A Method for 3D Reconstruction of Train Accident Scene Using Photos

Z. Tang[1], Y.Y. Nie[1], J. Chang[2], J.J. Zhang[1, 2]
[1]State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu, China
[2]National Centre for Computer Animation, Bournemouth University, Poole, UK

Abstract

Railway accident that usually cause numerous property and life losses occurred frequently in recent years all around the world. In general, plenty of resources such as financial supports and incident rescue programs are required to minimize the loss after an accident occurred. While due to poor information from the accident spot, most railway emergency management departments may face a predicament in setting up schemes to conduct the accident rescue. To provide more sufficient visual information about the accident field in drawing up a rescue planning, a realistic 3D virtual accident scene reconstruction technology is recommended which can present supplementary materials and information about the train accident, and assist relevant personnel to make a reliable rescue decision effectively. A photo-based 3D modelling framework of vehicles for measuring the position and pose of carriages involved in the accident is proposed. At last, we implement and examine two case studies to validate this reconstruction method and conclude that our methods perform well in the assigned reconstruction.

Keywords: Train accident reconstruction, accident rescue, Photo based modelling.

1 Introduction

A train accident, for its severity in life and property lose, always demands huge resources and supports from departments of railway emergency management. Many train accidents still happen worldwide every year and there is approximately one severe accident every 2 years[1] despite the fact that the railway industry attempts to ensure the safety. Once a train accident occurs, it requires executing an effective incident rescue planning to minimize the loss. While most related departments faced a dilemma in drawing up scheme of accident rescue for poor information from the scene of an accident after a train accident occurs. Additionally, the implementation of traditional emergency rescue plan is time-consuming.

In recent decades, computer vision technology has been utilised in railway industry[2-5]. 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 photos[6]. This technique has been applied in various disciplines, including virtual reality or augmented reality[7], architecture[8, 9], earth sciences[10], even in forensic infography[11] and archaeometry[12] 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.

In this paper, with using the photos took on the accident spot, we propose a framework which can rapidly and automatically construct a 3D virtual scene of a train accident. The framework uses a single photo-based modelling method and combine with CAD models to extract the position and pose information of carriages involved in an accident with high reconstruction accuracy and relatively efficient in computation cost. Meanwhile, adopts a geographic information system and the 3D visualization engine to model and display the landscapes and buildings in a train accident.

2 Methodology

It should be emphasized that the main aim of this work is to provide a visual reference for drawing up a rescue scheme quickly and automatically after an accident occurs. All the inputs we handle are images captured in the accident scene. Moreover, the realization of reconstruction depends on the quality of images. Therefore, reliable image pre-treatment process is essential for further analysis.
For the first step, we need to recognize a panorama that can describe the complete picture of a train accident scene. The photos taken from the accident scene may be fractional and each of them only contains the information related to a particular part of the scene. We apply the image automatic panorama stitching algorithm[13] which aligns and stitches photos into a seamless panorama.

2.1 Image stitching

The objective of image stitching is to find all matching images, then connect sets of image matches to produce a panorama. We recommend Lowe's Scale Invariant Feature Transform (SIFT) features for image matching thanks to its good performance in achieving reliable matching of multiple images and handling scale, rotation, blur, and affine changes of image than SURF and PCA-SIFT detector[14]. To automatically discover the matching relationships in unordered multi-images, the automatic panorama stitching algorithm proposed by Matthew Brown and David G. Lowe [13]is recommended in this paper.

By using this automatic stitching algorithm, a high-equality panorama of a train accident scene can be obtained, which contains all 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 feature extraction and matching is employed to identify the position of vehicles in the panorama, which is the critical information for rescue planning.

2.2 Feature extraction and matching

Once the panorama is obtained, the task of this section is confirming the correspondence relationship between the marked points on carriage model in world coordinates frame and its respond location in panorama. Hence we model it as a feature matching problem as shown in Figure 1.1. We can search within panorama for locations of vehicles, using image feature detecting and matching technology. At this point, the Scale Invariant Feature Transform(SIFT) algorithm [15] is adopted which performs well in recognizing the pixel coordinates of vehicles in a panorama. SIFT is robust in detecting features under the scale, rotation, blur, illumination and affine changes.

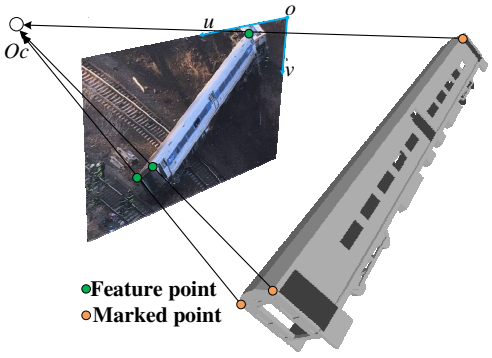

Figure 1.1. Feature matching process

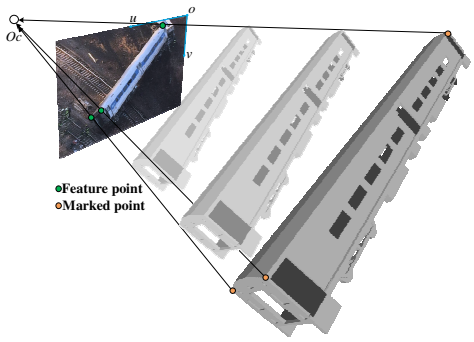

Figure 1.2. Singularity problem

From Figure 1.1 we denote the green dots as the feature points which are extracted by SIFT method, and orange dots as these fixed points marked on the CAD model of a specified carriage. All the work of SIFT method is to detect some designated points in panorama which should correspond to the same point in CAD model. The introduction of CAD model can not only solve the singularity which is a common problem in single-view 3D reconstruction(shown in Figure 1.2), but also greatly reduce the time-costing of our reconstruction algorithm. Because we only extract few point(no more than 7 points per carriage) that is much less than the conventional point-cloud based reconstruction methods.

2.3 3D modelling of vehicles

Hitherto 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 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[16, 17] used in Bundle Adjustment is no longer valid, and we employ the Trust Region algorithm[18] instead.

2.3.1 Projection Process

The projection process shows how to project a 3D model to a 2D image, which is modelled by a perspective projection consisting of translation, rotation and scaling operations. Here, we use the basic pinhole camera model to illustrate the camera mapping process[19], then utilize the finite projective camera model, which is the generalized model of a pinhole camera, to implement the photo-based 3D modelling. The 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.

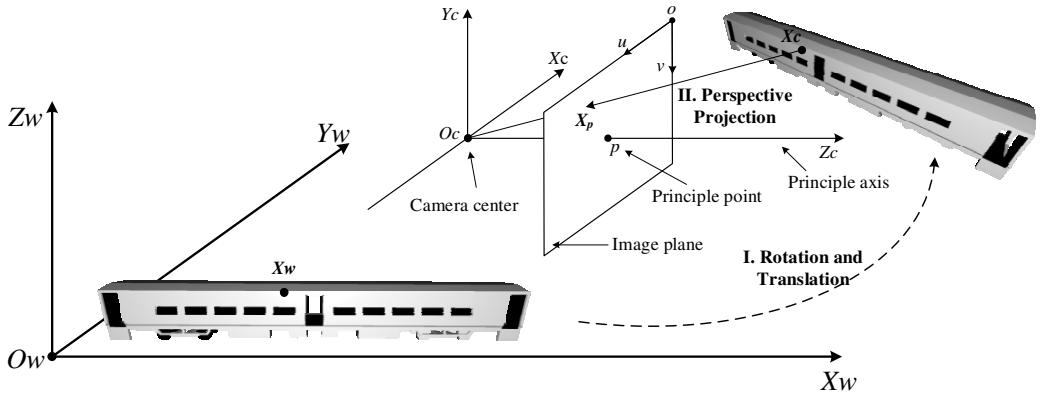

Figure 3. Camera model

We denote the points on a carriage using $X_w=(x_w, y_w, z_w) \in$ R3, in world coordinate frame, and $X_c=(x_c, y_c, z_c) \in \boldsymbol{R}^3$ in camera coordinate frame, and denote the pixel coordinates of projection points of carriage by $X_p=(u, v) \in$R2 in image coordinate frame. Based on the camera model[20, 21], the projection process from 3D points on carriage onto 2D image plane can be expressed as:

1. Rotation and translation process：

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

2. Lens distortion refine process：

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{u} + \delta u^{(r)} + \delta u^{(t)} \\ \tilde{v} + \delta v^{(r)} + \delta v^{(t)} \\ 1 \end{bmatrix}$$

Where $\begin{bmatrix} \delta u^{(r)} \\ \delta v^{(r)} \end{bmatrix} = \begin{bmatrix} \tilde{u}(k_1 r^2 + k_2 r^4 + k_5 r^6 ...) \\ \tilde{v}(k_1 r^2 + k_2 r^4 + k_5 r^6 ...) \end{bmatrix}$, $\begin{bmatrix} \delta u^{(t)} \\ \delta v^{(t)} \end{bmatrix} = \begin{bmatrix} 2k_3 \tilde{u}\tilde{v} + k_4(r^2 + 2\tilde{u}^2) \\ k_3(r^2 + 2\tilde{v}^2) + 2k_4 \tilde{u}\tilde{v} \end{bmatrix}$

3. Perspective projection process：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}, \qquad K = \begin{bmatrix} f/dx & s & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

In the previous formulas, s, [k_1, k_2,…] are denoted as skew and distortion parameters, and we name K as the camera intrinsic parameters, [R | T] as the extrinsic parameters. For briefly expressing, the projection process is denoted as：

$$X_p = Proj(R \cdot X_w + T) \tag{1}$$

From this section, we can understand that the extrinsic parameters [R | T] contain the whole information of spatial position and pose of the point X_p in panorama. Hence, the target of this chapter is catch the extrinsic parameter of carriages and recover their spatial pose and position information further.

2.3.2 Position and Pose Estimation

After the feature extraction and matching stage, we can get the correspondence relationship between marked points on carriage CAD model and their feature points on image plane. Suppose that there are m carriages in a train accident scene, and n_i features on the ith carriage, and we denote the coordinates of the jth point on the ith carriage CAD model in world coordinate frame by $X_{i,j}^w \in \boldsymbol{R}^3$, and its corresponding feature point in image coordinates frame by $X_{i,j} \in \boldsymbol{R}^2$ (i=1, 2, …, m, j=1, 2, …,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 ith carriage model, we denote its rotation matrix and translation vector by R_i and T_i respectively as shown in(1). We can then measure the projection error by expression:

$$X_{i,j}^p = Proj(R_i \cdot X_{i,j}^w + T_i), e_{i,j} = X_{i,j} - X_{i,j}^p$$

Let us denote the coordinates of all points on the ith carriage model in world coordinates frame by $X_i^w = \{X_{i,1}^w, X_{i,2}^w, \ldots, X_{i,n_i}^w\} \in \mathbf{R}^3$. For every carriage, we augment X_i^w with the coordinates of couplers $C_{i,1}^w, C_{i,2}^w$, which represent the front and rear coupler respectively. While $X_i^w = \{X_{i,1}^w, X_{i,2}^w, \ldots, X_{i,n_i}^w, C_{i,1}^w, C_{i,2}^w\}$ and the couplers need not to be matched the image feature. The main idea of this algorithm is illustrated in Figure 4.

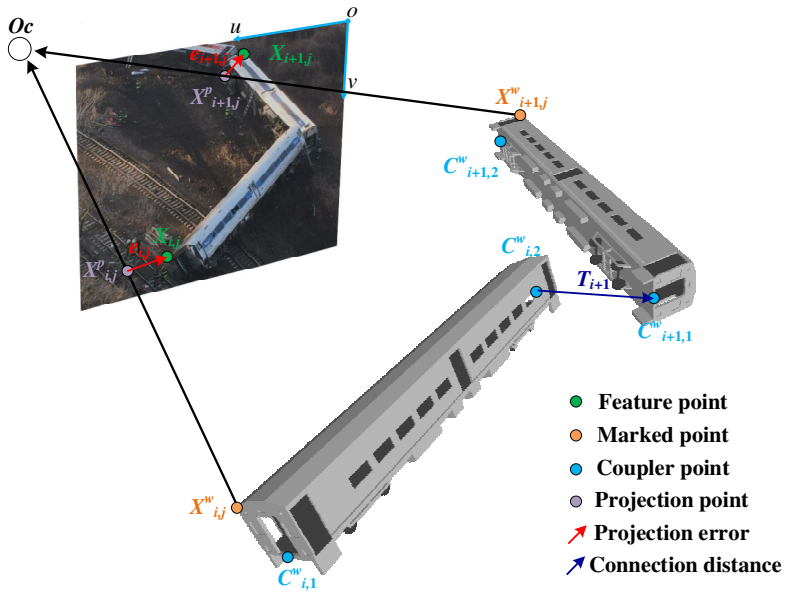

Figure 4. Position and pose estimation process

Then the main target is to estimate the extrinsic parameters $[R_i \mid T_i]$(i=1, 2, …, m). In this stage, we denote the position and pose information of the first carriage as:

$$\tilde{X}^w_{1,j} = R_1 \cdot X^w_{1,j} + T_1,\ \tilde{X}^w_i = \{\tilde{X}^w_{i,1}, \tilde{X}^w_{i,2}, \ldots, \tilde{X}^w_{i,n_i}, \tilde{C}^w_{i,1}, \tilde{C}^w_{i,2}\}$$

and the i+1 th carriage's pose and position can be expressed as:

$$\tilde{X}^w_{i+1} = R_{i+1}(X^w_{i+1} - C^w_{i+1,1}) + T_{i+1} + \tilde{C}^w_{i,2}$$

The coordinates of the i+1 th carriage is based on the location of rear coupler of the ith carriage, and $[R_i \mid T_i]$(i=1, 2, …, m) is the variables which will be optimized as follows.

Then we deduce the multi-carriage position and pose measuring problems to a non-linear least square minimization of the error in perspective projection. In addition, considering the coupler connection relationship between carriages, we need add more constraint conditions to resolve the actual measurement of locations and poses. We use the Trust Region algorithm[18] instead of the Levenberg-Marquart algorithm[16, 17] , which is adopted in BA[22], to solve the constrained non-linear least square problem. A reliable location and pose of a 3D model of accident vehicle can be obtained by solving the optimal $[R_i^* \mid T_i^*]$ (i=1, 2, …, m) in (2):

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

$$subject\ to\ \left\|T_i\right\|^2 \le \rho_i, i = 2,\ \ldots, m \qquad (2)$$

where $I_{i,j}$ is the indicator function, which represents the feature matching result. If the jth feature on the ith carriage is visible in the panorama, $I_{i,j}$ = 1, otherwise, $I_{i,j}$ = 0. $N_i = \sum_{j=1}^{n_i} I_{i,j}$. ρ_i is the maximum distance between the i-1 th carriage and the i th

carriage. Once the optimization problem has been solved, the position and pose of the *i*th carriage in camera coordinates frame can be expressed as $R_i^*\{P_i^w\}+T_i^*$.

2.4 Modelling based on calibrated or uncalibrated camera case

So far, the main methodology of the 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 $[R_i^* \mid T_i^*](i=1, 2, \ldots, m)$ for recovering the position and pose information of carriages. In other words, only the intrinsic matrix K, skew parameter s and distortion parameter kc are obtained 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[21]; 2. Consider intrinsic parameters as optimization variables and solve them in formula (2) using simplified camera model to maintain the robustness of our algorithm, such as the pinhole model.

3 Case study and discussion

To validate and evaluate our framework in reconstructing the real rail deraiment accident scene, we conducted two case studies, which are occurred in Bronx, New York City, Dec 1st, 2014[23] and Philadelphia on May 12, 2015[24].

In this section, we reconstruct the train accident scene which displays that several carriages have derailed. The photo and corresponding simplified CAD carriage model is shown in Figure 5 and Bronx and Philadelphia

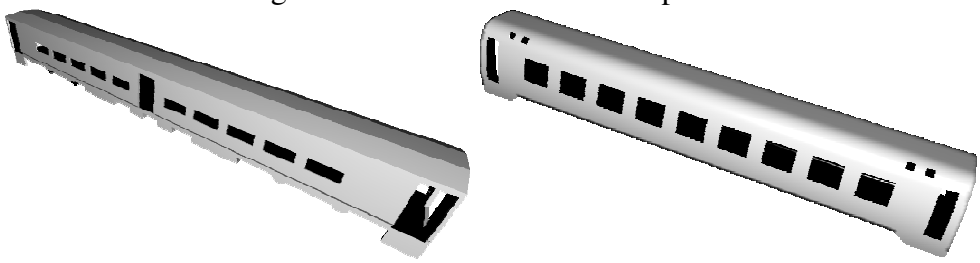

Figure 6.

Figure 5. Rail derailment accident occurred in Bronx and Philadelphia

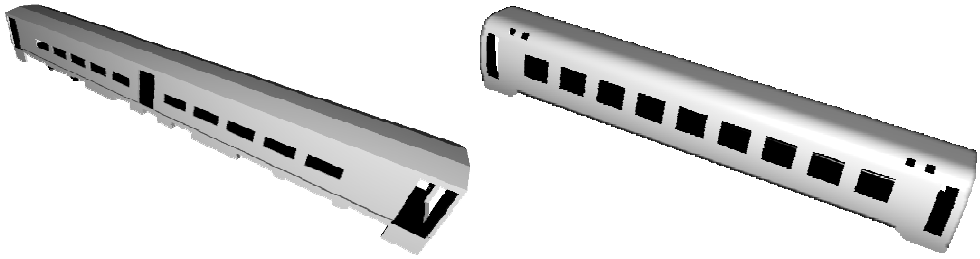

Figure 6. Corresponding CAD models

In the real derailment scene, 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.

Figure 7. Feature extraction results

In this case, it is clear that we cannot obtain the camera parameters and there are not enough significant texture information on the surface of carriages. So we need to utilize a simplified camera model (mentioned in Section 0) and manual matching instead of SIFT detector to accomplish the reconstruction work (Figure 7). 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 cancelled.

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

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

Step4. Substitute all parameters acquired in previous 3 steps into equation (2), 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 (Figure 8). Then integrate it into a proper 3D virtual geographical environment created by 3D GIS and visualization engine. A digital 3D train accident scene is created, which is shown in Figure 9-Figure 11. The reconstructed 3D digital accident scene allows users to view the 3D scene from a different angle and get a big picture of the relative locations and configurations of different objects, allowing further interactions with objects and making a rescue plan with 3D planning software.

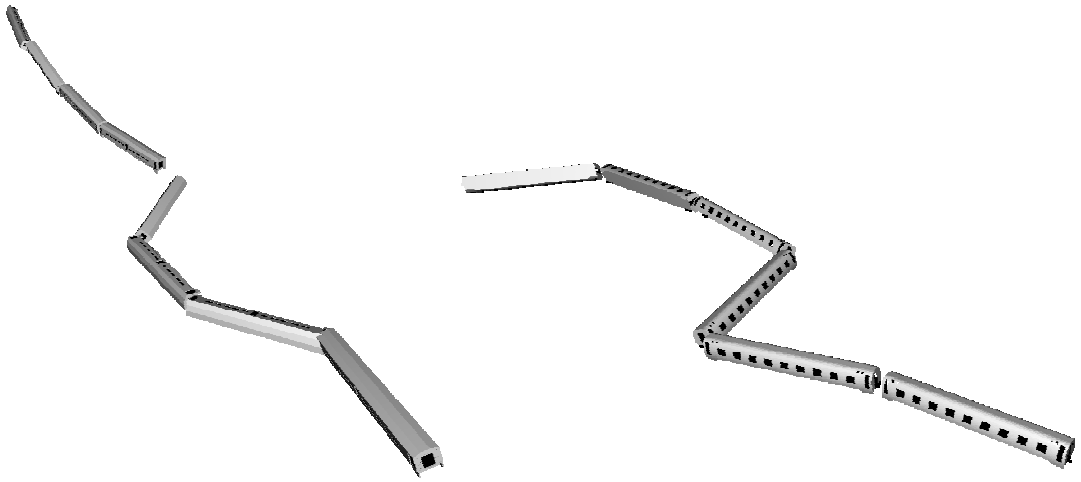

Figure 8. Pose and Position Reconstruction results.

Figure 9. Reconstructed accident scene in Bronx

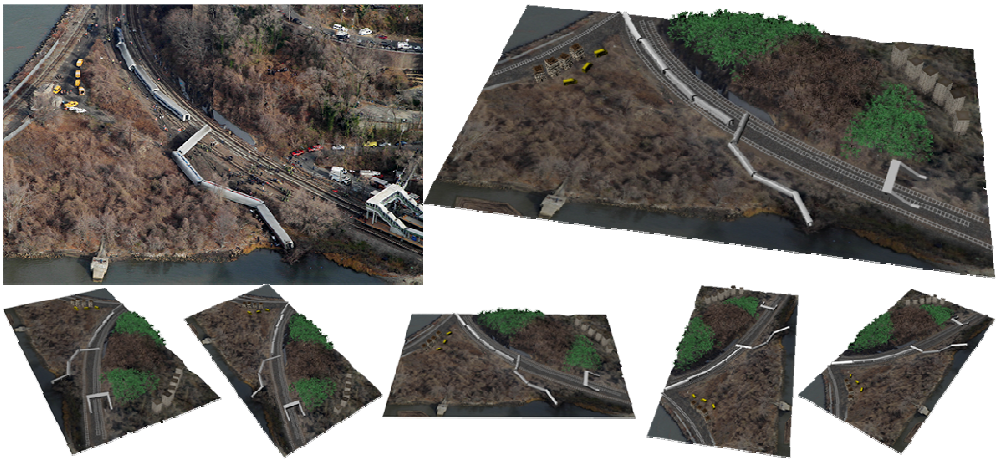

Figure 10. Comparison between real and reconstructed accident scene

Figure 11. Reconstructed accident scene in Philadelphia

3 Conclusion

In this paper, we propose a framework to rapidly reconstruct the 3D digital scene of a train accident, and develop a 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) To meet the particular needs of railway industry, we tailor a 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.

(2) 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, or there is no significant features that can be recognized (Section 0). Fortunately, by the introduction of geometric constraints and topographic information, we can ensure the accuracy of our algorithms even in the worst situations.

(3)With proposed method and applying 3D visualization technology to rendering the environment, the 3D scene 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.

References

1. Zarboutis, N. and N. Marmaras, *Searching efficient plans for emergency rescue through simulation: the case of a metro fire.* Cognition, Technology & Work, 2004. **6**(2): p. 117-126.
2. Aytekin, C., et al., *Railway Fastener Inspection by Real-Time Machine Vision.*
3. Lai, Y.C., 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**(3): p. 353-364.
4. Sawadisavi, S., *Machine-vision inspection of railroad track*. 2009, University of Illinois at Urbana-Champaign.
5. Schlake, B.W., et al., *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**(F5): p. 499-511.
6. Dai, F. and M. Lu. *Photo-based 3D modeling of construction resources for visualization of operations simulation: case of modeling a precast façade*. in *Proceedings of the 40th Conference on Winter Simulation*. 2008. Winter Simulation Conference.
7. Yang, M.-D., et al., *Image-based 3D scene reconstruction and exploration in augmented reality.* Automation in Construction, 2013. **33**: p. 48-60.
8. Bhatla, A., et al., *Evaluation of accuracy of as-built 3D modeling from photos taken by handheld digital cameras.* Automation in construction, 2012. **28**: p. 116-127.
9. Udayan, J.D., H. Kim, and J.-I. Kim, *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**(1): p. 12-27.
10. Frankl, A., et al., *Detailed recording of gully morphology in 3D through image-based modelling.* Catena, 2015. **127**: p. 92-101.

11. Zancajo-Blazquez, S., et al., *An Automatic Image-Based Modelling Method Applied to Forensic Infography.* PloS one, 2015. **10**(3): p. e0118719.
12. Fernández‐Hernandez, J., et al., *Image‐Based Modelling from Unmanned Aerial Vehicle (UAV) Photogrammetry: An Effective, Low‐Cost Tool for Archaeological Applications.* Archaeometry, 2015. **57**(1): p. 128-145.
13. Brown, M. and D.G. Lowe, *Automatic panoramic image stitching using invariant features.* International journal of computer vision, 2007. **74**(1): p. 59-73.
14. Juan, L. and O. Gwun, *A comparison of sift, pca-sift and surf.* International Journal of Image Processing (IJIP), 2009. **3**(4): p. 143-152.
15. Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision, 2004. **60**(2): p. 91-110.
16. Levenberg, K., *A method for the solution of certain problems in least squares.* Quarterly of applied mathematics, 1944. **2**: p. 164-168.
17. Marquardt, D.W., *An algorithm for least-squares estimation of nonlinear parameters.* Journal of the Society for Industrial & Applied Mathematics, 1963. **11**(2): p. 431-441.
18. Moré, J.J. and D.C. Sorensen, *Computing a trust region step.* SIAM Journal on Scientific and Statistical Computing, 1983. **4**(3): p. 553-572.
19. CityEngine, E. and *ESRI CityEngine* *http://www.esri.com/software/cityengine*. 2015 [cited 2015; Available from: http://www.esri.com/software/cityengine.
20. Heikkila, J. and O. Silvén. *A four-step camera calibration procedure with implicit image correction*. in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. 1997. IEEE.
21. Bouguet, J.-Y., *Camera calibration toolbox for matlab.* 2004.
22. Triggs, B., et al., *Bundle adjustment-a modern synthesis.* Vision Algorithms: Theory and Practice. International Workshop on Vision Algorithms. Proceedings (Lecture Notes in Computer Science Vol. 1883), 2000: p. 298-372.
23. Peralta, E., *Commuter Train Derails In The Bronx Killing 4.* 2013.
24. PHILLIPS, S. *What's in those tank cars near the Amtrak derailment?* 2015 [cited 2015 MAY 13, 2015]; Available from: http://stateimpact.npr.org/pennsylvania/2015/05/13/whats-in-those-tank-cars-near-the-amtrak-derailment/.