

Photo-based Automatic 3D Reconstruction of Train Accident Scene

Zhao Tang¹, Yinyu Nie¹, Jian Chang², Jian Jun Zhang^{1, 2}, Feng Jia Liu¹

¹State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu, China

²National Centre for Computer Animation, Bournemouth University, Poole, UK

Abstract: Railway accidents place significant demands on resources and support from departments of railway emergency management. Once an accident occurs, an efficient incident rescue plan needs to be given as early as possible to reduce the loss of life and property. However, in railway system, most related departments currently face a challenge in drawing up rescue scheme effectively and accurately with insufficient information collected from the scene of a train accident. To assist with the rescue planning, we propose a framework which can rapidly and automatically construct a 3D virtual scene of a train accident by utilizing the photos of the accident spot. The framework uses a hybrid 3D reconstruction technology to extract the position and pose information of carriages involved in an accident. It adopts a geographic information system (OSG Earth) and a 3D visualization engine to model and display the landscapes and buildings in a train accident. In order to assess and validate our prototype, we quantitatively evaluate our main algorithm and demonstrate the usage of our technology with two case studies including a simulated scene with in-lab setting and a real train derailment scene from on-site pictures, and both results are accountable with high accuracy and represent the ability of timely modelling and visualisation of a train accident scene.

Keywords: Train accident; scene reconstruction; accident rescue; photo-based modelling; railway safety;

1 Introduction

Despite the fact that the railway industry attempts to ensure the safety, many train accidents still happen worldwide every year and there is approximately one severe accident every 2 years¹. Once a train accident occurs, it is important to gain a full understanding and comprehensive awareness of the accident situation, which helps

achieve an optimal decision-making of train accident rescue. At the same time, critical information about the accident, like the positions and poses of vehicles in the scene, needs to be acquired as soon as possible. Currently, railway rescue operation mainly depends on visual observations and survey by the rescuers to make decision, suffering from the lack of detailed information of an accident. In addition, taking lessons from the previous accidents is an effective way to prevent the accident happening again. What we can learn from an accident depends on our understanding to the cause and phenomenon of the accident². Reconstruction of the accident scene will improve our understanding with visual presentation and 3D location information, developing insight about the cause of the accident.

Apparently, a realistic 3D virtual accident scene can provide meaningful materials and as much relevant information as possible about the train accident, which can not only help with decision-making for rescue but also analyze the cause of the accident. However, rapidly constructing an accurate 3D digital accident scene is challenging because it involves a large number of different objects, such as vehicles, tracks and buildings, as well as the large-scale surrounding environment, which requires great modelling efforts.

In recent decades, computer vision technology has been applied to railway industry³⁻⁶. These works have mainly focused on operational inspection rather than the reconstruction of a train accident scene. A popular method in computer vision, named photo-based (or image-based) modelling, provides an automatic approach of 3D objects modelling by deriving geometric information from photos⁷. This technique has been applied in various disciplines, including virtual reality or augmented reality⁸, architecture^{9, 10}, earth sciences¹¹, even in forensic infography¹² and archaeometry¹³ etc. However, to our best knowledge, literatures and practices of applying the photo-based 3D modelling technique to the railway industry for assisting accident rescue are rarely seen.

Most photo-based 3D modelling techniques can sketch outlines of objects in a 3D space, but it resorts to multi-view inputs of the objects. It normally needs at least two photos from different camera angles to calculate the spatial position of a given point. This is the basis of almost all multi view-based 3D reconstruction methods. In an example, the reconstruction of a bridge used over 100 photos from different view angles⁹. However for a rescue operation of a train accident, it may not be always practical to take so many photos with precise settings around the vehicles involved in

the accident because some landscapes may prevent taking photos from a particular angle. Moreover, in many situations, the modellers, who are the beholders of the special knowledge on how to set the camera, just cannot be present at the scene of the accident and take photos. In addition, the multi view-based approaches may not be adequate to construct digital scenes of train accidents due to the fact that the response time of the railway accident is generally very short, not more than a few hours. We cannot acquire images with 360 degree of view that contain all accident objects in such a short response time.

Since 2000, researchers have been attempting to develop the single view-based reconstruction method which can produce convincing results^{14, 15}. But, the existing single view solution may degenerate the reconstruction with possible loss of information¹⁵.

Besides the photo-based method, other surveying methods such as aerial photogrammetry and laser scanning can also provide accurate measurement and rich information. But most of them rely on expensive equipment and require users to have relevant skills and knowledge of surveying theory, which may not be always available for the reconstruction of a train accident.

In this paper, we propose a framework to construct a 3D virtual accidental scene automatically. This framework uses a single view photo-based method to measure the position and pose of a train coach in the accident, and then utilizes 3D geographic information system (GIS) and visualization technology to construct the environment where the accident occurs.

The remainder of the paper is structured as follows. The overview of this framework is presented in section 2. Section 3 describes the theory of our framework in photo-based automatic 3D reconstruction. It is worth mentioning that all the inputs we use are images captured at the accident scene. Section 4 describes the automatic construction of environment, including terrain, buildings and tracks parts. Finally, two case studies are studied in section 5. It shows the photo-based 3D reconstruction of a train accident scene with scaled models in lab and a real derailment accident occurring in Philadelphia on May 12, 2015. Section 6 concludes the paper.

2 Overview

As shown in Figure 1, our framework consists of four modules: photo capturing

module, image recognition module, 3D modelling module and 3D visualization module.

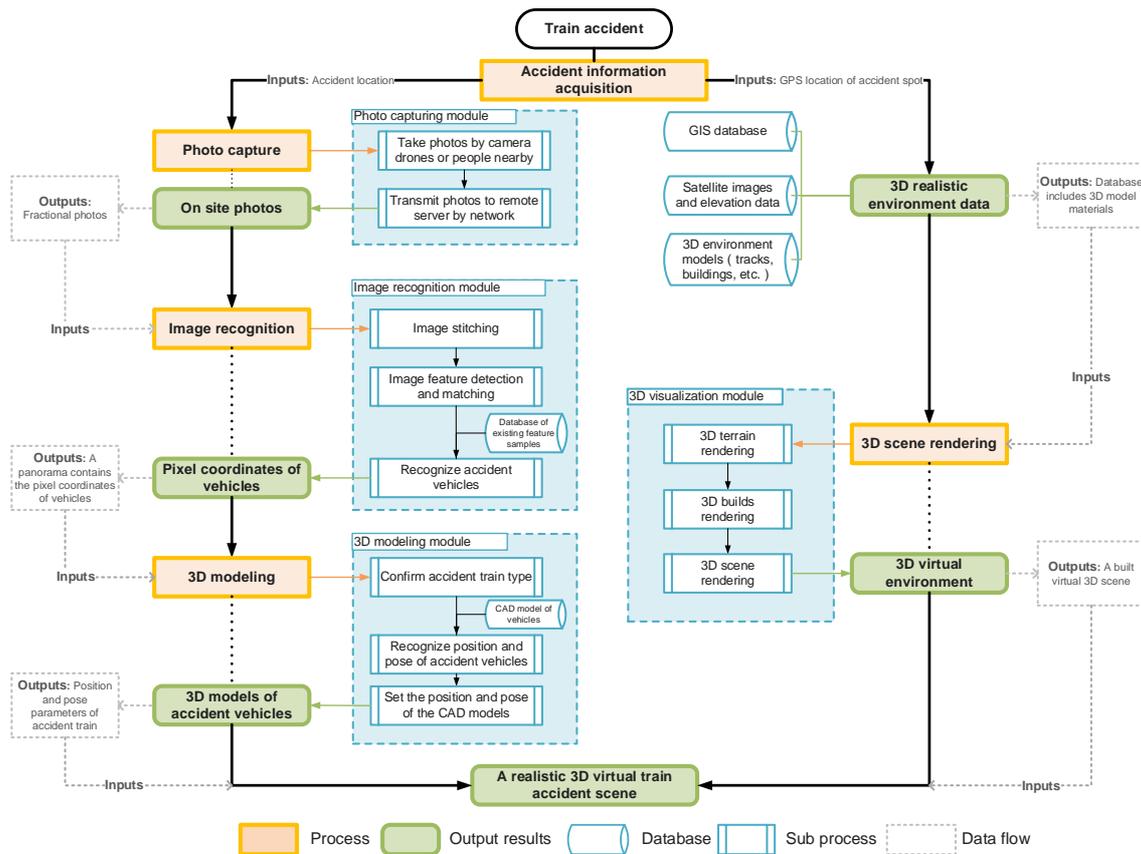


Figure 1. Framework to construct a 3D virtual train accident scene

The primary function of photo capturing module is on-site photo taking which can be done by drones or staff nearby. The captured photos are transmitted to back-end server for further processing and are also the main inputs of our framework which need to be provided by end-users.

The objective of image recognition module is to build a panorama that can register primary objects involved in the accident, such as vehicles, train coaches, tracks and buildings. Image alignment and image stitching algorithms are necessary to discover the correspondence and relationships among images. We adopt the image feature detecting and matching method named Scale Invariant Feature Transform (SIFT)¹⁶ to search the 2D image location of vehicles in the panorama, and benefit from its good performance in recognizing the pixel coordinates of objects in panorama and robustness when detecting features in spite of the changes of scale, rotation, blur, illumination and affine. With SIFT, we can locate the coordinates of objects in the 2D image plane.

In 3D modelling module, the spatial pose and position of carriages involved in an accident are measured by minimizing the distance errors in projections. Inspired by

Bundle Adjustment algorithm¹⁷, we resolve the multi-carriages position and pose measuring problem with a non-linear least square expression. It involves perspective-projection calculation and projection-error minimization. Differing from Bundle Adjustment algorithm, a hybrid photo-based 3D modelling method is proposed to extract multiple carriages from a single-view image. Instead of reconstructing the full geometric details from a photo/photos, the hybrid photo-based 3D modelling method only computes the poses and positions of the carriages and describes the detailed geometric information by utilizing pre-existing CAD models which are stored in our database.

3D visualization module consists of three sub-modules (3D terrain modelling, 3D buildings modelling and 3D scene modelling) which render the terrain, buildings and other objects respectively. This module generates a 3D virtual environment by using 3D realistic environment data (GIS database, satellite images, elevation data and environment models). These data are stored in database in advance and integrated into the scene in real-time. Ultimately, by embedding the 3D CAD model of accident vehicles into this virtual environment according to the pose and position generated by 3D modelling module, a realistic 3D virtual train accident scene is reconstructed.

3 Photo-based measurement and 3D modelling for train accident scene

Manually sketching and modelling the 3D models in a train accident would be labor intensive, demanding tedious and repetitive efforts of modelers or researchers. It is also error-prone. In this section, a photo-based measurement and 3D modelling method is proposed to automatically determine the spatial position and pose of multiple carriages in a train accident.

3.1 Image stitching

The objective of image stitching is to find all matching images, then connect sets of image matches to produce a panoramas. We use Lowe's Scale Invariant Feature Transform (SIFT) features for image matching thanks to its better performance in achieving reliable matching of multiple images and handling scale, rotation, blur, and affine changes of image than SURF and PCA-SIFT detector¹⁸. To automatically discover the matching relationships in unordered multi-images, the automatic panorama stitching algorithm proposed by Matthew Brown and David G. Lowe¹⁹ is recommended.

The main flow of this part is shown in Figure 2.

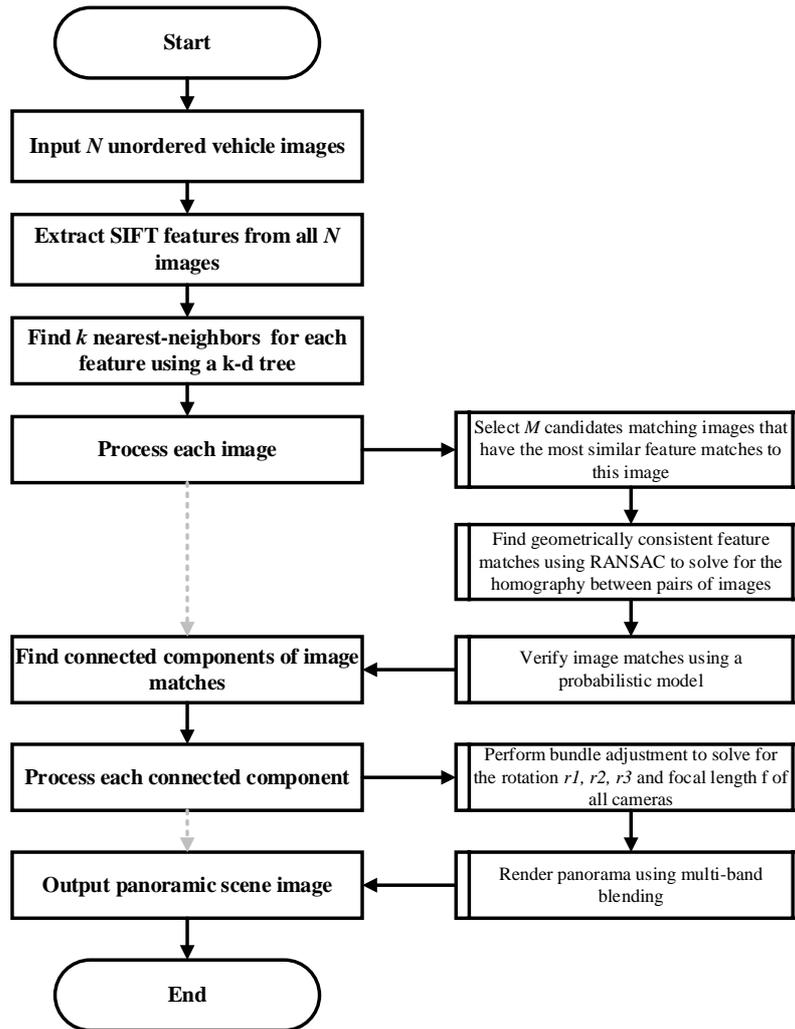


Figure 2. The flowchart of automatic image stitching

By using this automatic stitching algorithm, a high-quality panorama of a train accident scene can be obtained, which contains all the things in the accident including the damaged coaches/carriages, surrounding buildings and other environmental elements. From this panorama, we can also have a better awareness of the whole picture of the accident situation. This set a basis for the next stage where a photo-based modelling technique is employed to identify the position and pose of vehicles involved in the accident, which is the critical information for rescue planning.

3.2 Hybrid Photo-based 3D modelling of vehicles

Up to now, we have obtained the panorama of the train accident scene. In the next step, we need to locate the spatial position and pose of the coaches/carriages contained in the panorama. We propose a novel hybrid photo-based 3D modelling method, which

models this problem as a constrained non-linear least square optimization, then use perspective-projection calculation and projection-error minimization to solve it. Furthermore, based on the coupler connection relationship between coaches/carriages, geometric constraints are taken into account when solving the minimization problem. For such a constrained non-linear least square optimization problem, the conventional Levenberg-Marquart algorithm^{20, 21} used in Bundle Adjustment is no longer valid. Instead, we employ the Trust Region algorithm²².

As to the hybrid method, pre-existing 3D models like CAD models generated by 3D CAD software is a prerequisite, which can normally be satisfied in the railway sector because generally there are limited types of trains in a country. For example, in china, the high-speed EMU is mainly the CRH series, we can acquire the CAD models involved in an accident from factories or the government easily and store them in database in advance. Combining the CAD model of objects to the photo-based modelling method can compensate the depth information about the objects contained in the panorama, as well as conquer the singularity problem existing in traditional single view photo-based modelling method. The hybrid modelling method consists of three parts, projection process, feature extraction and matching and position and pose estimation. In following sections, we will discuss this method in more details.

3.2.1 Projection Process

The projection process shows how to project a 3D model to a 2D image, which is modeled by a perspective projection consisting of translation, rotation and scaling operations. Here, we use the basic pinhole camera model to illustrate the camera mapping process²³, then utilize the finite projective camera model, which is the generalized model of a pinhole camera, to implement the hybrid photo-based 3D modelling. The pinhole camera model is shown in Figure 3. A 3D point can find its projection on the image plane as the intersection of the image plane and a line defined by camera centre and this point.

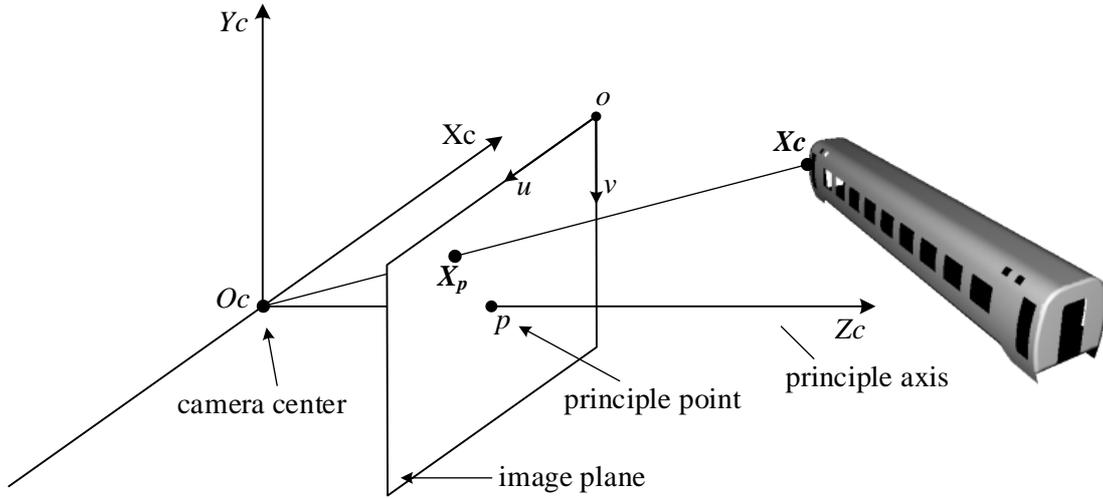


Figure 3. Pinhole camera model

We denote the points on a carriage using $\mathbf{X}_w=(x_w, y_w, z_w) \in \mathbf{R}^3$, in world coordinate frame, and $\mathbf{X}_c=(x_c, y_c, z_c) \in \mathbf{R}^3$ in camera coordinate frame, and denote the pixel coordinates of projection points of carriage by $\mathbf{X}_p=(u, v) \in \mathbf{R}^2$ in image coordinate frame. Based on the pinhole model, the projection process from 3D points on carriage onto 2D image plane can be expressed as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f/dx & 0 & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (1)$$

where $f \in \mathbf{R}$ is the focal length of camera, $dx, dy \in \mathbf{R}$ represents the width and height of a pixel. f/dx and f/dy represent the focal length measured by pixel dimensions in horizontal axis and vertical axis respectively. The principle point is not at the origin of image coordinate frame, and we offset principle point by $(u_0, v_0) \in \mathbf{R}^2$. The intrinsic matrix \mathbf{K} can be defined as:

$$\mathbf{K} = \begin{bmatrix} f/dx & 0 & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

In general, the points in camera coordinate frame can be expressed in world coordinate frame through rotation and translation, and we can denote the two transformation by a rotation $\mathbf{R} \in \mathbf{R}^{3 \times 3}$ and a translation $\mathbf{T} \in \mathbf{R}^3$. The equation can be written as:

$$X_c = R \cdot X_w + T \quad (3)$$

where

$$R = \begin{bmatrix} \cos \theta \cos \psi & \cos \phi \sin \psi + \sin \phi \sin \theta \cos \psi & \sin \phi \sin \psi - \cos \phi \sin \theta \cos \psi \\ -\cos \theta \sin \psi & \cos \phi \cos \psi - \sin \phi \sin \theta \sin \psi & \sin \phi \cos \psi + \cos \phi \sin \theta \sin \psi \\ \sin \theta & -\sin \phi \cos \theta & \cos \phi \cos \theta \end{bmatrix}$$

$$T = [t_1, t_2, t_3]^T$$

ϕ , θ , ψ are the rotation angle around x, y, z axis respectively, and $[R | T]$ is defined as the extrinsic matrix of camera. We can then express the projection of 3D carriage surface in world coordinates frame onto image coordinates frame:

$$X_p = \frac{1}{z_c} K \cdot (R \cdot X_w + T) \quad (4)$$

Starting from the pinhole model, we can generalize the projection by considering the skew effect and distortion of camera lens, which describes a skewing of pixel in CCD array because the x and y axis, as well as radial and tangential distortion coefficient of camera lens, are not perpendicular.

Taking a skew effect into consideration, the intrinsic matrix K can be rewritten as:

$$K = \begin{bmatrix} f / dx & s & u_0 \\ 0 & f / dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

where s denotes the skew parameter.

At this point, we adopt the model proposed by Heikkila²⁴ to introduce the skew and distortion of camera lens. The distortion model can be expressed as:

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \tilde{u} + \delta u^{(r)} + \delta u^{(t)} \\ \tilde{v} + \delta v^{(r)} + \delta v^{(t)} \end{bmatrix} \quad (6)$$

where

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} x_c \\ y_c \end{bmatrix}, \begin{bmatrix} \delta u^{(r)} \\ \delta v^{(r)} \end{bmatrix} = \begin{bmatrix} \tilde{u}(k_1 r^2 + k_2 r^4 + k_5 r^6 \dots) \\ \tilde{v}(k_1 r^2 + k_2 r^4 + k_5 r^6 \dots) \end{bmatrix}, \begin{bmatrix} \delta u^{(t)} \\ \delta v^{(t)} \end{bmatrix} = \begin{bmatrix} 2k_3 \tilde{u} \tilde{v} + k_4 (r^2 + 2\tilde{u}^2) \\ k_3 (r^2 + 2\tilde{v}^2) + 2k_4 \tilde{u} \tilde{v} \end{bmatrix}$$

$$\tilde{r} = \sqrt{\tilde{u}^2 + \tilde{v}^2}$$

In this equation, the distortion coefficients are represent by vector $kc=(k_1, k_2, \dots, k_n)^T$.

Once the skew parameter and distortion coefficients are considered, the projection model can be written as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/dx & s & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \tilde{u} + \delta u^{(r)} + \delta u^{(t)} \\ \tilde{v} + \delta v^{(r)} + \delta v^{(t)} \\ 1 \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ 1 \end{bmatrix} = \frac{1}{z_c} (R \cdot X_w + T)$$

From equation(7), any point X_w on the carriage surface in the world coordinates can have its pixel projection on image plane calculated by knowing the related parameters, like camera focal length, skew parameter and distortion coefficients. In order to express clearly, we denote the projection from world coordinates frame onto image coordinates frame processed by:

$$X_p = Proj(R \cdot X_w + T) \quad (8)$$

3.2.2 Feature extraction and matching

In the last section, we have presented the projection model that describes the relationship between points on carriages surface and their projection on image plane. It means that the crucial work of estimating the position and pose of a carriage boils down to determining the extrinsic parameters $[R | T]$ used in equation(8).

To solve these parameters, the task of this section is confirming the correspondence relationship between the selected points on carriage model in world coordinates frame and its location in a panorama. Hence we model it as a feature matching problem as shown in Figure 4.

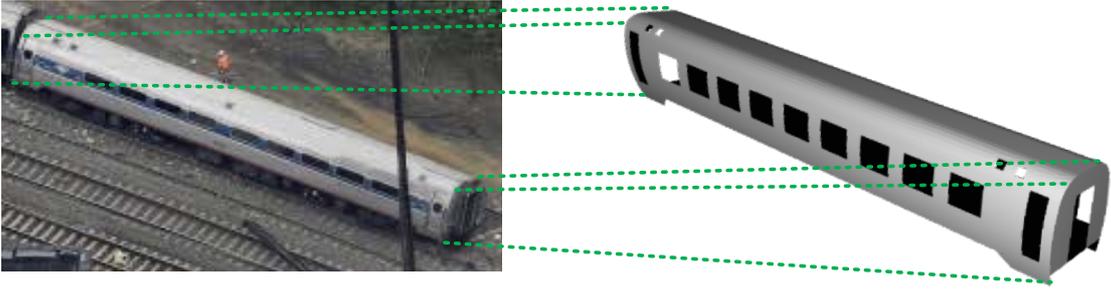


Figure 4. Feature extraction and matching

We adopt the SIFT detector¹⁶ to extract the inherent pattern on the carriage surface in panorama, and mark the location of these inherent patterns onto the 3D carriage model. SIFT detector is robust in detecting features of these patterns under the scale, rotation, blur, illumination and affine changes. We can then build a one to one

correspondence relationship between carriage points in panorama and in the world coordinate frame.

Once the correspondence relationship is confirmed, for any selected point $X_w \in \mathbf{R}^3$ on carriage surface, we know its feature point $X \in \mathbf{R}^2$ in image plane, then the parameters in equation(7) can be determined through solving an optimization problem, where the expected parameters minimise the overall errors among all the pairs of selected matching features.

3.2.3 Position and Pose Estimation

After the feature extraction and matching stage, we can get the correspondence relationship between selected points on carriage surface in 3D space and their feature points on image plane. Let us suppose that there are m carriages in a train accident scene, and n_i features on the i th carriage. We denote the coordinates of the j th point on the i th carriage CAD model in world coordinate frame by $X_{i,j}^w \in \mathbf{R}^3$, and its corresponding feature point in image coordinates frame by $X_{i,j} \in \mathbf{R}^2$ ($i=1, 2, \dots, m, j=1, 2, \dots, n_i$). It is worth emphasizing that the coordinates $X_{i,j}^w$ in world frame is only an initial position of carriage model which needs to be adjusted to the camera frame by rotation and translation. For the i th carriage model, we denote its rotation matrix and translation vector by R_i and T_i respectively as shown in(7). We can then measure the projection error by expression(9):

$$error = \sum_{i=1}^m \frac{1}{N_i} \left(\sum_{j=1}^{n_i} d(Proj(R_i \cdot X_{i,j}^w + T_i), X_{i,j}) \cdot I_{i,j} \right) \quad (9)$$

where $d(\cdot, \cdot)$ is the distance measure, and it is always defined by Euclidean norm. $I_{i,j}$ is the indicator function, which represents the feature matching result. If the j th feature on the i th carriage is visible in the panorama, $I_{i,j} = 1$, otherwise, $I_{i,j} = 0$. $N_i = \sum_{j=1}^{n_i} I_{i,j}$.

Hence, the target of estimating the position and pose of carriages boils down to solving a non-linear least square optimization problem if we define $d(x, y) = \|x - y\|^2$, as shown in expression(10). The intrinsic matrix K , skew parameter s and distortion parameter kc should be solved by camera calibration process which is discussed in following section, hence all the optimization variables are extrinsic matrix $[R_i | T_i]$ ($i=1, 2, \dots, m$).

$$\min \sum_{i=1}^m \frac{1}{N_i} \left(\sum_{j=1}^{n_i} \| \text{Proj}(R_i \cdot X_{i,j}^w + T_i) - X_{i,j} \|^2 \cdot I_{i,j} \right) \quad (10)$$

The Levenberg-Marquart method can be adopted to solve this non-linear least square optimization problem, and all projection information can be obtained if the optimal parameters, $[R_i^* | T_i^*]$ ($i=1, 2, \dots, m$) are searched, including the position and pose information of all the carriages.

Let us denote all mesh points on the i th carriage CAD model in world coordinates frame by $\{P_i^w\} \in \mathbf{R}^3$, then the position and pose of the i th carriage in camera coordinates frame can be expressed as $R_i^* \{P_i^w\} + T_i^*$. Some rotation and translation operator can be utilized if it needs to be transferred to a specific coordinates frame.

Formula(10) is a non-linear least square problem without any constraints. Each pair of optimal parameters $[R_i^* | T_i^*]$ are only decided by minimizing

$$\sum_{j=1}^{n_i} \| \text{Proj}(R_i \cdot X_{i,j}^w + T_i) - X_{i,j} \|^2 \cdot I_{i,j} \quad (i=1, 2, \dots, m).$$

Hence, the formula (10) can be rewritten as a multi-object optimization form, which means that there is no connection between carriages and the reconstruction process of its pose and position is independent. In other words, $[R_i^* | T_i^*](i=1, 2, \dots, m)$ is independent in different carriages. However, in reality this is not the case. On a common train accident spot, the passenger carriages are usually connected by the coupler. If we take these connections into account, it performs better in estimating the entire pose and position of a train than the method expressed in(10), which only has a good performance in estimating separate carriages.

Let us denote the coordinates of all points on the i th carriage model in world coordinates frame by $X_i^w = \{X_{i,1}^w, X_{i,2}^w, \dots, X_{i,n_i}^w\} \in \mathbf{R}^3$. For every carriage, we add the coordinates of couplers $C_{i,1}^w, C_{i,2}^w$ to X_i^w , which represent the front and rear coupler respectively, while $X_i^w = \{X_{i,1}^w, X_{i,2}^w, \dots, X_{i,n_i}^w, C_{i,1}^w, C_{i,2}^w\}$ and the couplers need not to be matched the image feature.

For any couple of carriages X_i^w, X_{i+1}^w the pose and position of the front carriage have an effect on the rear carriage. The pose and position of the rear carriage can be written as:

$$X_{i+1}^w = R_{i+1} (X_{i+1}^w - C_{i+1,1}^w) + T_{i+1} + C_{i,2}^w \quad (11)$$

R_{i+1}, T_{i+1} are the rotation and translation operator of the $i+1$ th carriage relative to the i th carriage. And T_{i+1} describes the translation from the rear coupler of the i th carriage to the front coupler of $i+1$ th carriage. If we require $\|T_{i+1}\|^2 \leq \rho$, the geometric constraints can be added. Moreover, all the carriages have a constrained pose and position except the first carriage. Then we can rewrite the optimization problem by a constrained non-linear least square programming as shown in formula(12).

$$\min \frac{1}{N_1} \sum_{j=1}^{n_1} \|Proj(R_1 \cdot X_{1,j}^w + T_1) - X_{1,j}\|^2 \cdot I_{1,j} + \sum_{i=2}^m \frac{1}{N_i} \left(\sum_{j=1}^{n_i} \|Proj(R_i \cdot (X_{i,j}^w - C_{i,1}^w) + C_{i-1,2}^w + T_i) - X_{i,j}\|^2 \cdot I_{i,j} \right)$$

subject to $\|T_i\|^2 \leq \rho_i, i = 2, \dots, m$ (12)

To solve this problem, we adopt the genetic algorithm with a small number of individuals and generations to generate initial value of optimization variables rapidly, and use trust-region method²² to solve it, which performs well in handling constrained optimization problems.

Once the optimization problem has been solved, the position and pose of the i th carriage in camera coordinates frame can be expressed as $R_i^* \{P_i^w\} + T_i^*$. If it needs to base some specific coordinates frame, the rotation and translation operator can be utilized.

3.3 Modelling based on calibrated or uncalibrated camera case

So far, the main methodology of the hybrid photo-based 3D modelling has been discussed in detail. The only thing that still has not been given is the way of obtaining the camera intrinsic parameters, from which we can calculate extrinsic matrix $[R_i^* | T_i^*](i=1, 2, \dots, m)$ for recovering the position and pose information of carriages. Meanwhile, as previously mentioned, only when we obtain the intrinsic matrix K , skew parameter s and distortion parameter kc beforehand can we solve the optimization problem. Considering actual application, here, we divide the method of acquiring intrinsic parameters into two branches: 1. Calibrate the camera before an accident occurs which is discussed in section 3.3.1; 2. Consider intrinsic parameters as optimization variables and solve them in formula(12) which is detailed in section 3.3.2.

3.3.1 Modelling based on calibrated camera

In general, a camera needs to be calibrated before an accident occurs then can be used to photograph the accident scene and calculate the $[R_i^* | T_i^*]$ further. We follow the four-step camera calibration method²⁴ and use the Camera Calibration Toolbox for Matlab²⁵ to realize the calculation of the camera intrinsic matrix K , skew parameter s and distortion parameter kc in lab. The related case study is discussed in section 5.1.

3.3.2 Modelling based on uncalibrated camera

In some cases, the camera intrinsic parameters of photos captured in an accident scene cannot be obtained by using calibration process. For instance, the pictures are acquired from the internet, or captured by passengers, etc. In this case, we also consider parameter K , s , kc as optimization variables and search the optimal value of them by minimizing the error in formula(12) as the way we optimize $[R_i^* | T_i^*](i=1, 2, \dots, m)$ in previous section. In practice, thanks to the solid-state manufacturing techniques, the skew parameter of camera is negligible. And in most cases, the pixel size is approximately square, which means $dx=dy$. Therefore, we do not need to solve all parameters. If these pictures are captured by high-accuracy camera, we can even suppose that the focal point is in the center of image plane, and distortion parameter is also negligible. Then the projection model reduce to a simple version, and the only variable to be solved is the focal length. We will discuss more details when we conduct a case study in section 5.2, which is based on a single image of train accident from the internet.

4 Modelling other components in a virtual train accident scene

The main aim of this work is to automate the reconstruction of a 3D virtual scene of a train accident in a short time. In the previous sections, we present an automatic modelling method for train carriages. However, except railway vehicles, the environment including terrain, buildings, catenary and tracks are also very important elements in a realistic 3D digital train accident scene.

4.1 Terrain

Railway lines were built through a very large-scale area. In general, A virtual globe engine is required on rendering such wide geological features. Although Google Earth²⁶

is a very popular virtual globe and is applied in many sectors, we find that the functions open to developers are limited. So we adopt an open-source solution. OSG Earth is a scalable open-source terrain rendering toolkit for OpenSceneGraph²⁷. It is a 3D high performance graphics toolkit for applications such as industry simulation and visualization.

4.2 *Railway track, Catenary and building*

The railway track and catenary are in regular shapes and patterns. We develop an automatic modelling script to generate the geometry models of railway track and catenary based on a CAD system named CATIA. The main components of track and catenary including sleeper, ballast, support, and stay are automatically generated. The information to be input to the software tool for track and catenary modelling, such as curve radius, superelevation, can be specified from a known standard and also based on the observation at the accident spot.

Buildings are another possible part of the accident scene. Accordingly, these buildings around the accident spot also need be generated automatically. Some commercial software packages like CityEngine²³ can do this very well. So we utilize this commercial package to model the buildings if there are many buildings around the accident spot.

5 Case study

To validate and evaluate our framework in reconstructing the train accident scene, we conducted two case studies. The first case is to reconstruct the simulated train accident scene in lab, and the second one is the reconstruction of a real derailment accident scene from the internet image which occurred in Philadelphia on May 12, 2015.

5.1 *Reconstruction of in-lab scene*

In this stage, we mimic the real train accident scene by placing 3 small-size carriage models in different pose and location, and measure the coordinates of key points in the surface of carriage models to access the accuracy of our main reconstruction algorithms. In terms of in-lab case study, it is feasible for us to simulate the photo-based 3D modelling based on calibrated camera(Section 3.3). Hence we should calibrate the camera parameters beforehand.

The camera model we calibrated is NIKON D610, and the intrinsic matrix is

$$K = \begin{bmatrix} 6177.4 & -7.1992 & 2993.6 \\ 0 & 6175 & 1941.2 \\ 0 & 0 & 1 \end{bmatrix}$$

distortion coefficients $kc=[-0.020511, 0.1966, -0.0032969, -0.0065837, 0]$ ($n=5$) as expressed in Camera Calibration Toolbox for Matlab²⁸, and the panorama of the in-lab scene is shown in Figure 5.

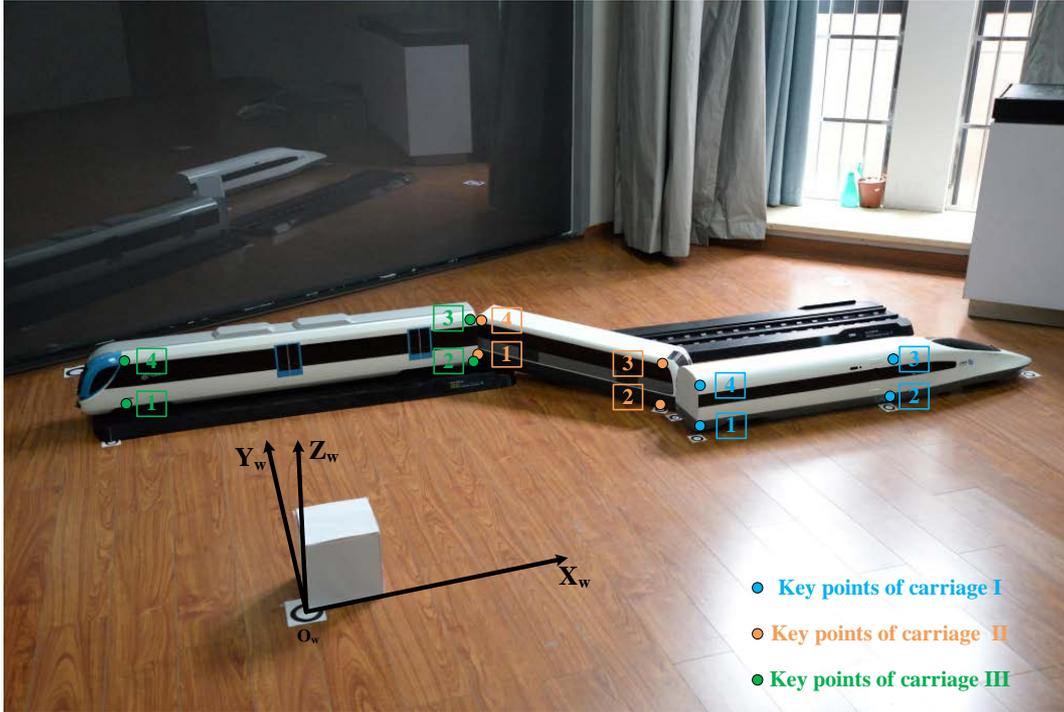


Figure 5. Panorama of the in-lab scene

In world coordinates frame (O_w - X_w - Y_w - Z_w), we denote the coordinates of the j th key point on the i th carriage by $P_{i,j}(x_{i,j}, y_{i,j}, z_{i,j})$. The measured coordinates of 12 key points are listed in Table 1.

Table 1. Coordinates of key points

Train No.	Key point ID	X(mm)	Y(mm)	Z(mm)
I	$P_{1,1}$	1266	521	10
	$P_{1,2}$	1962	521	10
	$P_{1,3}$	1962	521	170
	$P_{1,4}$	1266	521	170
II	$P_{2,1}$	756	1245	51
	$P_{2,2}$	1242.8	768.99	5.8529
	$P_{2,3}$	1252.9	763.49	173.17
	$P_{2,4}$	766.16	1239.5	218.37

III	$P_{3,1}$	-381.04	1196.5	43
	$P_{3,2}$	756	1196.5	43
	$P_{3,3}$	756	1202.2	230.58
	$P_{3,4}$	-381.03	1202.2	230.66

While, as shown in Figure 5, the texture information is not obvious around these key points and cannot be easily distinguished from the background, hence it is a tough task to recognize these points by using feature matching and detection algorithms. To detect these locations of key points on carriages automatically, we put markers on the carriage as shown in Figure 6. Then the SIFT detector is applied to recognize the pixel coordinates of each marker, and the recognition result is shown in Figure 7. To reconstruct the whole train accident scene, CAD models of each carriage should be prepared in database (Figure 8).



Figure 6. Panorama of marked carriages



Figure 7. Recognition result of each marker

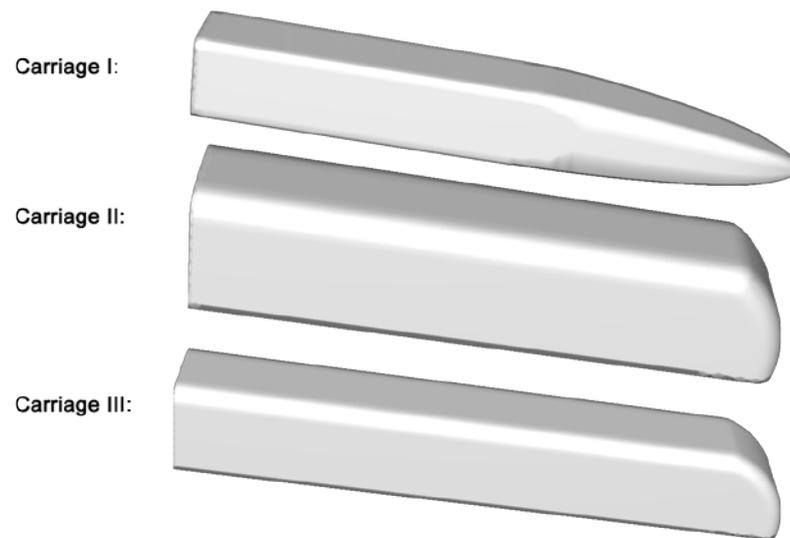


Figure 8. CAD models of 3 carriages

Once the previous work has been finished, the reconstructed scene can be obtained with solving the optimization problem in(12), and the result is shown in Figure 9.



Figure 9. Reconstructed scene

Figure 10 shows a comparison chart of position accuracy between real carriages in lab and the reconstructed one. The red region in this chart illustrates the location of real in-lab carriages, while the grey region denotes the reconstructed carriages. The box at the bottom-left position of this scene roles as an origin point, which is used to unify all points to a world coordinates system. It is intuitive that the proposed reconstruction algorithms achieve a good performance.

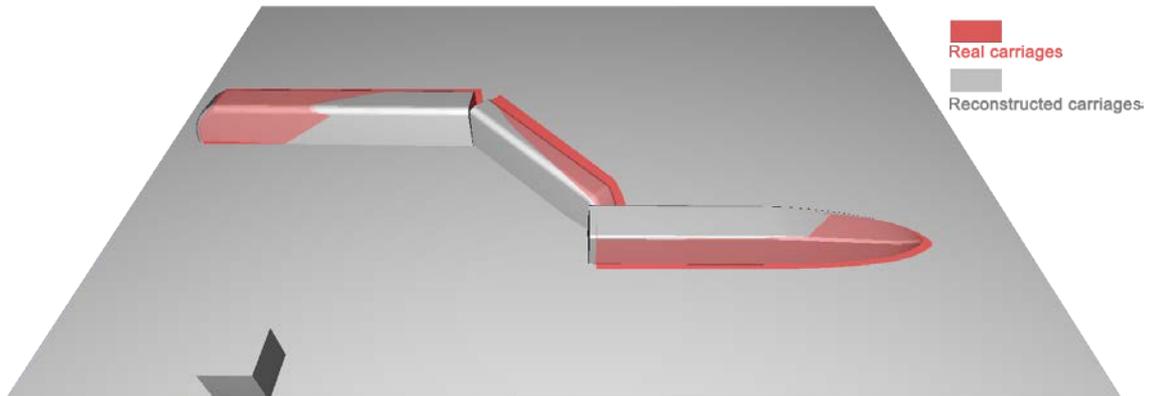


Figure 10. Comparison chart of real and reconstructed carriages

To access the accuracy of the whole reconstruction in Figure 10 quantitatively, we have listed the reconstructed coordinates of key points (as denoted in Figure 5), and made a comparison with the measured value of real in-lab carriages in Table 2. The errors are listed in Table 3

Table 2. Comparison results

Real coordinates				Reconstructed coordinates			
Train No.	X(mm)	Y(mm)	Z(mm)	Train No.	X(mm)	Y(mm)	Z(mm)
I	1266	521	10	I	1290	520.95	24
	1962	521	10		1962	520.95	24
	1962	521	170		1962	520.95	156
	1266	521	170		1290	520.95	156
II	756	1245	51	II	693.4	1101.5	46.127
	1242.8	768.99	5.8529		1176.2	640	59.279
	1252.9	763.49	173.17		1232.8	702.13	161.06
	766.16	1239.5	218.37		750.01	1163.6	147.91
III	-381.04	1196.5	43	III	-427.27	1012.2	85.335
	756	1196.5	43		691.96	1100.4	59.581
	756	1202.2	230.58		684.3	1227	160.38
	-381.03	1202.2	230.66		-434.93	1138.8	186.13

Table 3. Absolute errors

Train No.	Key point ID	X(mm)	Y(mm)	Z(mm)
I	$P_{1,1}$	24	0.05	14
	$P_{1,2}$	0	0.05	14
	$P_{1,3}$	0	0.05	14
	$P_{1,4}$	24	0.05	14
II	$P_{2,1}$	62.6	143.5	4.873
	$P_{2,2}$	66.6	128.99	53.4261
	$P_{2,3}$	20.1	61.36	12.11
	$P_{2,4}$	16.15	75.9	70.46
III	$P_{3,1}$	46.23	184.3	42.335
	$P_{3,2}$	64.04	96.1	16.581
	$P_{3,3}$	71.7	24.8	70.2
	$P_{3,4}$	53.9	63.4	44.53
	Mean	37.44333	64.87917	30.87626

From Table 2 and Table 3, the cumulative absolute errors of 3 carriages in different directions can be calculated, which are shown in Figure 11, display that the algorithm has a better performance in reconstructing carriage I than others. Figure 12 denotes the cumulative absolute errors of different carriages in 3 directions, and it shows a relatively large reconstruction errors in y-axis direction, which is closely related to the depth in camera coordinates space. Generally it is hard for single-view reconstruction method to estimate the depth information from one image.

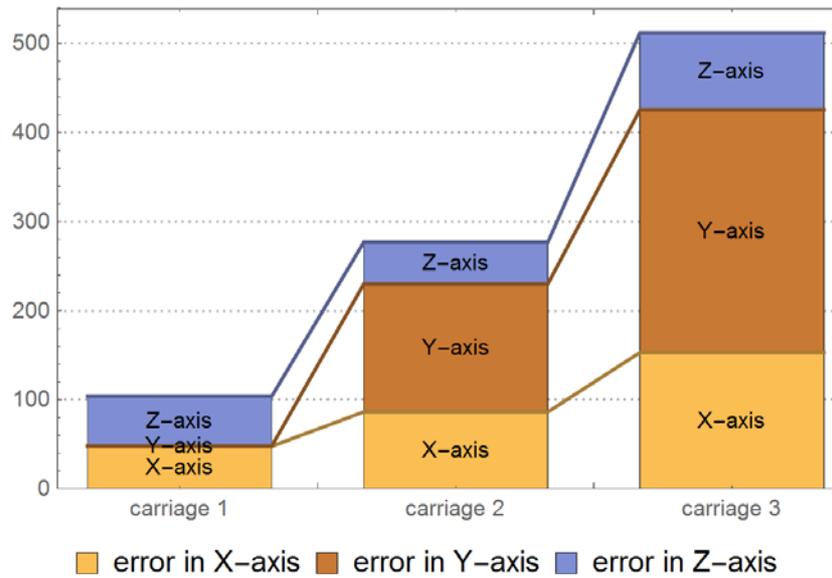


Figure 11. Cumulative absolute errors distribution of 3 carriages

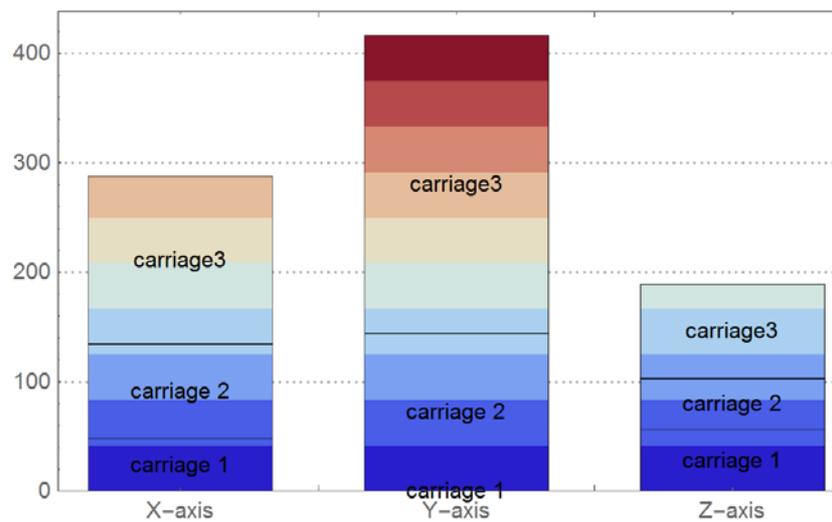


Figure 12. Cumulative absolute errors of different carriages in 3 directions

The reconstruction process in this case is based on the assumption that there are some inherent texture or artificial markers on the surface of carriages, which can instruct the feature detector to recognize. It is impossible to complete the reconstruction work without the information of pixel location about carriages. However, in most real cases, the photos may generally lack the necessary texture information about carriages, for instance, when the resolution of photo is low or the captured object is smooth or transparent. Hence, some manual intervention in recognizing the pixel location of

carriages is necessary. If the SIFT detector cannot perform or leads to a wrong match, manual intervention can improve the effectiveness and accuracy.

5.2 *Reconstruction of real derailment scene*

A real train accident site differs from an ideal laboratory environment in terms of constraints on accessibility, line-of-sight, and illumination. In the in-lab scene, the camera intrinsic parameters like the focal length, distortion etc. and extrinsic parameters (i.e., position and orientation) can be pre-known or determined with calibration process.

In this section, we reconstruct the train accident scene which occurred in Port Richmond neighborhood of Philadelphia, Pennsylvania, on May 12, 2015. The photo is shown in Figure 13 which displays that 6 carriages have derailed²⁹. The type of the passenger car involved in this accident is Amfleet I series passenger cars, and its simplified CAD model is shown in Figure 14.

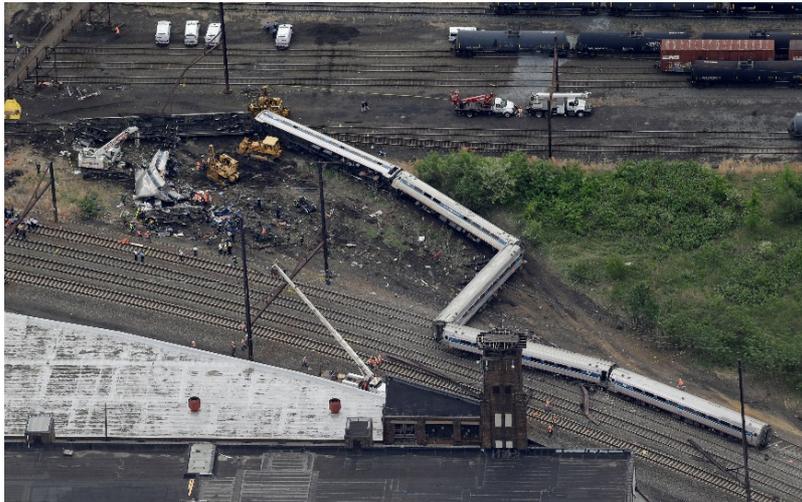


Figure 13. Accident scene occurring in Port Richmond

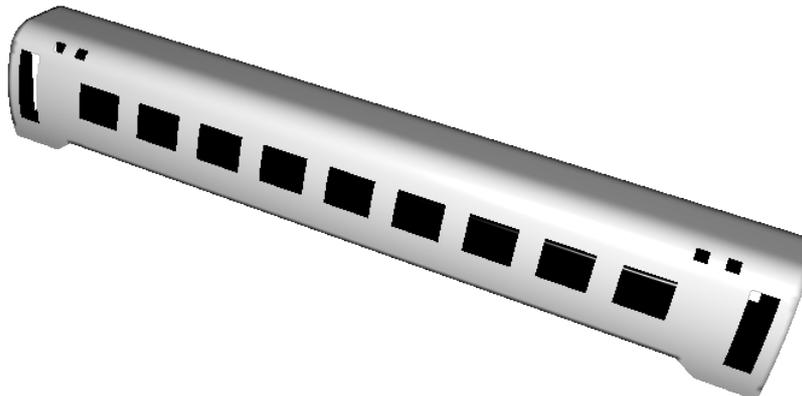


Figure 14. Simplified CAD model of Amfleet I

As mentioned in the previous sections, the SIFT detector may fail in extracting a feature where there is no necessary texture information. In that case, appropriate manual intervention can improve the performance of the pose and position recognition algorithm.

In this case, it is clear that we cannot obtain the camera parameters and there are not enough significant feature information on the surface of carriages. So we need to utilize a simplified camera model (mentioned in Section 4.2) and manual matching instead of SIFT detector to accomplish the reconstruction work. Because the skew and distortion effects are ignored, briefly it may cause a large number of errors. In this case, the topographic information of the accident scene is utilized in this paper which is a key factor to improve the reconstruction performance. We summarize the reconstruction as the following steps:

Step1. Estimate pose information of each carriage in case that geometric constraints are canceled.

Step2. Move each carriage to the zero potential energy surface (surface of terrain in the spot of accident scene)

Step3. Introduce the geometric constraints, but keep the rotation matrix of each carriage unchanged which is solved in step1 and only optimize the other variables (translation vector, focal length, etc.).

Step4. Substitute all parameters acquired by previous 3 steps into equation (3), the pose and position information of each carriage are obtained.

Once all the steps are accomplished, the 3D reconstruction result of carriages can closely represent the real case, which is shown in Figure 15.

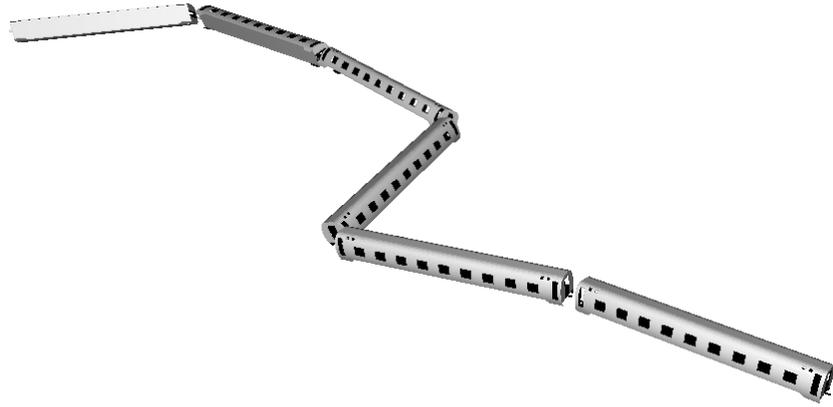


Figure 15. Reconstructed carriages

Afterwards, the reconstructed carriages are integrated into the 3D virtual geographical environment created by 3D GIS and visualization engine. A digital 3D train accident scene is created, as shown in Figure 16.

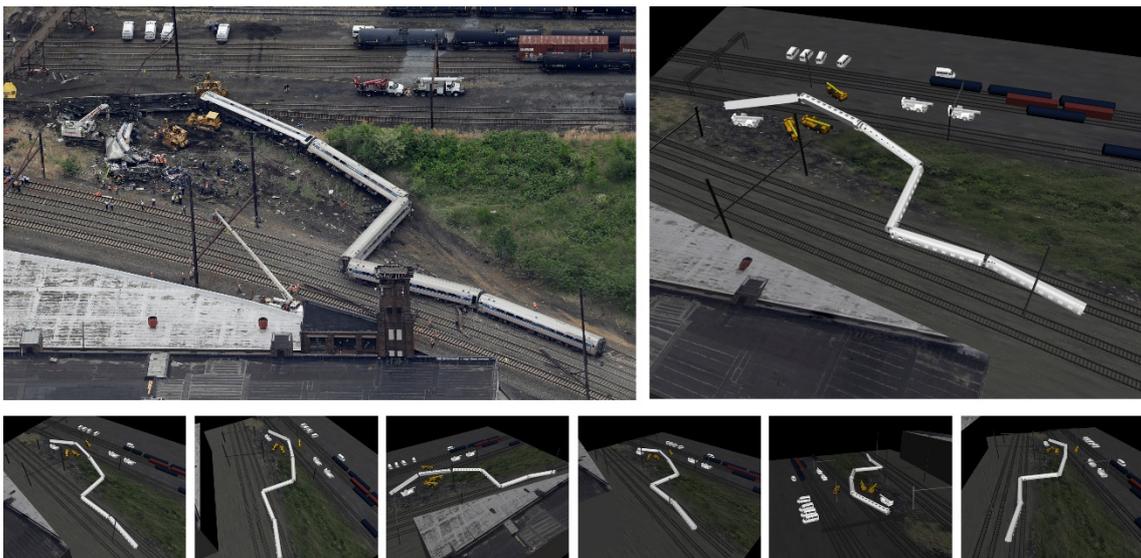


Figure 16. Simulated train accident scene.

The reconstructed 3D digital accident scene allows users to view the 3D scene from a different angle and know exactly the relative locations and configurations of different objects, allowing further interactions with objects and making a rescue plan with 3D planning software.

6 Conclusion and future work

In this paper, we propose a framework to rapidly reconstruct the 3D digital scene of a train accident, and develop a hybrid photo-based 3D modelling method for vehicles involved in an accident, which helps make a development in using computer vision and virtual reality technique for railway emergency rescue and offers benefits to railway industry. The main conclusions and contributions of this work are as follows:

- (1) We offer a cost-efficient, flexible, effective framework to construct 3D scene of a train accident, which will make the related rescue planning more efficient and transparent, thereby minimizing the loss of lives and properties.
- (2) To meet the particular needs of railway industry, we tailor a hybrid photo-based 3D modelling method for the reconstruction of a train accident scene. Differing from conventional photo-based 3D modelling method, it can generate the 3D digital scene rapidly after a train accident occurs without resort to significant manual interventions. In extreme condition, this method only needs a single panoramic image of the accident scene and do not prone to the singularity problem which is common in the traditional single view photo-based modelling.
- (3) The errors of the whole reconstruction process may be caused by the following reasons:
 - a) The calibrated camera model is an approximation with simplified model of the complete camera projection. An accurate camera model is essential if we require a precise projection.
 - b) There is a disparity between the CAD model of carriages and the real one in accidents scene, especially when a carriage severely damaged.
 - c) There may be a bias between the recognized feature on the image plane and its respective point on the carriage model. This part is the main error source in our practice. Limited human intervention with minor effort can greatly reduce the error.

The accumulation of errors may lead to a bad reconstruction, so we should control the main error sources. While in some special cases, it is not possible to avoid these errors, e.g. we may not know the calibrated camera parameters (which is mentioned in Section3.3), or there is no significant features that can be recognized (Section 5.2). Fortunately, by the introduction of geometric constraints and topographic information, we can ensure the accuracy of our algorithms even in the worst situations (Section5.2).

(4) Two case studies, in-lab small-scale train accident scene and real derailment scene in America, are conducted. With proposed method and applying 3D visualization technology to rendering the environment, two different digital 3D scenes of train accidents are obtained. By comparison, we find that the proposed framework has a good reconstruction performance in both accuracy and operability. For real derailment scene, the images used for reconstruction are from the internet, while the intrinsic parameters of camera are unavailable. But, we also achieve plausible result.

In the future, we envisage that the framework will be used in train accident rescue planning as well as for accident analysis by providing useful information, as well as helping make a reliable rescue decision effectively. There are also still many works to be done to improve the accuracy of 3D scene modelling and cover more application scene, such as at night or in foggy weather. In addition, more aided software tools, including modelling and visualization, also need to be developed for the proposed framework can be better applied in train accident rescue or education and training.

7 Acknowledgements

This research was supported by National Natural Science Foundation of China (51405402) and the Self-developed Research Project of the State Key Lab of Traction Power(2015TPL_T06).

8 References

1. Zarboutis N, Marmaras N. Searching efficient plans for emergency rescue through simulation: the case of a metro fire. *Cognition, Technology & Work*. 2004; 6: 117-26.
2. Fan Y, Li Z, Pei J, Li H, Sun J. Applying systems thinking approach to accident analysis in China: Case study of “7.23” Yong-Tai-Wen High-Speed train accident. *Safety science*. 2015; 76: 190-201.
3. Aytakin C, Rezaeitabar Y, Dogru S, Ulusoy I. Railway Fastener Inspection by Real-Time Machine Vision.
4. Lai YC, Barkan CPL, Drapa J, et al. Machine vision analysis of the energy efficiency of intermodal freight trains. *Proceedings Of the Institution Of Mechanical Engineers Part F-Journal Of Rail And Rapid Transit*. 2007; 221: 353-64.
5. Sawadisavi S. Machine-vision inspection of railroad track. University of Illinois at Urbana-Champaign, 2009.
6. Schlake BW, Todorovic S, Edwards JR, Hart JM, Ahuja N, Barkan CPL. Machine vision condition monitoring of heavy-axle load railcar structural underframe components. *Proceedings Of the Institution Of Mechanical Engineers Part F-Journal Of Rail And Rapid Transit*. 2010; 224: 499-511.
7. Dai F, Lu M. Photo-based 3D modeling of construction resources for

visualization of operations simulation: case of modeling a precast façade. *Proceedings of the 40th Conference on Winter Simulation: Winter Simulation Conference*, 2008, p. 2439-46.

8. Yang M-D, Chao C-F, Huang K-S, Lu L-Y, Chen Y-P. Image-based 3D scene reconstruction and exploration in augmented reality. *Automation in Construction*. 2013; 33: 48-60.

9. Bhatla A, Choe SY, Fierro O, Leite F. Evaluation of accuracy of as-built 3D modeling from photos taken by handheld digital cameras. *Automation in construction*. 2012; 28: 116-27.

10. Udayan JD, Kim H, Kim J-I. An image-based approach to the reconstruction of ancient architectures by extracting and arranging 3D spatial components. *Journal of Zhejiang University SCIENCE C*. 2015; 16: 12-27.

11. Frankl A, Stal C, Abraha A, et al. Detailed recording of gully morphology in 3D through image-based modelling. *Catena*. 2015; 127: 92-101.

12. Zancajo-Blazquez S, Gonzalez-Aguilera D, Gonzalez-Jorge H, Hernandez-Lopez D. An Automatic Image-Based Modelling Method Applied to Forensic Infography. *PLoS one*. 2015; 10: e0118719.

13. Fernández - Hernandez J, González - Aguilera D, Rodríguez - González P, Mancera - Taboada J. Image - Based Modelling from Unmanned Aerial Vehicle (UAV) Photogrammetry: An Effective, Low - Cost Tool for Archaeological Applications. *Archaeometry*. 2015; 57: 128-45.

14. Criminisi A, Reid I, Zisserman A. Single view metrology. *International Journal of Computer Vision*. 2000; 40: 123-48.

15. Hoiem D, Efros AA, Hebert M. Geometric context from a single image. *Computer Vision, 2005 ICCV 2005 Tenth IEEE International Conference on: IEEE*, 2005, p. 654-61.

16. Lowe DG. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004; 60: 91-110.

17. Triggs B, McLauchlan PF, Hartley RI, Fitzgibbon AW. Bundle adjustment—a modern synthesis. *Vision Algorithms: Theory and Practice International Workshop on Vision Algorithms Proceedings (Lecture Notes in Computer Science Vol 1883)*. 2000: 298-372.

18. Juan L, Gwon O. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*. 2009; 3: 143-52.

19. Brown M, Lowe DG. Automatic panoramic image stitching using invariant features. *International journal of computer vision*. 2007; 74: 59-73.

20. Levenberg K. A method for the solution of certain problems in least squares. *Quarterly of applied mathematics*. 1944; 2: 164-8.

21. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*. 1963; 11: 431-41.

22. Moré JJ, Sorensen DC. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*. 1983; 4: 553-72.

23. CityEngine E, . ESRI CityEngine <http://www.esri.com/software/cityengine>. 2015.

24. Heikkila J, Silvén O. A four-step camera calibration procedure with implicit image correction. *Computer Vision and Pattern Recognition, 1997 Proceedings, 1997 IEEE Computer Society Conference on: IEEE*, 1997, p. 1106-12.

25. Bouguet J-Y. Camera calibration toolbox for matlab. 2004.

26. Google INC. Goolgle Earth <https://www.google.com/earth/>.

27. OSG. OSG www.openscenegraph.org.

28. Bouguet J-Y. Camera Calibration Toolbox for MATLAB, 2013, http://www.vision.caltech.edu/bouguetj/calib_doc. 2014.

29. PHILLIPS S. What's in those tank cars near the Amtrak derailment? STATEIMPACT: SUSAN PHILLIPS, 2015.