cameras, radio frequency identification devices (RFID), and other sensing approaches to ensure quality and completeness of the payload before delivery (Zhong et al., 2017).

Similar scenarios are also observed in construction and demolition waste (CDW) management. A huge amount of CDW is generated globally every day (Kaza et al., 2018). The generated waste materials are transported by truck trailers to designated disposal facilities, and different levies are charged according to their compositions. In Australia, for example, the state authority of New South Wales has released clear guidelines (NSWEPA, 2020) and criteria (NSWEPA., 2018) to levy different disposal fees for waste loads with different composition. In Hong Kong, CDW, according to their non-inert and inert properties, should be directed to landfills and public fills, respectively, while waste mixture with more than 50% inert content should be transported to off-site sorting facilities (HKEPD, 2019). Therefore, the truck payloads need to be estimated to ensure waste materials have reached the designated destinations. Despite the clarity and explicitness of the control criteria, how to operatize them in field is a challenge. Physical separation by digging into each waste dump can surely provide precise amount of specific waste types, but is difficult to implement due to efficiency issues.

In reality, the criteria have always been, more or less, simplified in engineering practice to ensure operability. For example, the Hong Kong Environmental Protection Department (HKEPD) materialized the “50% inert content requirements” into two easy-to-measure indices based on the weight and depth of the waste dumps. A bunch of sensors (e.g., weighbridge, range finders, and cameras) have been set up at the entrance of the waste disposal facilities to collect relevant data. However, the simplification might have led to significant deviations as the simplified indices do not directly reflect the amount of different material type in the waste mixture. This is confirmed by a report from the Hong Kong Audit Commission (Audit Commission, 2016), which revealed that the waste dumps of many truck loads accepted have not met the inert content requirement. The deficiency of existing approach calls for a new quantification approach for truck-loaded CDW, which should not only be efficient enough to implement in field, but also have sufficient granularity to

precisely sense the volume of the diverse materials in waste mixtures.

The recent development in computer vision (CV) and artificial intelligence (AI) provides opportunities to tackle the conundrum of CDW quantification. This is especially true as the stereo vision technology greatly expands the capability of CV, from the original 2D images to the understanding and perception of 3D information (Liu and Aggarwal, 2005). The technological advancement has fostered the development of research fields such as robotics, and is becoming a catalyst for new industries such as autonomous driving (Singh, 2020a) and augmented reality (Pollefeys, 2017). The technology has also entered people's daily life with functionalities unimaginable in the past. For example, people can easily measure objects' dimension with the cameras of their smart phones (Singh, 2020b); the generation Z is enjoying videogames with an unprecedented immersive experience by using stereo cameras such as Kinect and RealSense.

It would be promising if stereo vision can be used to quantify CDW amount without direct contact or sorting of the waste materials. However, exiting stereo vision solutions either rely on binocular cameras or utilize the rational of structure-from-motion (SfM) to obtain a sense of depth from images. These requirements are difficult to meet in CDW disposal facilities, as the most common setting is monocular cameras installed in fixed places. Quantifying CDW from monocular vision is notably more challenging for the following reasons. First, there is a lack of semantic information on waste composition from the raw images, making it difficult to directly obtain detailed quantities of specific material components. Second, because only one single image can be captured, the depth dimension is lost, leading to the difficulty of recovering waste dumps' 3D geometry.

This research attempts to take the challenges of CDW quantification based on monocular vision, and proposes an automated volume estimation algorithm for truck-loaded waste dumps. It does so by leveraging deep learning-enabled semantic segmentation to recognize waste materials from mixed CDW, and compensating the loss of depth information with prior knowledge on truck dimension profiles. Our study reaffirms the power of computer algorithms. With a single camera and a bunch of sensors, 3D geometry of diverse materials in the truck-loaded CDW dumps can be reconstructed in a matter of seconds. The required sensors (i.e., camera, range finders, and weight-bridge) are all common configurations at existing waste disposal facilities. Without any additional hardware, the proposed algorithm can be easily deployed to existing facilities, benefiting policy makers, waste processing contractors and other stakeholders with a valuable information on CDW volume to help determine admissibility of waste truckloads.

The remaining of this paper is organized as follows. Subsequent to this introductory section, Section 2 describes the current state of construction waste quantification and vision-based

object dimension measurement. Section 3 illustrates the proposed monocular vision algorithm for CDW volume estimation, Section 4 delivers its experimental results, and Section 5 presents preliminary applications of the algorithm in a case study. Section 6 discussed general implications of the research findings and conclusions are drawn in Section 7.

2. Literature review

2.1 Estimating construction and demolition waste amount

Quantifying CDW has been investigated since around three decades ago (Gavilan and Bernold, 1994). CDW quantification can be roughly classified into two directions according to the difference of measuring unit, namely, quantification by weight (ton, kg, or lb) and by volume (m^3). While the former can provide accurate information on the amount of waste dumps as a whole, it is difficult to weigh each specific material component in the mixture without onerous physical separation. By comparison, quantification by volume might be a more operable approach as waste volume is directly related to its visual appearance, floor space, geometric profiles. This is especially the case in certain scenarios. For example, volume-based quantification was demonstrated to be more suitable when planning construction waste hauling trucks (Lu et al., 2021b).

Scholars have proposed various approaches for volume-based CDW quantification. Most previous studies focus on predicting project-level waste generation by utilizing historical statistics, because of the extensive project waste management demand. For example, Gavilan and Bernold (1994) developed a waste-management system to estimate the quantity of CDW in residential projects, based on the data collected through field observation. Llatas (2011) presented a model which allows technicians to estimate construction waste amount (by volume) according to building design documents. The basis of this model is the European Waste List. A similar study can also be found in (Lam et al., 2019). The mentioned studies aim to forecast future construction waste generation at the design stage. Based on the context of residential building construction in Spain, Villoria Sáez et al. (2011) developed a quantification tool to predict the volume of CDW generated during different project stages. Historical statistics and related indices such as GDP have also been used to predict the CDW generation volumes at a country level, such as the United States (Prairie Village, 1998), Greece (Fatta et al., 2003), and so forth.

To serve on-site waste processing, disposal and transport dispatching, research efforts have been made to estimate CDW volume in field. They did so by harnessing learned waste accumulation regularities (Katz and Baum, 2011), or by simplifying profiles of waste dumps as either pyramidal or rectangular shapes (Arumugam et al., 2020). In (Hoang et al., 2020; Lau et al., 2008), the authors assumed the waste materials are stored in the layout of either stockpiled, gathered, scattered or stacked. Based on the types of layouts, material volumes can be estimated accordingly. Despite the easiness of implementation of such methods, they

tend to oversimplify the cluttered and heterogeneous characteristics of CDW mixture, and fail to provide accurate material composition information with sufficient granularity. Thus, a better CDW volume estimation approach is in need, which should not only be precise enough to sense fine-grained compositions of waste mixtures, but also be efficient and operable enough to implement in field.

2.2 Measuring object dimension by computer vision

Measuring dimensions of objects in photographs is not a new problem; rather, it has been researched for over a century, and was first addressed with the term "Photogrammetry" in 1870s. The basic rationale of photogrammetry is the geometrical-mathematical model of how light travels from the objects, distorted when passing through camera lens, and finally casts onto the sensor (Aber et al., 2010). In its early days, the technique of photogrammetry was implemented manually and relied on delicately designed instruments for the purpose of topographic mapping and architectural survey (The Center for Photogrammetric Training, 2008). The invention of computers made it possible to process photographs in batch, which significantly improved the efficiency and push the discipline to a new phase called analytical photogrammetry (The Center for Photogrammetric Training, 2008). In addition, the subsequent advancements of computer vision further expand the capability of photogrammetry by augmenting it with the intelligence to recognize objects in images.

In computer vision, an active research field is computer stereo vision, which aims to obtain 3D information from digital images based on the basic principles of photogrammetry. In the most common setting, two cameras installed at different known locations are used to capture photographs of the same scene; thus, depth perception can be achieved by triangulating features correspondences between the two photographs. Torr and Zisserman (2000) proposed a robust method to estimate multiple view relations from point correspondences. Hamdan et al. (2021) demonstrated the viability to extract 3D geometry of structural damages from multiple overlapped photographs in the lab environments. Other than infrastructure monitoring, stereo vision has been used in environmental engineering as well for the monitoring of deep-sea ecosystems (Aguzzi et al., 2019). Where the "two (or multiple) cameras" setting is not available, multiple photographs taken from different locations and with different postures are required to extract the 3D information of a stationary object. This is widely known as the "structure-from-motion (SfM)" technique. Ost et al. (2021) presented such a SfM system that used overlapping secondary electron images to reconstruct three-dimensional topography models of soil microaggregates. Isailović et al. (2020) applied sequential aerial images to reconstruct point cloud models of bridges, from which geometry of spalling damages recognized by deep neural networks can be extracted.

Compared with the common settings in stereo vision, object measurement based on monocular vision is notably more challenging as it lacks diverse views of the scene to generate a sense of

depth. To compensate that, prior knowledge or certain assumptions of the scene are required by existing studies (Li et al., 2018; Xue et al., 2018; Zaheer et al., 2018). In (Zaheer et al., 2018), the regularities of man-made structures (e.g., orthogonal and parallel lines) were leveraged to reconstruct 3D profiles of multi-planar scenes from a single view. Xue et al. (2018) considered parametric models of common building elements as a priori, and developed a derivative-free optimization approach to generate building information models from 2D images. Lim et al. (2014) assumed cracks are on the same plane with inspection robots, and proposed an algorithm to extract defect geometry on bridge deck by the camera line-of-sight with the deck surface. Similar approaches can be used for waste amount estimation as well. As truck manufacturers all follow designated assembly standards and regulations, it would not be too difficult to make certain assumptions on the distribution of truck dimensions. Based on such assumptions, the problem of monocular vision can be simplified, making it solvable to recover 3D geometry of truck-loaded wastes from single images.

3. Methods

3.1. Construction waste semantic segmentation

The rapidly developing deep learning (DL) makes it possible to train computer algorithms to detect objects from images with a pixel-level granularity. This pixel-level image recognition task is known as semantic segmentation in the computer science community. Semantic segmentation based on DL can be used to recognize specific waste components from images of mixed CDW dumps as well. There are many off-the-shelf semantic segmentation techniques that have achieved solid performance on public datasets, e.g., Mask R-CNN (He et al., 2017), U-Net (Ronneberger et al., 2015), and the DeepLab series (Li, 2020). In our previous work (Lu et al., 2021a), a DeepLabv3+ model was trained to segment 9 types of CDW materials and relevant objects. Fig. 1 shows some examples of CDW semantic segmentation, where the meanings of encoded colors in the semantic maps have been explained in the table at the right side. Note that the area of trucks is also recognized by the model, which lays the foundation to detect boundary of the dump bucket and further calculate its dimension in Section 3.3.

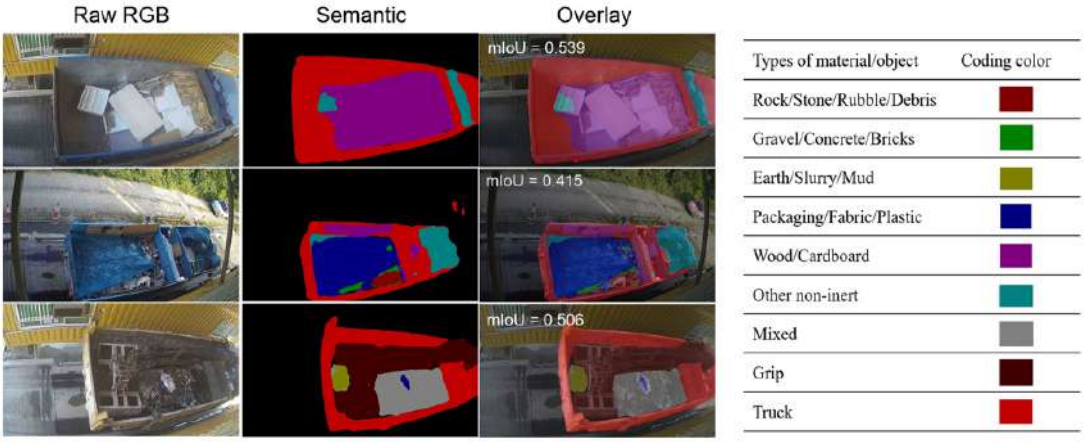

Fig. 1. CDW semantic segmentation.

3.2. Creation of truck dimension profiles

A major challenge for 3D reconstruction from monocular vision is the loss of depth information. Previous researches have more or less utilized some sort of prior knowledge or assumptions on the objects of interests to compensate the loss of the depth dimension, e.g., the principle of perspective (Asadi et al., 2019), and the Manhattan world assumption for built environment scene reconstruction (Li et al., 2018).

In 3D geometry reconstruction of truck-loaded wastes, similar priori on truck dimensions might also be leveraged to confine the solution space, making the problem solvable. We notice that there is a certain level of flexibility in bucket dimensions, as waste haulers can customize their dump buckets. Nevertheless, the front cabins of trucks are relatively stable, and the bucket size shall generally be in line with the cabin to abide relevant transportation rules and regulations. According to Code of Practice for Loading of Vehicles (CPLV) released by the HONG KONG Transport Department (HONG KONGTD, 2019) , the truck side mirrors should extend beyond the its payload so that driver can the traffic condition behind. This means the truck bucket has roughly the same width as front cabin. Meanwhile, to load as much cargo as possible in a single trip, waste haulers tend to extend their buckets to as high as the cabin.

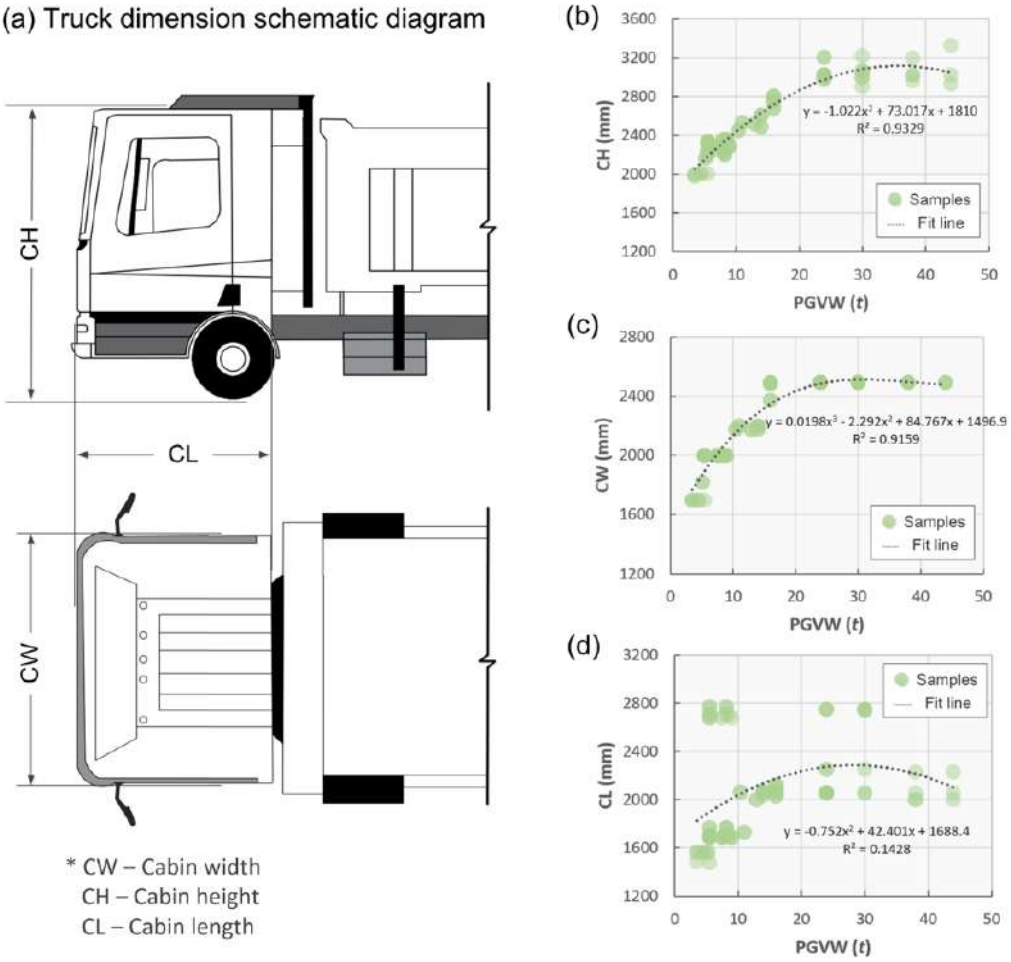

Fig. 2. (a) Schematic diagram showing how truck dimensions are defined; (b) Relationship between CH and PGVW; (c) Relationship between CW and PGVW; (d) Relationship between CL and PGVW.

Here we develop a mathematical model called truck profiles to describe the variations of truck dimensions for different types of trucks. The basic rationale is based on the concept of user profiles, which are frequently used in social media, online shopping, and other Internet sectors to retrieve useful patterns from users' behaviors and features. To create the truck profile model, we have investigated more than 90 truck models from 3 main manufacturers in the CDW management industry of HONG KONG, i.e., Isuzu, Hino, and Fuso. The investigated factors include cabin height (CH), cabin width (CW), cabin length (CL), and permitted gross vehicle weight (PGVW) of the respective truck model, which were collected from the official HONG KONG websites of the three manufacturers.

Fig. 2 (a) shows how the truck CH, CW, and CL are defined by a schematic diagram, while Fig. 2 (b)~(d) present the scatter plots between the cabin dimensions with the PGVW of all investigated trucks. It was observed that CH and CW demonstrate significant correlation with the PGVW, both with their coefficients of determination larger than 0.9. The findings implicate the viability to estimate the height and width of the cabin (and thus the bucket) according to the truck's PGVW, which is a major truck manufacturing parameter that can be easily obtained.

3.3. Construction waste volume estimation

3.3.1. Camera calibration

The purpose of camera calibration is to obtain both its intrinsic and extrinsic parameters. The intrinsic parameters are those that describe the internal characteristics of a camera, e.g., focal length, resolution, and distortion coefficients. The extrinsic parameters, on the other hand, describe the camera's posture relative to a global coordinate system. The acquisition of the two aspects of parameters is a basic prerequisite in many photogrammetric applications, including CDW volume estimation in our case.

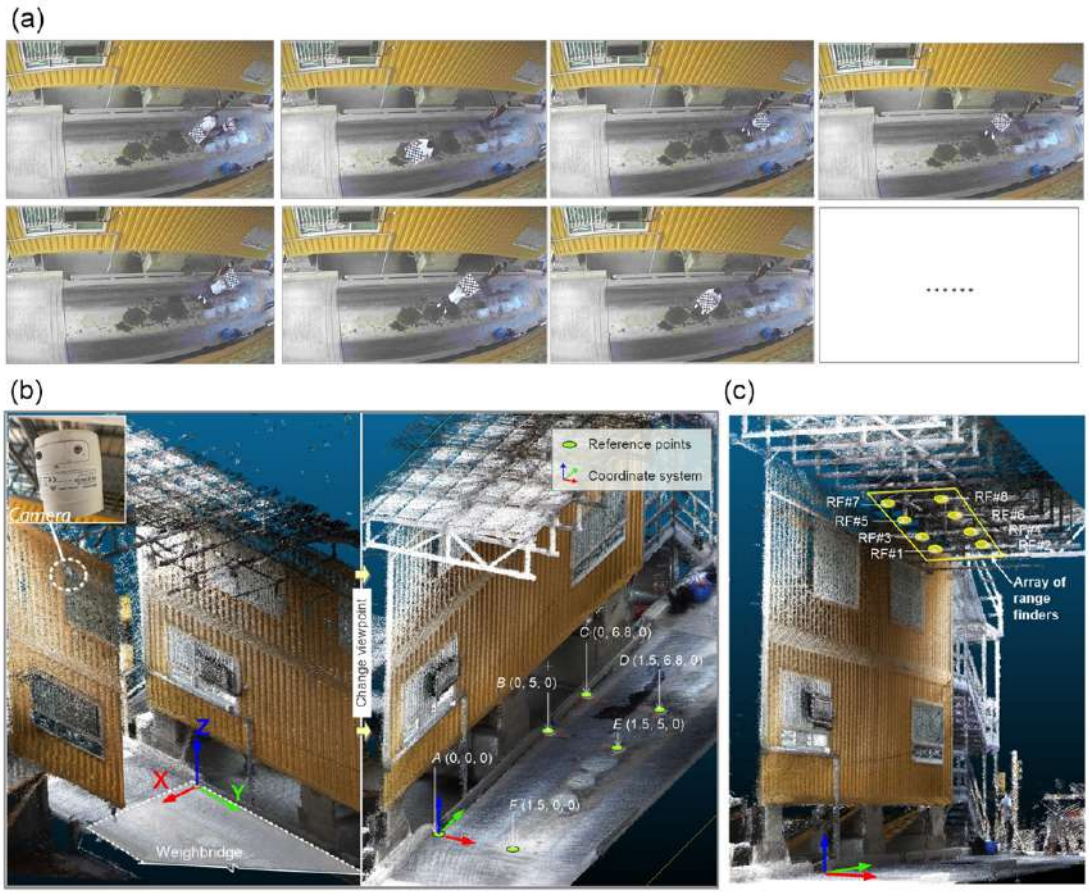

Fig. 3. (a) Example photos of checkboard pattern used for camera calibration Point cloud; (b) The created global coordinate system and some of the selected reference points for camera extrinsic parameter calculation; (c) Layout of ranger finders on top of the toll gate entrance.

(1) Intrinsic parameters

The camera to be calibrated is a DS-2CD2025FWD-IHONG KONG 4mm from the Hikvision Digital Technology. The camera is installed around 6 m above the ground at the tollgate of an off-site sorting facility in HONG KONG, as shown in Fig. 3 (b). We used the Zhang's method (Zhang, 2000) to calibrate the camera. As the camera is relatively distant from the ground, the checkboard pattern was printed on an A2 paper to ensure its clear presence on the corresponding photos. Fig. 3 (a) presents part of the collected photos, which were then input to the MatLab toolbox for calibration. The second column of Table 1 lists the output intrinsic parameters.

Table 1. Details information of the calibrated camera.

Model	Intrinsic parameters	Extrinsic parameters
DS-2CD2025FWD-IHONG KONG 4mm	• Distortion coefficients: [-0.523, 0.278, 0, 0] • Intrinsic matrix: $\begin{bmatrix} 1530.2 & 0 & 0 \\ 0 & 1550.5 & 0 \\ 843.3 & 635.1 & 1 \end{bmatrix}$	• Position (mm): [3151, 575, 6455] • Orientation matrix: $\begin{bmatrix} 0.185 & 0.924 & 0.336 \\ 0.9654 & -0.107 & -0.238 \\ -0.184 & 0.369 & -0.911 \end{bmatrix}$

(2) Extrinsic parameters

To calculate the camera pose, i.e., extrinsic parameters, a global coordinate system needs to be set up first. As shown in Fig. 3 (b), we use the upper-left corner of the weighbridge as origin. The system follows the basic rules of Cartesian coordinates, and its X and Y axes are respectively along the short and long side of the weighbridge. On the weighbridge, we measured the actual coordinates of around 8 points in the coordinate system. With these reference points, the extrinsic parameters of the camera can be obtained by solving the classic Perspective-n-Points (PnP) problem in photogrammetry. The third column in Table 1 lists the derived position and orientation matrix of the camera.

3.3.2. Truck bucket boundary detection

Different from irregular heterogeneous CDW dumps, a truck bucket is a hollow cubic box with explicit regularity. Thus, it is sensible to first detect and measure the boundaries of truck buckets, and then estimate volumes of CDW contained inside. Based on the semantic maps provided by CDW semantic segmentation in Section 3.1, the procedure of bucket boundary detection includes four steps, i.e., distortion rectification, candidate point detection, line fitting and line refinement.

(1) Distortion rectification

As the camera has a wide-angle lens, the radial distortion is too significant to neglect. The camera intrinsic parameters can be used to rectify the distortion. Fig. 4 shows photos of an identical scene before and after distortion rectification, from which it can be observed that the bended bucket boundaries have been successfully undistorted to straight lines. The same rectification is also applied to the semantic maps as they will be used for line detection in later steps.

Fig. 4. Comparison between photos before and after distortion rectification.

(2) Candidate point detection

The bucket boundaries comprise four lines extending along its edges. Facing the direction of truck front cabin, they can be defined as front line l_{front}, left line l_{left}, rear line l_{rear}, and right line l_{right}, respectively, as shown by Fig. 5 (a). To obtain mathematical expression of l_{front}, l_{left}, l_{rear}, and l_{right}, we first find collections of candidate points, i.e., $\{P_i^{front}\}$, $\{P_i^{left}\}$, $\{P_i^{rear}\}$, and $\{P_i^{right}\}$, scattered along the four lines. As the camera is in the rear-right direction of the trucks, the rear and right end of the bucket can be detected by detected and reflected in the semantic maps in most cases. Thus, the acquisition of $\{P_i^{rear}\}$ and $\{P_i^{right}\}$ is straightforward: subsample the leftmost pixels (corresponding to the bucket rear end) and bottommost pixels (corresponding to the bucket right end) of areas recognized as truck in the semantic map. As illustrated by the first diagram in Fig. 5 (a), the subsampling is conducted according to a predetermined distance interval (e.g., every 20 pixels), and should avoid starting regions of the leftmost pixels as those pixels are corresponding to the vertical lines of the bucket rear end.

The detection of $\{P_i^{left}\}$ is comparatively more complicated as the payload in the bucket could occlude the bucket left boundary, as illustrated by the second diagram in Fig. 5 (a). The algorithm described in Fig. 6 is adopted to address the issue. First, a similar procedure of iterative subsample is performed to the uppermost pixels, and if the corresponding semantic labels of the subsampled pixel is "Truck", the pixel coordinates [u,v] will be added to the $\{P_i^{left}\}$. After iteration, the number of points in $\{P_i^{left}\}$ will be compared with a preset threshold th (e.g., th = 5). If it is greater than th, either there is no occlusion or the occlusion is not severe enough to hide most of the boundary points. In such case, the $\{P_i^{left}\}$ will be preserved for later line fitting. Otherwise, a compromise solution will be used. The solution assumes that even though the CDW can occlude most area of the left boundary, it will not normally cover the upper-left corner of the bucket and the upper-right area of the truck. Therefore, the "truck" pixels on the upper-left corner and upper-right area can be detected and added to $\{P_i^{left}\}$ as candidate fit points.

The detection of $\{P_i^{front}\}$ is different from the others because the bucket front boundary does not explicitly show in the "Truck" labels of semantic maps, as illustrated by the third diagram in Fig. 5 (a). Bucket boundaries after rectification are straight lines. This empirical observation can be used to detect the front boundary candidate points. First, Hough transform is performed to detect lines in the truck region. Among the detected results, we only consider the vertical (or nearly vertical) ones in the front half of the truck region. The lines closest to the upper-left corner of the image are most likely to the bucket front boundary. Thus, their starting and end points are selected to be included in $\{P_i^{front}\}$.

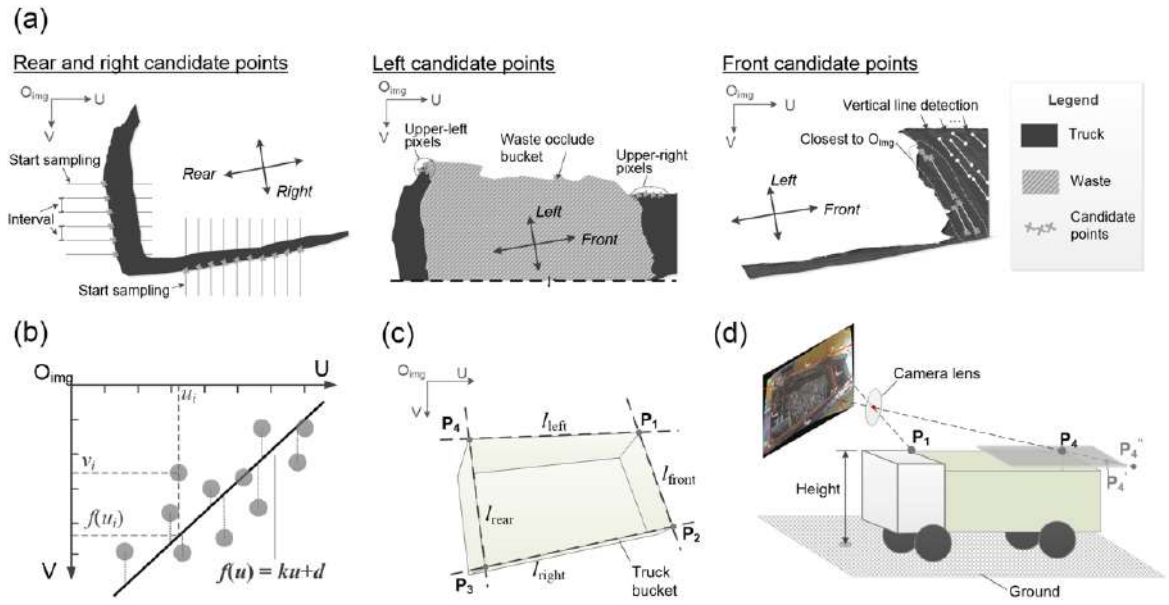

Fig. 5. (a) Schematic diagram for boundary candidate points detection; (b) Line fitting based on least square method; (c) Bucket vertices calculation; (d) Rationale of estimating world coordinates from 2D pixel coordinates.

Algorithm: Bucket left boundary candidate points detection

Input: Semantic segmentation map I_{seg}, e.g., $I_{seg}(u,v)$="truck" means the pixel at (u,v) is part of truck.
Output: Candidate point collection $\{P_i^{left}\}$
Procedure:

```
{P_i^left} = [];                                  # Initialize as an empty matrix.
# Define the searching range and interval along image horizontal axis
u_start = a, u_end = b, inv = c;
# Part 1: Iterative subsampling
for u in range(u_start, u_end, inv) do
    v = min(I_seg(u, :) != None);     # Get the vertical coordinate of upmost non-background point
    if I_seg(u,v) == "truck" then
        {P_i^left}.append([u,v])      # Include in {P_i^left} when the point is "truck"
 end
# Part 2: Handling occlusion
th = d;                                           # Set threshold to determine occlusion
if length({P_i^left}) <= th then
    {P_i^left} = [];                              # Re-initialize as an empty matrix
    [u_ul, v_ul] = searchUpLeft(I_seg);           # Search for upper-left pixel of "truck" region
    {P_i^left}.append([u_ul, v_ul]);
    [u_ur, v_ur] = searchUpRight(I_seg);          # Search for upper-right pixel of "truck" region
    {P_i^left}.append([u_ur, v_ur]);
```

Fig. 6. Pseudo code for detecting bucket left boundary candidate points.

(3) Line fitting

Mathematical expressions of bucket boundaries l_{front}, l_{left}, l_{rear}, and l_{right} can be obtained by fitting them to candidate points $\{P_i^f\}$, $\{P_i^l\}$, $\{P_i^{re}\}$, and $\{P_i^{ri}\}$, respectively. The fitting process is essentially an optimization problem aiming to minimize the square error between the predictive and observation values, as shown by Fig. 5 (b). Suppose k_{f} and d_{f} are the slope and intercept of l_{front} respectively, the fitting of $l_{\text{front}}(k_{\text{f}}, d_{\text{f}})$ is mathematically formulated as:

$$\begin{cases} v = f_{\text{front}}(u) = k_{\text{f}}u + d_{\text{f}} \\ \text{s.t.} \ \min\limits_{(k_{\text{f}}, d_{\text{f}})} \sum_{i=1}^{N} [v_i - f_{\text{front}}(u_i)]^2 \end{cases} \quad (1)$$

Where, (u_i, v_i) is an observed value in the collection of candidate point $\{P_i^f\}(i = 1,2,\ldots, N)$. Similarly, mathematical expressions of the other boundaries can be obtained and denoted as $l_{\text{left}}(k_{\text{l}}, d_{\text{l}})$, $l_{\text{rear}}(k_{\text{re}}, d_{\text{re}})$, and $l_{\text{right}}(k_{\text{ri}}, d_{\text{ri}})$.

(4) Line refinement

The fitted lines need to be further refined to obtain bucket boundaries as precise as possible. The refinement includes two aspects. First, as the camera is at the rear-right direction of the trucks, the original $l_{\text{rear}}(k_{\text{re}}, d_{\text{re}})$ and $l_{\text{right}}(k_{\text{ri}}, d_{\text{ri}})$ only reflects the bottom boundaries of the bucket. They thus should be translated from the bottom to the upper edge. The refinement can be implemented by iteratively changing the values of intercepts (i.e., d_{re} and d_{ri}) until certain termination conditions are met. Second, empirical rules are used to rectify $l_{\text{front}}(k_{\text{f}}, d_{\text{f}})$ and $l_{\text{left}}(k_{\text{l}}, d_{\text{l}})$ that contradict with common sense. For example, as an effect of perspective projection, l_{front} and l_{rear}, and l_{left} and l_{right} should intersect somewhere in the negative V axis and positive U axis direction respectively. If the pattern is not followed, slope k of violating lines will be adjusted accordingly to ensure intersection at the correct direction. In addition, the bucket front line l_{front} should neither be too close nor too far away from the camera viewpoint, and if it happens, its intercept d_{f} would be adjusted to put l_{front} in a proper place (e.g., 1/6 of total truck length to the tuck front end). Fig. 7 shows some examples of detected bucket boundaries after refinement.

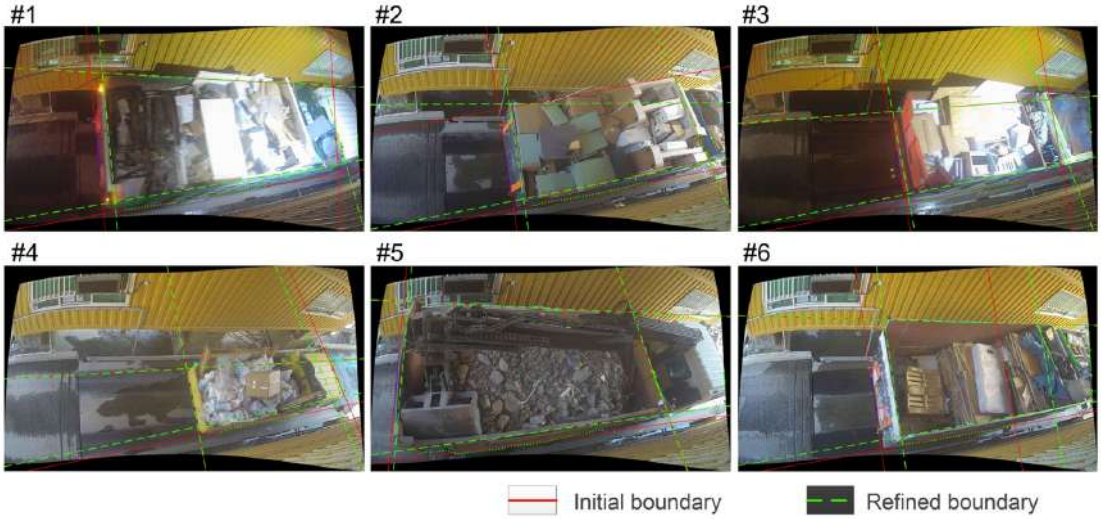

Fig. 7. Examples of detected boundary lines before and after refinement.

3.3.3. Bucket dimension calculation

The pixel coordinates of bucket vertices on respective images are calculated first. As shown by Fig. 5 (c), suppose P_1, P_2, P_3, and P_4 are vertices of four corner of the bucket, their pixel coordinates in the image coordinate system can be obtained by calculating intersections between the boundary lines:

$$\begin{cases} k_{\mathrm{l}}u_1 - v_1 + d_{\mathrm{l}} = 0 \\ k_{\mathrm{f}}u_1 - v_1 + d_{\mathrm{f}} = 0 \\ k_{\mathrm{f}}u_2 - v_2 + d_{\mathrm{f}} = 0 \\ k_{\mathrm{ri}}u_2 - v_2 + d_{\mathrm{ri}} = 0 \\ k_{\mathrm{ri}}u_3 - v_3 + d_{\mathrm{ri}} = 0 \\ k_{\mathrm{re}}u_3 - v_3 + d_{\mathrm{re}} = 0 \\ k_{\mathrm{re}}u_4 - v_4 + d_{\mathrm{re}} = 0 \\ k_{\mathrm{l}}u_4 - v_4 + d_{\mathrm{l}} = 0 \end{cases} \quad (2)$$

Where, (u_1, v_1), (u_2, v_2), (u_3, v_3), and (u_4, v_4) are pixel coordinates of P_1, P_2, P_3, and P_4 respectively.

Let P_j, (u_j, v_j), and (x_j, y_j, z_j) (j = 1,2,3,4) denote any one of the bucket vertices and its corresponding pixel coordinates and world coordinates in the global coordinate system. According to the coordinate transformation model of pin-hole cameras (see Fig. 5 (d)), we have:

$$z_c\begin{bmatrix} u_j & v_j & 1 \end{bmatrix}^T = K[R|T][x_j \; y_j \; z_j \; 1]^T \quad (3)$$

Where K and $[R|T]$ are the camera intrinsic and extrinsic matrix respectively; z_c is the scale factor. Let $CaMAT$ denote the overall camera matrix after the multiplication of K and $[R|T]$, and assume $CaMAT = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}$. Then Eq. (3) can be transformed to the following:

$$\begin{bmatrix} z_c u_j \\ z_c v_j \\ z_c \\ 1 \end{bmatrix} = CaMAT \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} \quad (4)$$

Multiply both ends with the inverse of $CaMAT$, and we can obtain:

$$\begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} = CaMAT^{-1} \begin{bmatrix} z_c u_j \\ z_c v_j \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_c u_j \\ z_c v_j \\ z_c \\ 1 \end{bmatrix} \quad (5)$$

In Eq. (5), the scale factor z_c represents how far the object point corresponding to pixel (u_j, v_j) is away from the camera along the principal optic axis, or the so-called depth. Not knowing z_c, Eq. (5) is unsolvable because numerous points in the real world can be projected onto an identical image pixel, as the P_4' and P_4'' in Fig. 5 (d).

The analysis of trunk dimension profile in Section 3.2 indicates the viability to estimate bucket height based on PGVW of the truck. Leveraging this priori, z_c can be calculated as follows:

$$z_j = f_{\mathrm{truckDim}}(PGVW) \quad (6)$$

$$z_c = \frac{f_{\mathrm{truckDim}}(PGVW) - n_{34}}{n_{31}u_j + n_{32}v_j + n_{33}} \quad (7)$$

Where $f_{\mathrm{truckDim}}(x)$ is the regression model in Fig. 2 (b) that estimates bucket height from PGVW; n_{31}, n_{32}, n_{33} and n_{34} are values in the third row of $CaMAT^{-1}$.

With the calculated z_c, the world coordinates of P_1, P_2, P_3, and P_4 can be easily obtained by via Eq. (5) Finally the calculation of bucket dimensions (i.e., bucket length (BL) and bucket width (BW))

is straightforward, which is simply the Euclidean distance between the bucket vertices:

$$BL = [dist(P_1, P_4) + dist(P_2, P_3)]/2 \quad (8)$$

$$BW = [dist(P_1, P_2) + dist(P_3, P_4)]/2 \quad (9)$$

Where *dist*(*A*, *B*) gives the Euclidean distance between world coordinates of point A and B.

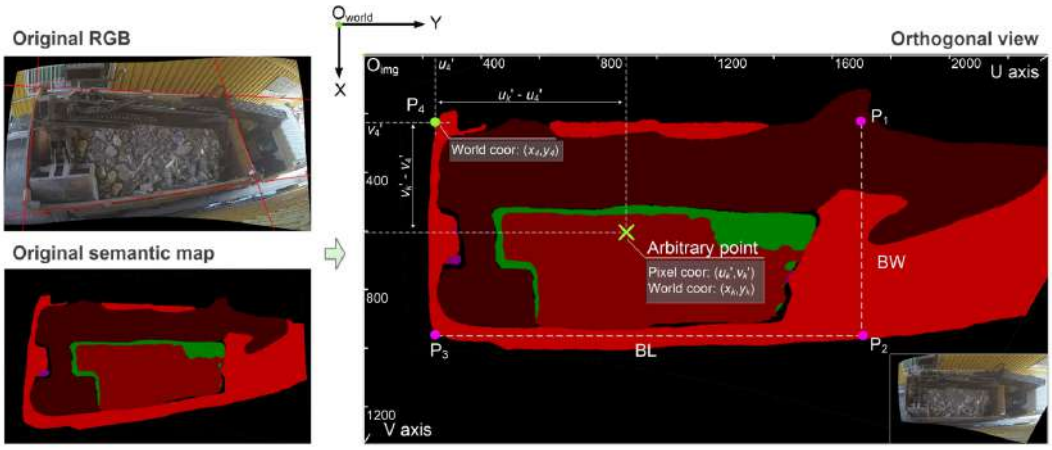

Fig. 8. Orthogonal rectification and subsequent world coordinate calculation for arbitrary points in truck bucket.

3.3.4. Orthogonal rectification

With the pixel coordinates of bucket vertices, the original image and its semantic map can be rectified to an orthogonal view by applying homography transformation (Kim et al., 2019), as shown by Fig. 8. The rectification mitigates the effect of perspective projection, and thus allow dimension measurement of the waste materials in the bucket. Based on the bucket dimensions obtained in previous step, it is viable to derive world coordinates corresponding to arbitrary pixels inside the bucket area, or vice versa:

$$x_k - x_4 = \frac{BW}{v_3' - v_4'}(v_k' - v_4') \quad (10)$$

$$y_k - y_4 = \frac{BL}{u_1' - u_4'}(u_k' - u_4') \quad (11)$$

Where (u_4', v_4') and (u_k', v_k') are pixel coordinates after rectification of P_4 and arbitrary point inside the bucket respectively; and (x_4, y_4) and (x_k, y_k) are their corresponding X and Y components of world coordinates.

3.3.5. Waste volume estimation

The range finders installed at the entrance of the sorting facility provide further information to precisely estimate waste volumes. As shown by Fig. 3 (c), there are totally 8 range finders, denoted by RF#1~#8 respectively. Table 2 lists installation position of the sensors in relative to the global coordinate system. The range finders have been calibrated in prior so that the data given by the sensors directly represents the distance of targets from the weighbridge.

Table 2. Range finder installation position.

ID	RF#1	RF#2	RF#3	RF#4	RF#5	RF#6	RF#7	RF#8
X (m)	1.0	2.0	1.0	2.0	1	2	1	2
Y (m)	6.0	6.0	4.5	4.5	3	3	1.5	1.5
Z (m)	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5

Let $\Omega(P_4, BL, BW)$ denote a rectangular region with P4 as its upper-left corner and BL and BW as length and width of along the Y and X axis respectively. It represents the horizontal space covered by the truck bucket. Then range finders that fall within this range can be refer to as $M = \{\mathrm{RF}\#k | \mathrm{RF}\#k \in \Omega(P_4, BL, BW)\}$. Thus, the waste depth corresponding to RF#k is calculated by:

$$D_{\mathrm{RF}\#k} = H_{\mathrm{RF}\#k} - H_0 \tag{11}$$

Where $H_{\mathrm{RF}\#k}$ is the height data collected by sensor RF#k, and H_0 is the height of the bucket bottom, which has been pre-stored in the record of the facility system.

With a collection of $\{D_{\mathrm{RF}\#k}\}$, the waste depth of arbitrary points in $\Omega(P_4, BL, BW)$ can be interpolated, as demonstrated by Fig. 9 (a). The $\Omega(P_4, BL, BW)$ is turn into mesh grid according to a preset interval (e.g., Δd = 150 mm) to determine positions of interpolation point PI_{ij}.

$$\begin{cases} x_{\mathrm{PI}ij} = x_4 + \Delta d \times i, \ (i \in N^*, i \le \frac{BW}{\Delta d}) \\ y_{\mathrm{PI}ij} = y_4 + \Delta d \times j, \ (j \in N^*, i \le \frac{BL}{\Delta d}) \end{cases} \tag{11}$$

Where ($x_{\mathrm{PI}ij}$, $y_{\mathrm{PI}ij}$) is the coordinates of PI_{ij}, and N^* denote a set of positive integers. The predictive waste depth $\widehat{D}_{\mathrm{PI}ij}$ at after interpolation is:

$$\widehat{D}_{\mathrm{PI}ij} = \sum_M \lambda_k D_{\mathrm{RF}\#k} \tag{12}$$

Where $M = \{\mathrm{RF}\#k | \mathrm{RF}\#k \in \Omega(P_4, BL, BW)\}$, and λ_k is the weight factor of observed waste depth at RF#k on the predictive waste depth. Different interpolation methods calculate λ_k differently. With Kriging interpolation, for example, λ_k is determined by the relative distance of predictive points, spatial layout of observation points, and other factors.

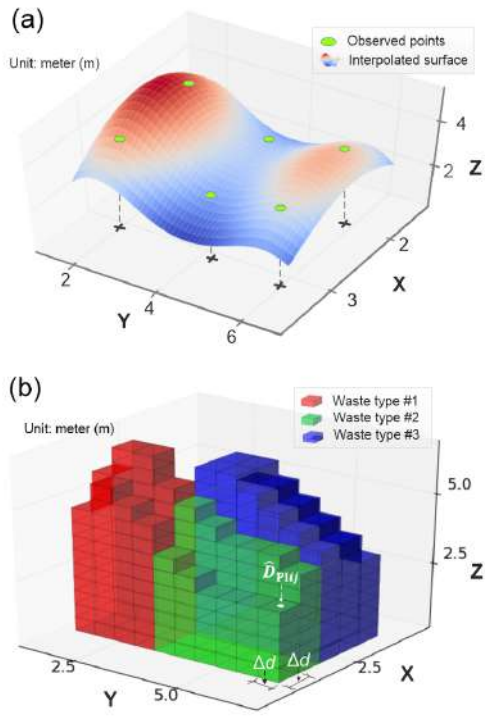

Fig. 9. Schematic diagrams of (a) interpolation and (b) waste volume estimation.

Via Eq. (10) and (11), the type of waste material corresponding to PI_{ij} can also be obtained from the semantic map, and be denoted by $T_{\mathrm{PI}ij}$. Hence, geometry and semantic information for all interpolation points can be obtained as $\{x_{\mathrm{PI}ij}, y_{\mathrm{PI}ij}, \widehat{D}_{\mathrm{PI}ij}, T_{\mathrm{PI}ij}\}$. Assuming that waste materials at the bottom are consistent with the surface materials, the volume of any specific type or the entire dump of CDW can be then estimated. This is done by simple accumulation of column volumes corresponding to certain waste materials, as shown by Fig. 9 (b). Eq. (13) shows mathematical formula to calculate the volume of a waste dump.

$$V_{all} = \sum_{i=1}^{I_{max}} \sum_{j=1}^{J_{max}} \widehat{D}_{\mathrm{PI}ij}\, \Delta d^2 \tag{13}$$

Where $I_{max} = \left\lfloor \frac{BW}{\Delta d} \right\rfloor$ and $J_{max} = \left\lfloor \frac{BL}{\Delta d} \right\rfloor$, and $\lfloor x \rfloor$ gives the floor integer of x.

4. Results and analysis

This section analyzes the results of the proposed algorithms in an off-site sorting facility in HONG KONG. The algorithms were implemented with MatLab and Python on an OptiPlex 7080 computer with Intel(R) Core (TM) i7-10700 CPU and NVIDIA GeoForce RTX 2070 SUPER GPU.

4.1. Results of bucket dimension calculation

We first evaluated bucket dimensions calculated by the proposed algorithm. As mentioned in Section 3.2, the width of dump buckets is supposed to be roughly the same with cabin width so as to comply relevant regulations. According to truck models and their PGVW, the width data of 78 trucks passing through the facility entrance was collected from official websites of corresponding

manufacturers. Fig. 10 (a) shows a comparison between the estimated bucket width and the observed ground-truth values, and the relative error for each sample is represented by a gray bar at the bottom. The general trend of the predictive values is consistent with the observed values. As summarized by Fig. 10 (b), the average relative error is 0.065, which is satisfactory.

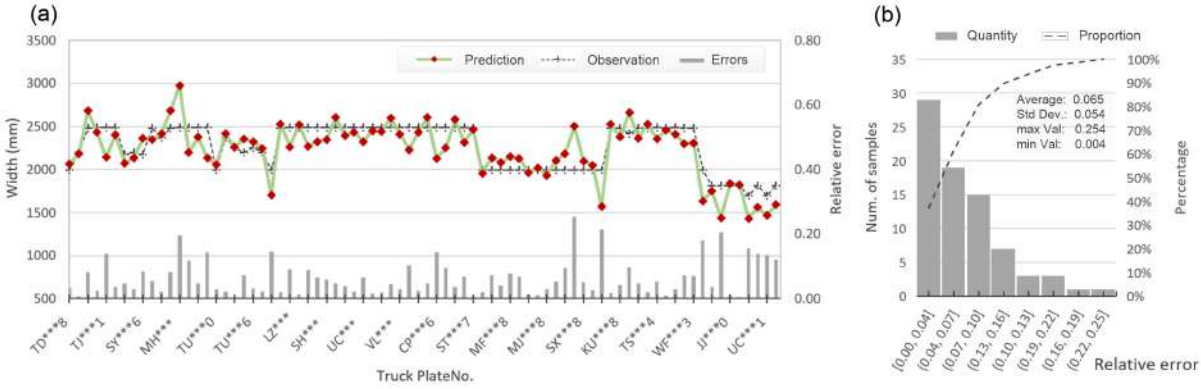

Fig. 9. Evaluation of bucket dimension calculation performance: (a) Comparison between predictive and observed bucket width of 78 investigated trucks; (b) Histogram distribution of relative errors.

4.2. Results of waste volume estimation

The satisfactory precision of bucket dimensions calculation lays the foundation for 3D reconstruction of truck-loaded waste and further volume estimation. Fig. 11 shows some example results. The first column is the original RGB images, and on top of each image the corresponding transaction number, truck plate number, and the estimated bucket dimensions are presented. In the second column are the reconstructed 3D geometric models with real texture applied onto the waste dump surface. In the last column are the reconstructed 3D semantic models with different CDW types encoded by different colors. The estimated waste volumes are also listed on the right hand side.

It can be found that the estimated geometry, semantic types, and volumes of the wastes are aligned with the observations from RGB photos. Note that there might be waste areas covered by the grippers mounted to the truck buckets. The semantic properties of these areas are replaced with the waste types of their nearest voxels when estimating the CDW volumes (An example is the fourth row of Fig. 11). The average processing time of the proposed algorithm is 3.3 s per image, including 3.22 s for bucket dimensions calculation and 0.08 s for volume estimation.

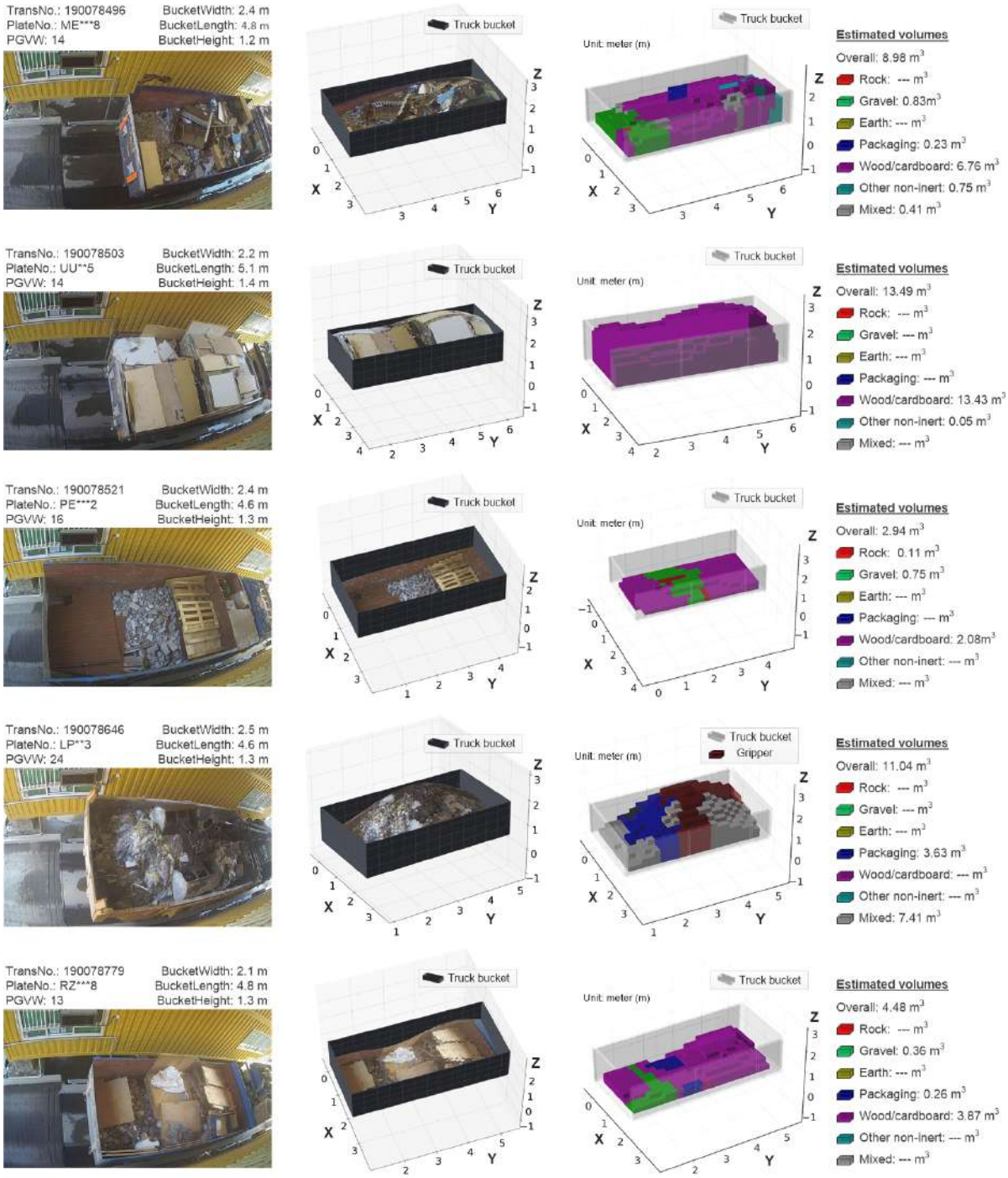

Fig. 11. 3D geometric and semantic models reconstructed from monocular vision.

5. Case study

The proposed algorithm was applied to analyze 2,914 transaction records happed in October 2019 at one lane of the sorting facility. Fig. 12 (a) demonstrates the daily CDW volumes arriving at the toll gate of the facility. It can be clearly observed that the waste volumes fluctuate periodically according to a cycle of roughly every seven days. Among the seven material types, wood/cardboard accounts for the largest proportion, followed by the "Mixed" waste which is a clutter combination of inert and non-inert materials. The three inert

materials, i.e., rock, gravel and earth, have the least proportion, possibly because the waste haulers elect to primarily send such waste types to the public fills for lower disposal price.

The facility toll gate received around 800.0 m^3 CDW per day, of which about 10.8 m^3 were rejected for not meeting the requirements of inert content proportion (HKEPD, 2019). Fig. 12 (b) and (c) show volume distribution over diverse materials accepted and rejected by the facility, respectively. Most of the rejected materials are non-inert wastes such as wood and packaging, implying the rejected waste dumps indeed have not reached the required proportion of inert materials. Meanwhile, non-inert materials also account for a great proportion in the accepted wastes. This is in line with the HONG KONG Audit Commission report (Audit Commission, 2016), which criticized that many truck loads accepted by the facility not actually meeting the inert-content requirements.

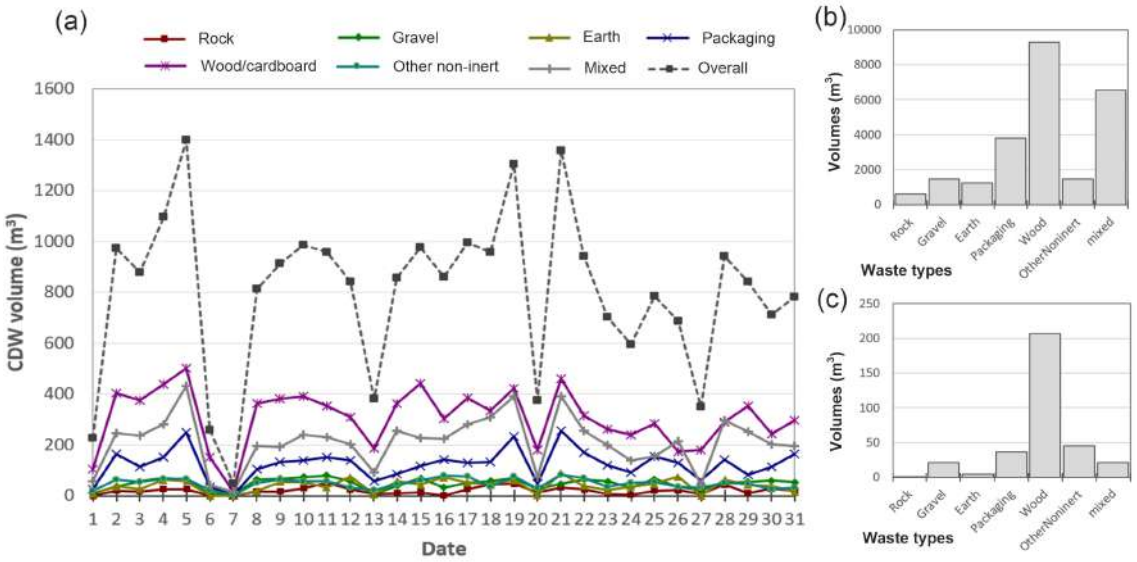

Fig. 12. Statistics of estimated waste volume in October 2019: (a) Amount of daily incoming waste dumps; (b) Distribution over different materials for the accepted waste dumps; (c) Distribution over different materials for the rejected waste dumps.

6. Discussion

"A picture is worth a thousand words". From an information technology perspective, the widespread adage reflects how immense magnitude of information can be conveyed by a single image. Indeed, a top-down photo of truck-loaded waste dump not only tells much about its composited material components, but also implicitly encodes information of its geometry. However, without effective algorithms and proper use of domain knowledge, such abundant information will remain undiscovered treasure inaccessible to those in need. This paper presents a monocular vision algorithm that integrates semantic segmentation and domain knowledge of truck dimension profiles to reconstruct 3D geometric and semantic models of CDW dumps from single images. The algorithm can estimate truck bucket dimensions with a relative error of 0.065, while only consumes 3.3 s to process each image. It

only requires common hardware configurations (e.g., camera and weight bridge) at existing waste disposal facilities, and can estimate CDW volumes in sufficient granularity without physical separation. The effectiveness of the algorithm makes it a promising solution to address the conundrum of balancing between "accuracy" and "operability" in waste quantification.

The algorithm is primarily implemented and validated in the HONG KONG context. However, it can be easily extended to other places of the world by simple recalibration of its core modules. For example, different countries/regions generate different types of construction wastes; thus, the CDW semantic segmentation model needs to be re-train or finetuned accordingly based on the specific contexts. In addition, the prevalent manufacturers and truck dimensions in other places might be different from those in HONG KONG, making it necessary to re-calibrate the "Cabin height-PGVW" regression model.

The broad concept of the proposed algorithm can also be generalized to other industries where similar truck payload amount estimation requirements and hardware setup are shared. An exemplary scenario is infrastructure construction, in which volumetric measurement is an important task for quality control. Similar systems can be set up at the entrance of such infrastructure projects. Then the proposed algorithm can be deployed to calculate volumes of incoming materials such as asphalt concrete and rockfill for compliance check or progress monitoring.

7. Conclusions

How to accurately quantify construction and demolition waste (CDW) with easy and operable approaches in field is an unsolved dilemma. Physically segregating the waste materials can certainly provide precise amount information, but is onerous and time-consuming. To tackle the dilemma, this paper presents a non-destructive and precise algorithm for truck-loaded CDW volume estimation based on monocular vision. In the proposed algorithm, a deep learning model is trained to extract semantic information of waste types (e.g., rock, gravel, and wood); prior knowledge of truck dimension profiles is used to achieve depth perception, and recover 3D geometry of the waste dumps. The algorithm achieved a relative error of 0.065 in estimating bucket dimensions while consumed 3.3 s to process each image. A case study was carried out to analyze 2,914 waste dumps received by an off-site sorting facility in Hong Kong during October 2019. It was found that the facility entrance received around 800.0 m^3 CDW per day, of which about 10.8 m^3 were rejected. Among the accepted materials, the non-inert wood/cardboard accounts for the highest proportion, revealing many CDW dumps accepted by the facility did not actually comply with the inert-content requirements.

Despite the demonstrated efficacy, future studies are recommended to address the following limitations. First, the algorithm might fail to work with oversize trucks of which the buckets

cannot be entirely covered by the camera field of view. Thus, further improvement is required to infer the location of invisible bucket boundaries from the images. Second, although materials on waste dump surface provide important clues on the types of underneath waste materials, they are not necessarily identical. Integration of physical features such as weight and bulk density of the waste dumps might be useful to further distinguish the waste materials beneath the surface. Such integration can be realized based on machine learning techniques such as Bayesian inference.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is jointly supported by the Strategic Public Policy Research (SPPR) Funding Scheme (Project No.: S2018.A8.010.18S) and the Environment and Conservation Fund (ECF) (Project No.: ECF 2019-111) of the Government of the Hong Kong Special Administrative Region.

References

Aber, J.S., Marzolff, I., Ries, J.B., 2010. Chapter 3 - Photogrammetry, in: Aber, J.S., Marzolff, I., Ries, J.B. (Eds.), Small-Format Aerial Photography. Elsevier, Amsterdam, pp. 23-39.

Aguzzi, J., Chatzievangelou, D., Marini, S., Fanelli, E., Danovaro, R., Flögel, S., Lebris, N., Juanes, F., De Leo, F.C., Del Rio, J., Thomsen, L., Costa, C., Riccobene, G., Tamburini, C., Lefevre, D., Gojak, C., Poulain, P.-M., Favali, P., Griffa, A., Purser, A., Cline, D., Edgington, D., Navarro, J., Stefanni, S., D'Hondt, S., Priede, I.G., Rountree, R., Company, J.B., 2019. New High-Tech Flexible Networks for the Monitoring of Deep-Sea Ecosystems. ENVIRON SCI TECHNOL 53, 6616-6631. 10.1021/acs.est.9b00409.

Arumugam, S.K., Muhamad, R., Yahya, K., 2020. Mapping of Construction Waste for Eco-Costs Per Value Ratio (Evr) Index Using Google My Maps in Shah Alam, Malaysia. IOP Conference Series: Materials Science and Engineering 849, 012046. 10.1088/1757-899x/849/1/012046.

Asadi, K., Ramshankar, H., Noghabaei, M., Han, K., 2019. Real-Time Image Localization and Registration with Bim Using Perspective Alignment for Indoor Monitoring of Construction. J COMPUT CIVIL ENG 33, 04019031. Artn 04019031 10.1061/(Asce)Cp.1943-5487.0000847.

Audit Commission, 2016. Management of Abandoned Construction and Demolition Materials, Chapter 4 of the Director of Audit's Report No. 67.

Fatta, D., Papadopoulos, A., Avramikos, E., Sgourou, E., Moustakas, K., Kourmoussis, F., Mentzis, A., Loizidou, M., 2003. Generation and Management of Construction and Demolition Waste in Greece—an Existing Challenge. Resources, Conservation and Recycling 40, 81-91. https://doi.org/10.1016/S0921-3449(03)00035-1.

Gavilan, R.M., Bernold, L.E., 1994. Source Evaluation of Solid Waste in Building Construction. J CONSTR ENG M 120, 536-552. doi:10.1061/(ASCE)0733-9364(1994)120:3(536).

Hamdan, A.-H., Taraben, J., Helmrich, M., Mansperger, T., Morgenthal, G., Scherer, R.J., 2021. A Semantic Modeling Approach for the Automated Detection and Interpretation of Structural Damage. AUTOMAT CONSTR 128, 103739. https://doi.org/10.1016/j.autcon.2021.103739.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-Cnn, Proceedings of the IEEE international conference on computer vision, pp. 2961-2969.

HKEPD, 2019. Management of Abandoned Construction and Demolition Materials. https://www.aud.gov.Hong Kong/pdf_e/e67ch04sum.pdf (Accessed 20 November 2020).

HONG KONGTD, 2019. Code of Practice for the Loading of Vehicles. https://www.td.gov.Hong_Kong/filemanager/en/publication/cop_loading_of_vehicles_eng.pdf (Accessed August 15 2021).

Hoang, N.H., Ishigaki, T., Kubota, R., Tong, T.K., Nguyen, T.T., Nguyen, H.G., Yamada, M., Kawamoto, K., 2020. Waste Generation, Composition, and Handling in Building-Related Construction and Demolition in Hanoi, Vietnam. Waste Management 117, 32-41. https://doi.org/10.1016/j.wasman.2020.08.006.

Isailović, D., Stojanovic, V., Trapp, M., Richter, R., Hajdin, R., Döllner, J., 2020. Bridge Damage: Detection, Ifc-Based Semantic Enrichment and Visualization. AUTOMAT CONSTR 112, 103088. https://doi.org/10.1016/j.autcon.2020.103088.

Katz, A., Baum, H., 2011. A Novel Methodology to Estimate the Evolution of Construction Waste in Construction Sites. Waste Management 31, 353-358. https://doi.org/10.1016/j.wasman.2010.01.008.

Kaza, S., Yao, L., Bhada-Tata, P., Van Woerden, F., 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. World Bank Publications.

Kim, D., Liu, M., Lee, S., Kamat, V.R., 2019. Remote Proximity Monitoring between Mobile Construction Resources Using Camera-Mounted Uavs. AUTOMAT CONSTR 99, 168-182. https://doi.org/10.1016/j.autcon.2018.12.014.

Lam, P.T.I., Yu, A.T.W., Wu, Z., Poon, C.S., 2019. Methodology for Upstream Estimation of Construction Waste for New Building Projects. J CLEAN PROD 230, 1003-1012. https://doi.org/10.1016/j.jclepro.2019.04.183.

Lau, H.H., Whyte, A.A., Law, P.L., 2008. Composition and Characteristics of Construction Waste Generated by Residential Housing Project. INT J ENVIRON RES 2, 261-268. 10.22059/ijer.2010.202.

Li, E.Y., 2020. Witnessing the Progression in Semantic Segmentation: Deeplab Series from V1 to V3+. https://towardsdatascience.com/witnessing-the-progression-in-semantic-segmentation-deeplab-series-from-v1-to-v3-4f1dd0899e6e (Accessed April, 7 2021).

Li, H., Yao, J., Bazin, J., Lu, X., Xing, Y., Liu, K., 2018. A Monocular Slam System Leveraging Structural Regularity in Manhattan World, 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2518-2525.

Lim, R., La, H., Sheng, W., 2014. A Robotic Crack Inspection and Mapping System for

Bridge Deck Maintenance. Automation Science and Engineering, IEEE Transactions on 11, 367-378. 10.1109/TASE.2013.2294687.

Liu, D., Cui, B., Liu, Y., Zhong, D., 2013. Automatic Control and Real-Time Monitoring System for Earth–Rock Dam Material Truck Watering. AUTOMAT CONSTR 30, 70-80. https://doi.org/10.1016/j.autcon.2012.11.007.

Liu, Y., Aggarwal, J.K., 2005. 3.12 - Local and Global Stereo Methods, in: Bovik, A.L. (Ed.), Handbook of Image and Video Processing (Second Edition). Academic Press, Burlington, pp. 297-308.

Llatas, C., 2011. A Model for Quantifying Construction Waste in Projects According to the European Waste List. Waste Management 31, 1261-1276. https://doi.org/10.1016/j.wasman.2011.01.023.

Loadscan Ltd., 2021. Load Volume Scanner (Lvs) System. https://www.loadscan.com/load-volume-scanner/ (Accessed August 19 2021).

Lu, W., Chen, J., Xue, F., 2021a. Using Computer Vision to Recognize Composition of Construction Waste Mixtures: A Semantic Segmentation Approach. Resources, Conservation & Recycling Manuscript submitted for publication.

Lu, W., Yuan, L., Xue, F., 2021b. Investigating the Bulk Density of Construction Waste: A Big Data-Driven Approach. Resources, Conservation and Recycling 169, 105480. https://doi.org/10.1016/j.resconrec.2021.105480.

NSWEPA, 2020. Protection of the Environment Operations (Waste) Regulation 2014. https://legislation.nsw.gov.au/view/html/inforce/current/sl-2014-0666#statusinformation (Accessed August 19 2021).

NSWEPA., 2018. Waste Levy Guidelines. https://www.epa.nsw.gov.au/-/media/epa/corporate-site/resources/wasteregulation/181272-waste-levy-guidelines.pdf (Accessed August 19 2021).

Ost, A.D., Wu, T., Höschen, C., Mueller, C.W., Wirtz, T., Audinot, J.-N., 2021. 4d Surface Reconstructions to Study Microscale Structures and Functions in Soil Biogeochemistry. ENVIRON SCI TECHNOL 55, 9384-9393. 10.1021/acs.est.1c02971.

Pollefeys, M., 2017. Computer Vision for Mixed Reality (Microsoft Hololens). https://www.youtube.com/watch?v=jbnoXS_rlTc (Accessed August 17 2021).

Prairie Village, K., 1998. Characterization of Building-Related Construction and Demolition Debris in the United States. The US Environmental Protection Agency Municipal and Industrial Solid Waste Division Office of Solid Waste.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234-241.

Singh, A., 2020a. How Neural Nets, Computer Vision and Autonomous Vehicles Are Related. https://towardsdatascience.com/stereo-vision-neural-nets-and-demand-in-autonomous-vehicles-6a9b6e6de41c (Accessed August 17 2021).

Singh, R., 2020b. Measure Anything with Your Phone Camera without Touching It. https://gadgetstouse.com/blog/2020/05/13/measure-anything-with-your-phone-camera-without-touching-it/ (Accessed August 17 2021).

The Center for Photogrammetric Training, 2008. History of Photogrammetry. https://ibis.geog.ubc.ca/courses/geob373/lectures/Handouts/History_of_Photogrammetr

y.pdf (Accessed August 15 2021).
Torr, P.H., Zisserman, A., 2000. Mlesac: A New Robust Estimator with Application to Estimating Image Geometry. COMPUT VIS IMAGE UND 78, 138-156.
Villoria Sáez, P., del Río Merino, M., Porras-Amores, C., 2011. Estimation of Construction and Demolition Waste Volume Generation in New Residential Buildings in Spain. WASTE MANAGE RES 30, 137-146. 10.1177/0734242X11423955.
Visco India, 2020. Standalone Weigh Enforcement System. https://www.viscoindia.com/ (Accessed August 19 2021).
Xue, F., Lu, W., Chen, K., 2018. Automatic Generation of Semantically Rich as-Built Building Information Models Using 2d Images: A Derivative-Free Optimization Approach. COMPUT-AIDED CIV INF 33, 926-942. https://doi.org/10.1111/mice.12378.
Zaheer, A., Rashid, M., Riaz, M.A., Khan, S., 2018. Single-View Reconstruction Using Orthogonal Line-Pairs. COMPUT VIS IMAGE UND 172, 107-123. https://doi.org/10.1016/j.cviu.2017.11.014.
Zhang, Z., 2000. A Flexible New Technique for Camera Calibration. IEEE T PATTERN ANAL 22, 1330-1334. 10.1109/34.888718.
Zhong, R.Y., Peng, Y., Xue, F., Fang, J., Zou, W., Luo, H., Thomas Ng, S., Lu, W., Shen, G.Q.P., Huang, G.Q., 2017. Prefabricated Construction Enabled by the Internet-of-Things. AUTOMAT CONSTR 76, 59-70. https://doi.org/10.1016/j.autcon.2017.01.006.