Railway Tracks Detection of Railways Based On Computer Vision Technique and GNSS Data

Adham, Mahmoud[1] , Mohamed, Mohamed G. [1], El Shazly, Adel[1]
[1]Civil Engineering, Faculty of Engineering, Cairo University
Giza Governorate 12613, Egypt
First. mahmoudadham4040@yahoo.com; Second. mohamedgomaa@cu.edu.eg
Third. Adel_shazly@hotmail.com

Abstract – Railway networks are major components of any country's infrastructure. In Egypt, the length of the railway network is 9570 km, where, 85% of the lines' movements are operated by mechanical signals. In July 2017, the Egyptian government launched a series of railway infrastructure projects aimed to modernize the Egyptian National Railways which includes the infrastructures rehabilitation for the network. This research focuses on developing an efficient low-cost framework using video camera and computer vision techniques for automatic railway track detection. A computer vision technique is used for automatic detection of railway tracks. Interior orientation parameters are extracted as part of the camera calibration task. Bundle adjustment calibration technique is used to compute the exterior orientation parameters based on selected ground control points. Hence, the eye fish effect of the images is removed and orthogonal images for the railway tracks are constructed using matching and feature extraction algorithms. The framework is tested on a dataset of a railway network with an approximate length of 20 km. The accuracy of the results is compared with a field survey data conducted to the same area using conventional surveying instruments such as Total station and Global Navigation Satellite System (GNSS). The proposed framework enables automatic extraction of railway tracks and its relationship with surrounding features, which contributes to quality control and assurance procedures for field collected data. The framework also offers a method for continuous and low-cost monitoring of the railway network. This will help to rapidly assess maintenance requirements for the network.

Keywords: Railway track detection, camera calibration, 3D reconstruction, video data acquisition, computer vision, and photogrammetry.

1. Introduction

Railway networks are major components of any country's infrastructure. Millions of citizens utilize the railway to travel and transport their belongings and goods. Egyptian National Railways (ENR) has a length of 9570 km. It is the second oldest railway network in the world, whereas, 85% of the lines' movements are operated by mechanical signals. In July 2017, the Egyptian government launched a series of railway infrastructure projects aimed to modernize the ENR including the infrastructures rehabilitation. This requires searching for an optimal solution for continuous monitoring, inspection, and maintenance of railway components such as track, switches, signs and all mechanical systems. Traditionally, railway tracks and other railway components are mapped using traditional surveying tools such as total station and Global Navigation Satellite System (GNSS). These methods provide accurate measurements, albeit expensive and time consuming.

The development of computer vision techniques and video cameras have paved the way for obtaining a low-cost data collection method; such as automatic detection of railway tracks. Automatic railway detection is used for position determination of the rail lines ahead of the train and is considered an essential task in a railway inspection framework. This task is implemented to determine any defects along the railway network. Computer vision using non-contact camera measuring is utilized to obtain considerable amount of information and has been widely applied to inspection, monitoring, system quality, and, safe tools [1]. Unmanned aerial vehicle (UAV), also known as drones, is one of the developments platforms. UAV can be mounted with more than one sensor to be used in civil purposes. One of the significant factors in imagery based-UAV compared with ground imagery is the vertical information for parcel captured. Many studies were conducted for monitoring and inspecting railway components based on mounted cameras using different platforms such as being mounted on the front of locomotives or manufactured carts.

One of the significant factors in imagery based-UAV compared with the terrestrial imagery is the vertical information for parcel captured. Singh et al. [2] presented an approach using computer vision to detect railway tracks using a camera fixed on a drone. The results for gauge measurements were reasonable for each image separately. On the other hand, Banic et al. [3] suggested an image processing algorithm with advanced drone imagery system for automated railway tracks monitoring and detection. Results from this method depends primarily on canny edge detectors applied on captured images to detect rail lines and classify the detected edges [4]. The detector uses the mean distance k-nearest neighbour (kNN) machine learning approach to indicate the detected line is a rail line or not.

Wohlfeil [5] presented an approach for railroad track extraction using image processing methods, however, this method was limited to only short distances along the track. Using a video camera mounted in the front of the train, Maire [6] demonstrated a procedure to detect the rail lines by transforming perspective images captured to a bird's-eye view to be simple in extraction for long distances. Nevertheless, every frame was treated separately while increasing the tested region. Karakose et al. [7] used some image processing algorithm on the captured images to extract the two rail lines by only one camera placed on the roof of a locomotive and determine the defects on these rails along the track based on the constant width of the track. Data georeferencing were not taken in account while they check width of the track on pixel system.

In literature, there are many proposed systems for detection of rail components. Li et al. [8] proposed an automatic vision for railway system inspection. They used four cameras mounted on the back of a truck that moved with a 16 kph on the railway track to collect the video data. Rail components detection were done based on Hough transformation and Sobel edge detection [9]. Another approach was performed to detect the damaged structure of the railway using four cameras and odometer instruments [10]. Where, Belkhade and Kathale [11] proposed a system to identify the missing railway track components such tie plates, spikes and anchors for installing new parts instead.

Recent researches in the literature used video imaging for moving platforms during the stage of collecting data. Video data were extracted to large numbers of frames based on the camera's rate of recording. Key frame extraction methods are used to minimize the number of frames. This process depends on automatic feature matching methods between every two consecutive frames. Several images matching algorithms can be found in the literature. The three main stages of matching two images are as follows:

1) Feature detection in each image as a Point of Interest (POI).
2) A distinguished descriptor define for each POI.
3) The different detected POI is matched using their descriptor.

One of the most often used method for feature extraction is Speeded-up Robust Features (SURF). SURF is a fast detector and descriptor of the performing scale and rotation invariant POI [12].

The main objective of the research is to propose a low-cost mapping system using video data and can be operated without any interruption of the trains circulation and schedules. As discussed in the literature, video cameras are more useful than images capturing as the film is recorded while the train moving at various speeds. For this study, we focus on extraction of the railway track from the orthophoto generated from sequences of perspective images captured using a camera mounted on the front of a locomotive. This paper presents the proposed algorithm based on image processing and computer vision methods to extract the georeferenced rail lines from the scenes captured and check the track width along the track.

2. Methodology

The methodology flowchart is summarized in Figure 1. The methodology is composed of three main components: video data analysis, 3D reconstruction model, and automated track detection framework. These components are presented in the following sections
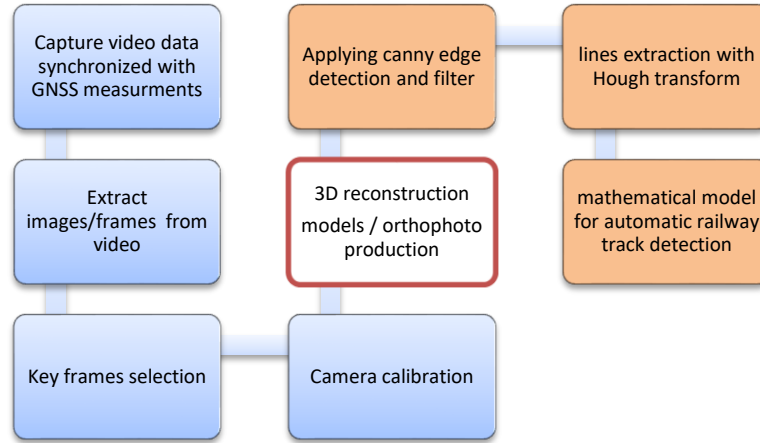
Fig.1: The methodology flowchart

2.1. Images Extraction from Video and Key Frames Selection

In order to achieve the objectives of this research, the images obtained from a recorded video were extracted using a developed MATLAB® code. Extracted images were proceeded to extract useful information from them. The video captures data on a rate of 30 frames per second (FPS) or more, implying 1800 images can be extracted in a 1-minute film. Dealing with all the extracted images leads to increasing the processing time. A key frame algorithm was developed to automatically select the minimum number of frames to achieve an acceptable overlap percentage between every two successive frames. The overlap percentage between the frames was calculated based on pixels count after applying feature points detection and matching using SURF technique. Matched points from the two consecutive frames used to determine the Images Transformation matrix (ITF). The first frame is used as a reference and the second frame is transformed with ITF. The transformed frame is compared with the reference using pixel coordinates and the overlap between the frames is calculated. Several overlapping ratios between each two consecutive images were tested to select the best ratio. Logically, decreasing the overlap ratio leads to fewer extracted key frames and as a result it could affect the accuracy of the produced maps. On the other hand, increasing the overlap ratio leads to increasing the number of key frames and thus, increasing the processing time.

2.2. Camera Calibration

For an accurate 3D-model based on a 2D imagery, the systematic error associated with the camera must be modelled and compensated prior to the 3D reconstruction process. Generally, there are two types of camera calibrations: geometric and radiometric. Geometric calibration refers to the process of determining the camera's focal length, principal point positions and lens distortion coefficients, which affects the positions of image points. These parameters are called interior orientation parameters (IOPs). The radiometric calibration refers to the process of determining the accuracy of the captured image's grey value, which affects the true reflectance in this section, we focus on the geometric camera calibration. In literature, many techniques can be applied. An open source (MATLAB® toolbox calibration algorithm) [13] utilized pinhole camera model expressed using equation (1)

$$w[xy1] = [XYZ1]\begin{bmatrix} R \\ T \end{bmatrix}\begin{bmatrix} fx & 0 & 0 \\ 0 & fy & 0 \\ Cx & Cy & 1 \end{bmatrix} \tag{1}$$

Where; (x, y): image coordinates of point, (X, Y, Z): world coordinates, W: scale factor, R: 3D camera rotation matrix, T: camera translation matrix to the world coordinates. Given that fx and fy can be computed using fx = F*Sx, fy = F*Sy. Where; F= focal length in millimetres; (Sx, Sy) = Size of the pixel in millimetres, (fx, fy) = Focal length in pixels, and, (Cx, Cy) = principal point in pixel.

Equation 1 is considered as an ideal pinhole camera model. However, there are many exceptions such as a) the light rays bend closer to the edges of the lens than at its optical center appear the radial distortion. b) The camera lens isn't parallel with the image plane appear the tangential distortion. The distorted points are shown in Equations (2) and Equation (3) respectively[14].

$$u_{distorted} = u(1 + k_1r^2 + k_2r^4 + k_3r^6) + (2p_1u\,v + p_2(r^2 + 2u^2)) \tag{2}$$
$$v_{distorted} = v(1 + k_1r^2 + k_2r^4 + k_3r^6) + (p_1(r^2 + 2v^2) + p_2u\,v) \tag{3}$$

Where; (u, v) = undistorted position of pixel, k_1, K_2 and k_3 = coefficients of radial distortion, p_1 and p_2= the coefficients of lens Tangential distortion, and $r^2 = x^2 + y^2$.

To obtain the above mentioned parameters of camera calibration, multi images have been captured for the standard checkerboard pattern with a large fill for the image form different camera poses.

2.3.3 D-Reconstruction Model

The photogrammetric algorithms and computer vision techniques are used to reconstruct an accurate 3D model. In this section, a Context Capture software is used for automatic 3D reconstruction (https://www.bentley.com/en/products/brands/contextcapture).Several steps are implemented to reconstruct the 3D model. The interest points are detected and matched using SURF feature detection algorithm. Bundle adjustment algorithm is used to relate the image coordinates of the interest points, their corresponding ground coordinates (control points), the IOPs of the camera, and the exterior orientation parameters (EOPs) of the captured images. Bundle adjustment is a non-linear least square estimation which depends on the collinearity equations to calculate the cost function for the mapping process. The model output is a random sparse point cloud which represents an approximate 3D model for the extracted images. The steps for georeferencing the 3D model includes selecting and measuring ground control points from the surveying maps. Intensive matching features are performed to generate dense point cloud that can be exported to 3D mesh containing all visible details in the original images. Many models were reconstructed based on different parameters to evaluate the most accurate model. The best 3D model was selected. And an orthogonal photo was generated and extracted into a Geographical Information System (GIS). The orthogonal photo of the model solved the problem of perspective projection and combines the extracted images into a georeferenced image similar to the map.

2.4. Detection of Railway Tracks

The generated orthophoto was classified into two main classes using a supervised classification: tracks or non-tracks. This procedure can be developed for detection of other several features. The classified image was converted to a grey scale image while operation of detection can be faster than colored image. Canny edge detection presented the most accurate results. The algorithm of canny passes through four main stages as follows [15];

1) Use Gaussian filter to smooth the input orthophoto and remove its noise.
2) Intensity direction and magnitude are calculated for the image to detect edges.
3) Eliminate unwanted edges by perform non-maximum suppression.
4) Thresholding method is performed to identify pixels that must counted as edges.

Morphological dilation function in MATLAB® can be performed to check and stretch the image shapes after canny detection. After edge detection procedure, Hough transformation is applied to determine the straight lines in the image.

It is noticed that Hough transformation faces a difficulty in dealing with pixel coordinates. Therefore, it is preferable to transform the lines into polar coordinate system (ρ, θ). Equation (5) shows the polar coordinate system, whereas, Equation (6) shows the traditional straight-line equation (Plane coordinate system).

$$\rho = x\cos(\theta) + y\sin(\theta) \tag{5}$$
$$Y = ax + b \tag{6}$$

Hough transformation algorithm depends on converting the point in a plane coordinate system into a polar coordinate system. All straight lines can be detected from the orthogonal image using a specific threshold. A constraint based on the slope of the detected lines for the track and the constant upper thickness (t) of the rail is utilized for better

detection of the tracks. By using the equation of the normal to line as shown in Equation (7) , we can identify all pixels in the image that are far away from the rail line by its thickness(t) as described in the flowchart shown in Figure (2).

$$D = (ax + by + c)/ \sqrt{a^2 + b^2} \tag{7}$$

Where; D= normal distance from railway track, a = slope of track, x & y = location of point in rail line, and,   b & c = coefficients.
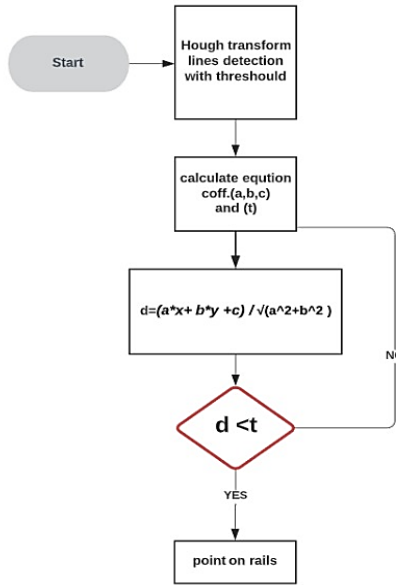

Fig. 2: Detection of track point's flow chart

3. Experimental Work and Results
3.1. Data collection and Camera Calibration
An action camera called a GOPRO hero 5 black camera is used for video data collection. A video was recorded with resolution 1920*1080 pixels on a rate of 30 FPS with digital video stabilization. A sample of the captured frame is shown in Figure 3-a. The camera system was mounted on the front of locomotive and GNSS on the center of the locomotive (refer to Figure 3-b). The chosen camera recorded the video for the railway while the train moved with different speeds. For performance evaluation with respect to the speed of the train, Video data is classified into small video clips based on different constant speed. For each clip, sequence frames were extracted. The framework is tested on a dataset of a railway network with an approximate length of 20 km. A traditional surveying using total station and GNSS was performed for all sectors.

Fig. 3: a) sample of captured frame   b) the used locomotive and GNSS system

Camera calibration using the MATLAB® toolbox is performed using a black and white checkerboard pattern with a squared length 30mm as shown in Figure 4 (a). A set of 20 images at different camera poses for the checkerboard were captured and the interior orientation parameters (IOPs) and lens distortion were calculated based on Zhang's method Table [1] summarizes the camera calibration parameters. Figure 4(a) shows the original checkboard image before and Figure 4(b) shows the image after remove lens distortion.
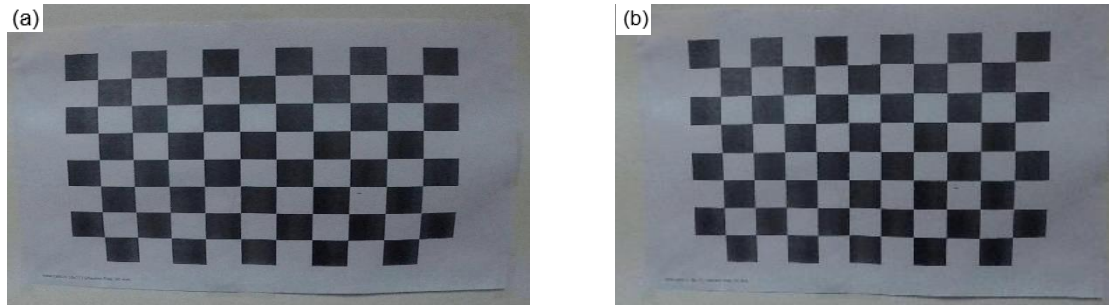

Fig. 4: a) original checkerboard image   b) undistorted checkerboard image

Table 1: camera calibration parameters

Focal length (pix)		Principle point location(pix)		Radial distortion coefficients			Tangential distortion coefficients	
Fx	Fy	X_o	y_o	K1	K2	K3	P1	P2
1014.40	1025.30	966.32	556.42	-0.2485	0.0546	-0.0889	-0.00045	-.00053

3.2. Results

To evaluate the proposed method, a video has been recorded to the railway using a single camera. A sequence of frames extracted from video. Interior orientation parameters and distortion parameters (I.O.Ps) were used as initial values. Extracted fames, IOPs and collected ground control points used to reconstruct the 3d model using context capture software. To select the optimum parameters, many 3D models have been reconstructed at different selected speed and different overlap. The quality of the 3D model was evaluated. The root mean square error (RMSE) for positions of check points is between 0.05 m to 0.32 m (refer to Figure 5). The video collected at speed 10 kph with the maximum overlap
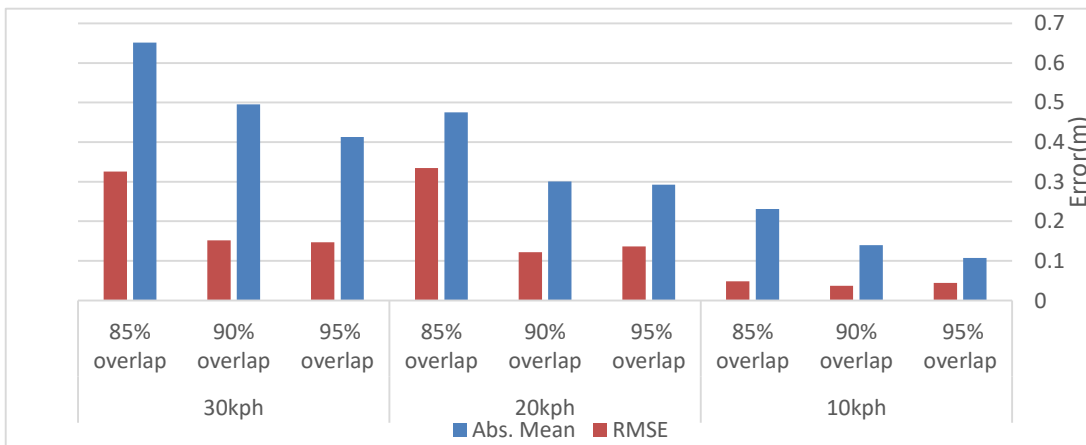

Fig. 5: position error comparison based on train speed and image overlap

90 % achieve the best 3D model. Figure 6 shows the 3D reconstructed models for the same position and speed of model and at different overlap.
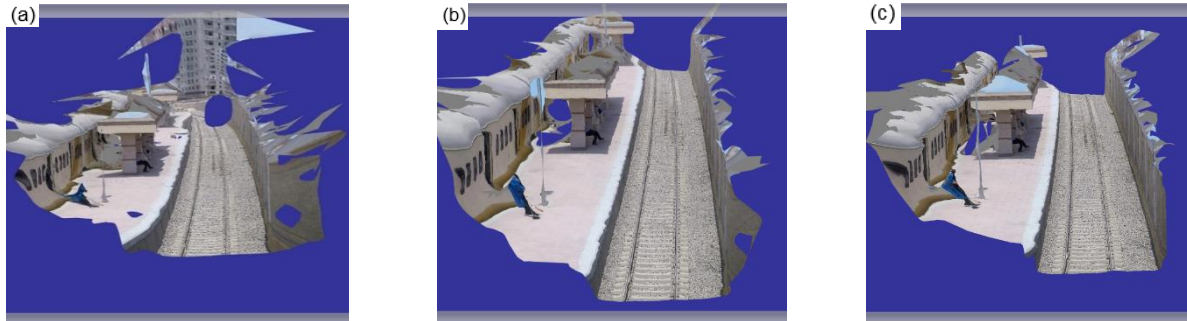

Fig. 6: Railway 3D-model at 10km/hr : a) with overlap85% b) with overlap90% c) with overlap95%

The orthophoto image from the 3D model was generated and shown in Figure (7-a). This image was imported into GIS as a georeferenced image. The supervised classification is applied to classify the data into track features and non-track features using GIS software (refer to Figure 7-b). Still, the classified image contains the non-track feature and classified into tracks. These features need to be removed for the automatic track detection.
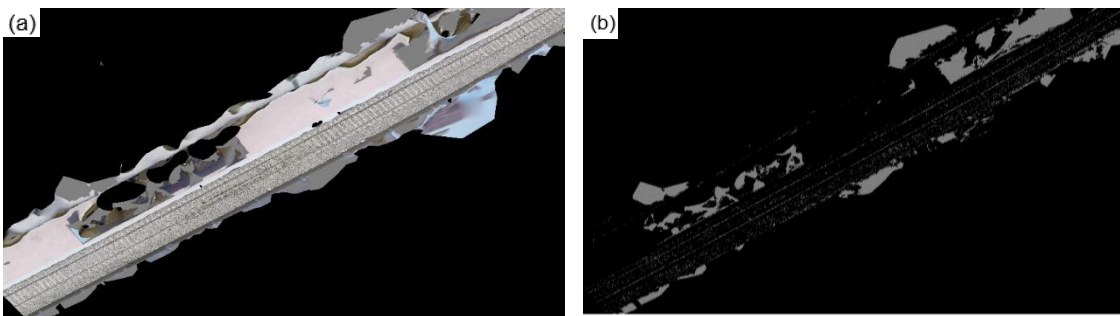

Fig. 7: a) orthogonal image   b) classified image

The automated track detection framework is applied to detect the railway tracks. The results of each step of the framework are shown in Figure 8. Figure 8 (a) shows the resulting image of canny edge detection method on the orthogonal photo. Dilate operation was performed to expand shapes in canny image in Figure 8 (b). The Hough transform algorithm was applied and the result is the green lined shown in and in Figure 8(c). Only small sectors of the tracks are extracted after this stage. Mathematical model with threshold constraint is applied for identify and classify the interested points as shown in the white. Points presented in Figure 8(d). Hough transform is applied on the previous image for final track detection. As shown in Figure 8 (e), the tracks are automated detected accurately with an RMSE equal 0.05 m.
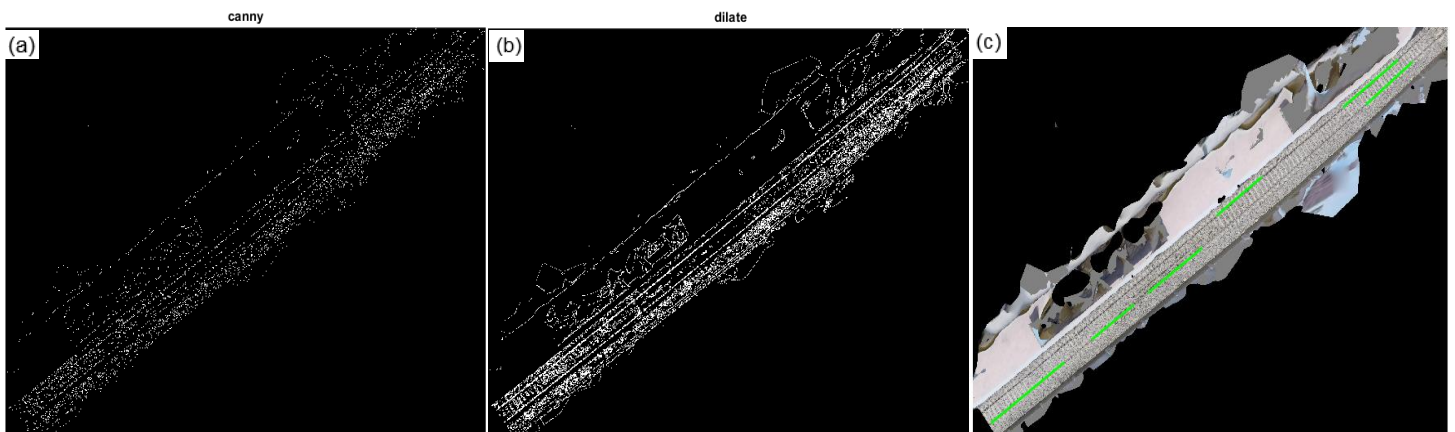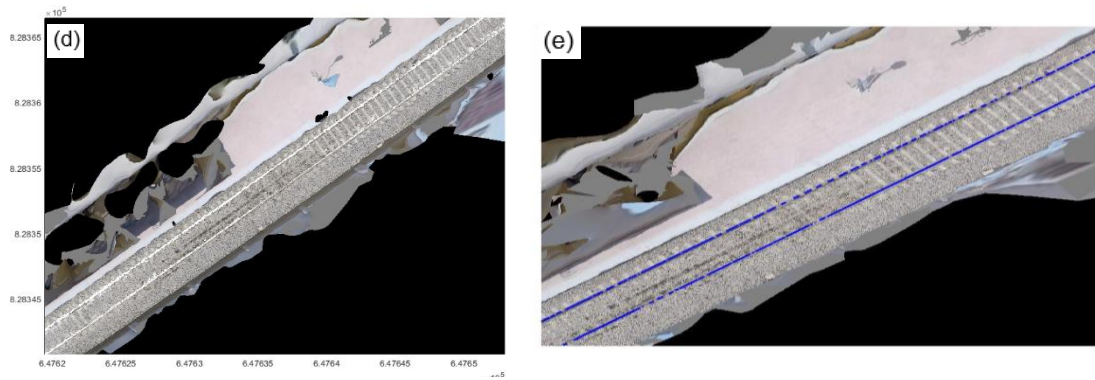
Fig. 8:a) image based canny edge detection b) image after dilate operation c) image with hough transform lines d)image based mathematical model e) image results based enhanced hough transform.

4. Conclusion and Discussion

In this paper, a low cost automatic track detection framework based on computer vision algorithm has been proposed. The proposed algorithm enables extracting railway tracks from images captured in the front of a moving locomotive. This method provides accurate, cost saving and fast approach of extracting the railway tracks. The RMSE of the extracted track is 0.05 m. The recommended train speed for reasonable video data collection is 10 kph which is considered relatively a slow motion. The proposed framework for imaging produces a georeferenced orthogonal photo to facilitate detection of the railway track with the actual dimensions. The framework also offers continuous and low-cost monitoring of the railway network, and consequently, rapidly assesses maintenance requirements for the network. In the future, improvement of the algorithm is required to provide more accurate extraction for railway tracks with high speeds and to detect different components of the railway.

References

[1] S. Zheng, X. Chai, X. An, and L. Li, "Railway track gauge inspection method based on computer vision," *2012 IEEE Int. Conf. Mechatronics Autom. ICMA 2012*, no. 1, pp. 1292–1296, 2012, doi: 10.1109/ICMA.2012.6284322.

[2] A. Kumar, A. Swarup, A. Agarwal, and D. Singh, "Vision based rail track extraction and monitoring through drone imagery," *ICT Express*, vol. 5, no. 4, pp. 250–255, 2019, doi: 10.1016/j.icte.2017.11.010.

[3] M. Banić, A. Miltenović, M. Pavlović, and I. Ćirić, "INTELLIGENT MACHINE VISION BASED RAILWAY INFRASTRUCTURE INSPECTION AND MONITORING USING UAV," vol. 17, pp. 357–364, 2019.

[4] M. Jiang, "Digital Image Processing," 2009.

[5] J. urgen Wohlfeil, "Vision based rail track and switch recognition for self-localization of trains in a rail network," no. Iv, pp. 1025–1030, 2011.

[6] F. Maire, "Obstacle-Free Range Determination for Rail Track Maintenance Vehicles," no. December, pp. 7–10, 2010.

[7] M. Karakose, O. Yaman, M. Baygin, K. Murat, and E. Akin, "A New Computer Vision Based Method for Rail Track Detection and Fault Diagnosis in Railways," vol. 6, no. 1, pp. 22–27, 2017, doi: 10.18178/ijmerr.6.1.22-27.

[8] Y. Li, S. Member, H. Trinh, N. Haas, C. Otto, and S. Member, "Rail Component Detection , Optimization , and Assessment for Automatic Rail Track Inspection," vol. 15, no. 2, pp. 760–770, 2014.

[9] T. Acharya and A. K. Ray, *Image processing: principles and applications*. John Wiley & Sons, 2005.

[10] N. Haas and S. Pankanti, "Multisensor Evidence Integration and Optimization in Rail Inspection," no. Icpr, pp. 886–889, 2012.

[11] A. Belkhade and S. Kathale, "AUTOMATIC VISION BASED INSPECTION OF RAILWAY TRACK : A REVIEW," pp. 12–15, 2014.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[13] Z. Zhang, "A Flexible New Technique for Camera Calibration," vol. 1998, no. 11, pp. 1330–1334, 2009.

[14] "What Is Camera Calibration_ - MATLAB & Simulink. (n.d.).," *Mathworks.com*. 2020. https://www.mathworks.com/help/vision/ug/camera-calibration.html?requestedDomain=

[15] G. A. (1994). D. image processing: principles and applications. W. N. Y. Baxes, Gregory ABaxes, *Digital image processing: principles and applications*. Wiley New York, 1994.