3.1 Introduction

Computer automation of RSE involves the determination of individual parts for examination to identify defects from the videos of a moving train. Video frame segmentation by using the CV-AC and CV-MDG-AC models results in a full-bogie binary image, making it impossible to track individual parts. Segmenting individual parts and tracking their shapes along the length of the train is a challenging task. This challenge can be overcome by using shape prior seeds (SP-CV-AC) as the destination contour from individual parts of the bogie in the CV-AC model. Spatial distances are used to propel the initial contour toward the final shape contour. The results demonstrate the quality of the video segmentation algorithm based on destination seed shape priors. The factual segmentation score (FSS) between the shape prior segments and manually segmented portions of the rolling stock was calculated to access the quality of the proposed segmentation algorithm. We further compared the shape prior segmentation model with ACs without shape priors to specify the importance of shape prior models in complex image processing tasks related to intelligent maintenance systems with computer vision.

Although RSE is a primary, regularly performed task, few systems have been developed for monitoring components of the rolling stock by using computer vision models [55, 56]. Computer vision-based RSE with the fuzzy weighted logarithmic least squares method has being investigated. The fuzzy model is based on the TFN. The weighted logarithmic least square method classifies the segmented rolling parts. The model evaluation process is described in [33].

In [34], the researchers proposed an embedded system for intelligent monitoring of rolling stock for enhancing safety during rail transit. The paper reviews conditions for controlling undesired rolling stock behavior during travel by using detection and prediction analysis methods.
The study reported in [35] explored PM scheduling to reduce the budget allocation for RSE. Genetic algorithm and simulated annealing were employed to determine optimal maintenance intervals to maximize the rolling stock life span. The study further computed the optimal number of spares that satisfied the requirements and availability.

A study reported in [36] reviewed and analyzed the application of wireless monitoring systems to existing RSE systems, and the results revealed potentially equivalent benefits of the prior model. A multihop mesh network as a standardized framework for railway rolling stock monitoring was developed in this study. A localized processing network of elements was used with configurable sensing to provide temporary and semi-permanent monitoring functions of rolling stock. Moreover, wireless management of energy harvesting power was implemented [36].

Rather than focusing on the entire rolling stock, in one study, the thickness of lining-type brakes was measured for automatic rolling stock monitoring. The Hough transform was used to detect the interest points, and the brake edge was explored in the region of interest to detect the circular outline of the disk lining brake. Results revealed that the proposed method measured the thickness of the lining-type brake with a precision of 1.15 mm at a distance of 1.0 m from the camera [37]. Studies [38-41] used similar computer vision and classification techniques to determine and analyze the shape of brake shoes by using a high-speed digital camera installed on railway tracks.

Finally, this chapter summarizes the core features of the TechnicAtome experiment that was developed as a demonstrator for RATP (a Paris subway company) based on interconnected digital systems. This demonstrator is currently in operation on an MF 88 train set and is still operated with conventional relay-based systems [42].

In our previous research, we focused on segmentation of train
rolling stock [55] as a whole to identify defects. The model used in our previous study did not yield accurate outcomes because identifying individual parts for further processing was difficult. Hence, the extraction of individual parts of the rolling stock is warranted. To achieve this, we propose shape prior-based segmentation of rolling stock parts for developing a fully ARSE system. The remainder of the chapter presents a brief review on ACs, followed by a description of the shape prior-based segmentation model. Finally, results of various frames of rolling stock videos are presented to validate the effectiveness of the algorithm by computing an FSS.

3.2 Shape Prior Active Contours

ACs are categorized as a model-based segmentation method. Image space is defined by the boundaries of objects in that particular image. Energy function controls the contour movement in the spatial domain. Earlier models of ACs are prone to topological disorders and are vulnerable to initial conditions. However, level sets can efficiently address changes in the objects. ACs are dependent on image gradients; Kass introduced a mathematical expression called snake. To achieve active segmentation, the energy function must be minimized, which was first achieved using the CV model. The energy or force function formulated using the CV-AC model was minimized using piecewise linear Mumford–Shah function [49] as described in Sections 2.3 and 2.4. The minimization problem was solved using Euler–Lagrange [84] equations and the level set function [46].

Combining the CV-AC [84] level set in Eq. 2.14 with the shape prior model proposed in [57], the shape-induced energy term can be represented as follows:

\[
E^{cv+shape} = E^{cv}(\Theta, \Phi_s^{(I)}, \Phi_s^{(E)}) + E^{shape}(\Theta, \Phi_s^{(I)}, \Phi_s^{(E)})
\]  

(3.1)
The first term denotes the data term obtained from the CV level set, and the second term denotes the prior shape energy model, which is defined as follows:

\[
E_{\text{shape}}(\Theta, \Phi^I_S, \Phi^E_S) = \int_{\Theta} (H(\Phi(x, y)) - H(\Phi_S(x, y)))^2 dxdy \tag{3.2}
\]

where \((\Phi_S(x, y))\) is the shape prior term independent of the position on the image. For multiple shape priors, the shape energy term is represented as follows:

\[
E_{\text{CV+shape}(n)} = -\log \left( \frac{1}{N} \sum_{j=1}^{N} e^{-\frac{d^2[H(\Phi),H(\Phi_{Sj})]}{2\sigma^2}} \right) \tag{3.3}
\]

Equation 3.3 can generate accurate segmentation results when the number of shape priors is typically small and capturing the statistical structure of different shapes in the observational space is difficult. In the present study, we primarily focused on single object extractions and performed simulations by using Eq. 3.2.

### 3.3 Results and Discussion

![Figure 3.1. Light Intensity variations at different times of the day.](image)

(a) Video Capture at 6.30AM  
(b) Video Capture at 12.30PM  
(c) Video Capture at 4.30PM

(a) at 6.30AM, (b)12.30 PM, (c) 4.30 PM
Figure 3.1 presents the video frames at three-time points of a day, morning, afternoon, and evening. Before inputting the video frame into the SP-CV-AC, a virtual image fusion-based contrast enhancement algorithm was implemented to improve the contrast of the image of the train bogie with the ballast (gravel or ground).

Shutter speed and F-stop in a high-speed digital camera controls the contrast during video capture. Shutter speed controls the quantity of echoed light that reaches the image sensor. F-stop controls the opening of the hole that transmits light to the sensor. Typically, the F-stop approximates to geometric sequences equivalent to the powers of root 2 [55]. This concept is used to produce computer-generated video frames from uneven contrast frames. These frames are fused onto one high-contrast image in the wavelet domain. The entire method is rapid and robust to brightness variations in the video frame, as proposed in our previous study [55].

The contrast-enhanced video frames for a sample frame are presented in figure 3.2.

Figure 3.2. Virtual image contrast enhancement for compensating light intensity variations at different times of the day. (a) at 6.30AM, (b) 12.30 PM, (c) 4.30 PM
The use of regular CV-ACs on contrast-enhanced video frames yields a good segmented frame [55] compared with many other edge detection techniques. Figure 3.3(a) presents the segmented output from the CV-ACs without shape priors. The initial contour covers the entire spatial domain for faster convergence of the initial contour toward the gradient magnitude edges of the rolling frame. The segmentation is not focused on a particular part, but covers the entire video frame. Visual observations reveal that the entire frame is covered during segmentation, thus making it impossible to extract major parts during RSE for defect identification. Classification algorithms also failed to elucidate the segmented rolling stock in figure 3.3(b).

Figure. 3.3(a). Chan Vese AC segmented video frame Segmentation process

Figure. 3.3(b) Final Result
Shape prior CV-ACs propel the contour toward the object of interest in the video frame. Our proposed model does not focus on the shape of the initial contour \( \Phi_0(x,y) \). The only constraint is that the seed contour should be placed near the rolling part to be observed by selecting a particular frame of interest. The shape contour \( \Phi_s(x,y) \) is precisely cut from the original video frame in which the full view of the bogie is visible. The spatial error between two contours acts as a controller to focus the initial contour toward the rolling part. The frame presented in figure 3.4 was selected for testing the algorithm.

Figure. 3.4 Test frame showing a bogie of train moving around 30kmph during rolling stock examination by our proposed high-speed camera model

Figure. 3.5. Proposed segmentation model is tested on parts pointed by arrows on the bogie of a rolling stock

Individual parts are manually extracted from the frame by using multimedia tools as shape contours. Figure 3.5 presents the parts of
the bogie that are extracted for testing the proposed shape prior AC model.

The frame is contrast enhanced and is inputted into the shape prior CV algorithm. Figure 3.6 presents the initial contour, which is a rectangular area comprising seed pixels for selecting the spatial pixels near the rolling part. By using the initial contour in figure 3.6, the extreme right binding screw was selected. According to railway engineers, this screw holds the bogie in place and does not allow the bogie to move during transit.

![Rolling Stock Frame](image)

Figure 3.6. Rolling stock frame showing the part to be segmented along with the initial contour acting as seed

A uniform distance transform provides a shape prior contour. From Eq. 3.2, the spatial error propels the initial seed contour toward the shape prior contour, thus yielding a focused segmentation result. Figure 3.7 presents the step-by-step process of binding screw extraction at the 15th, 18th, 29th, 51st, 73rd, 92th, 107th, 125th, 144th, 175th, 191th, and 217th iterations.

The initial contour propagates toward the shape prior model of the binding screw, as shown in figure 3.8.
Figure 3.7. Shape prior evolution of CV Active contour targeting the binding screw

Figure 3.8. Shape prior model of right binding screw

Figure 3.9 presents the segmentation outputs at the previously mentioned iterations. Figure 3.10 presents the final segment of the binding screw after morphological dilation was performed to close small unconnected pixels. Figure 3.11 is the complementary image of figure 3.10; it exhibits more efficient contrast and involves less computational processing when too many zeroes are involved.
Figure 3.9 Segmentation results of the binding screw at various iterations during propagation of initial contour.

Figure 3.10. Final Segment of the Bogie binding.

Figure 3.11. Compliment of figure 3.10.
A comparison between figure 3.11, which is the simulated output, and figure 3.8, which is the manually segmented output, demonstrated the usefulness of the shape prior model for segmenting bogie parts of a train moving around 30kmph.

Figure. 3.12. Testing on a different bogie at noon, under bright light

Figure. 3.13. Comparing the simulated results of proposed segmentation algorithm with ground
The mean square error between the two images is less than 7%, indicating that the simulated result and GT are 93% similar. Furthermore, Figure 3.13 presents a comparison between the simulated results and GT for each simulated part.

![Image of image comparisons](image-url)

**Figure 3.14.** Comparing the simulated results of proposed segmentation algorithm with ground truth for various parts of train rolling stock captured in figure 3.13

Figure 3.13 presents the similarities between the GT rolling parts and simulated rolling parts obtained using shape priors. The computation of mean square error between the two similar parts determines the similarity between the simulated part and the manually segmented part in the database. Mean square error calculation for an entire video of 10,000 frames with only seven dictionary parts produced an average mean square error of approximately 0.11, which indicates that the average similarity between the segmented parts was approximately 89%.
Furthermore, the proposed algorithm was tested on a video frame captured at 12.30 PM for six parts, as shown in figure 3.14. A comparison between the simulation segmented rolling stock parts and manually segmented GT parts revealed a mean square error of 9%.

This indicates that both the images were 91% similar. For almost all the parts, the similarity fluctuated by $\pm 2\%$. Testing on the video frames captured at 4.30 PM yielded similar results.

The first shortcoming of the proposed segmentation algorithm is the placement of the initial contour close to the region of interest. This reduces the number of iterations required for the initial contour to reach the final shape. Second, a shape prior dictionary must be created for every shape to be segmented at different positions of the train movement in the video.

3.4 Conclusion

In this chapter, a shape prior segmentation model was proposed for segmenting parts of a train rolling stock moving around 30 kmph. The CV-AC model was upgraded with the shape prior term. Seven parts of a train were segmented using a video with 10,000 frames. The similarity index with mean square error between the simulated and dictionary segments was approximately 89%, which indicates that the proposed algorithm is robust to brightness variations during video capture. However, the drawback of using this algorithm was that different shape priors must be provided for the entire length of the train. In the following chapter, we propose a level set that can address the deformation of parts of an RS frame with a single shape prior model for the entire length of the train.