IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 4, 2018, accepted October 4, 2018, date of publication October 23, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877712

Unifying Boundary, Region, Shape into Level Sets for Touching Object Segmentation in Train Rolling Stock High Speed Video

N. SASIKALA, P. V. V. KISHORE, (Senior Member, IEEE), CH. RAGHAVA PRASAD, E. KIRAN KUMAR, (Student Member, IEEE), D. ANIL KUMAR, (Student Member, IEEE), M. TEJA KIRAN KUMAR, (Student Member, IEEE), AND M. V. D. PRASAD

Biomechanics and Vision Computing Research Center, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur 522502, India

Corresponding author: P. V. V. Kishore (pvvkishore@kluniversity.in)

ABSTRACT Traditional level sets suffer from two major limitations: 1) unable to detect touching object boundaries and 2) segment partially occluded objects. In this paper, we present a model and simulation of a level set functional with unified knowledge of objects region, boundary, and shape models. The simulations of the proposed model were tested on high-speed videos of the train rolling stock for bogie part segmentation. The proposed model will resolve single- and multi-object segmentation of touching boundaries and partially occulted mechanical parts on a train bogie. Simulations on high-speed videos of four trains with 1 0720 frames have resulted in near perfect segmentation of 10 touching and occluded bogie parts. The proposed model performed better than the state-of-the-art level set segmentation models, showing faster and more accurate segmentations of moving mechanical parts in high-speed videos.

INDEX TERMS Level sets, train rolling stock, automated maintenance, hybrid high speed video segmentation, computer vision based condition monitoring.

I. INTRODUCTION

Rolling stock examination (RSE) is a manual checking of bogie parts, when train is moving at 30KMPH, to identify accident causing defects. Here manual checking involves a team of 3 humans putting their trained eyes, ears and intelligence to the limits. Fig.1 shows the under carriage of a moving train called as bogie in Indian Railways. The system operates on all passenger trains at every train station across Indian subcontinent. Manual RSE safeguards trains from accidents while moving at high speeds. This work is a first step towards building a system to replace trained human eyes with high speed cameras and intelligence with processing algorithms.

The objective of this work is to develop a simulation for segmenting bogie parts of a moving train. A High speed (HS) camera captures the moving bogies at 240 frames per second in a bogie video log of a train. The segmented bogie parts from the video frames are further inspected for maintenance. The best model for extracting individual parts in closely packed images was snakes [1]. The second generation snakes called as active contours were the most researched convex

FIGURE 1. Showing Manual Rolling examination in progress.

segmentation algorithms on complex natural images [2]. Previous algorithms related to segmentation of bogie parts and identification used shape prior models with shape invariance concept [3], [4]. In this work, we try to build on these algorithms to further improve the segmentation quality and speed.

There is a surge in computer vision based algorithms for human safety monitoring during transit. However, most of this research is on waterways, airways and roadways. Train rolling stock is one area that significantly dominates railway passenger safety, which is being monitored by humans 24 hours across the world with no or little automation. This work proposes to make inroads into this area by providing a vision based solution with multiple object detection algorithm on high speed video data. Largely, trains in Indian railways use fabbrica italiana de automobile torino (FIAT) train bogies and integral coach factory (ICF) bogies 1. Around 30% of the passenger trains use six operational models of FIAT and the remaining 70% of the Indian trains are fitted with ICF bogies, which are a trademark of Indian railway research board (IRRB) invention. Fig.1 shows the manual procedure followed for rolling stock examination with a train fitted with ICF bogie.

II. LITERATURE

Computer vision research is currently being driven by the software industry giants such as Google and Facebook. Research shows face recognition; image search; video compression and playback; deep learning algorithms are revenue generators for the tech sector companies. Vision research has made deep inroads into automobile with autonomous driving cars [5], transportation monitoring [6], structural quality assessment [7], agricultural [8] and manufacturing [9] industry in the last two decades. Specialized cameras such as high speed, laser and hyperspectral cameras are being used extensively in video capture for monitoring and testing applications of mechanical moving parts.

Narayanaswami [10] postulates a vision on the future of transportation in urban environments. The ideas conversed in this work helps to understand the link between essence of technology for automation in transportation save time and finances by avoiding accidents. Inspired from this work, we propose to apply computer vision models for detecting moving bogie parts in a train bogie videos. There are only a handful of works in computer vision based rail safety research, which are related to track monitoring, ballast monitoring and a little on rolling stock monitoring.

Sabato and Niezrecki [11] uses 3D digital image correlation (DIC) algorithms to inspect railroad tyres and ballast. Two cameras at a specific displacement were mounted on a rail car moving at 60Kmph to produce a 3D image. Deformation of railway tracks was identified with 3D DIC and pattern projection algorithms. US Federal Railroad Administration data between the years 2005 and 2015 showed 16000 derailment due to track, ballast and rolling stock failure.

Automation of Rolling stock was important for railways around the world for cutting costs and saving trains from derailment. Computer vision processing is being studied for monitoring train rolling stock to identify defects in bogies or check the bogie part health after maintenance. Hart *et al.* [12] used multispectral imaging to extract bogie segments for inspection. They recorded both RGB videos and thermal videos of a moving train and processed the frames using panoramic representation and correlation. The focus was on detecting high temperature regions on the undercarriage parts such as brake shoes, Axel box, air conditioning blowers and wheel joints. Though, the work is of great use to rail companies, motion blur makes it difficult to identify non-heating parts from heating parts.

Kim and Kim [13], transforms the rolling stock brake inspection problem into an image curve fitting problem. The methods developed use a pit hole setting on the undercarriage of the train to record bogie movements and use the cure fitting tools to identify brake configuration characteristics during train movement. The cost of the setup poses a big drawback for such a system to be implemented in real time. The patent from Sanchez-Revuelta and Gomez [14], shows the use of computer vision for rolling stock examination which was two decades prior to our work. This patent says, artificial vison will be used to monitor rolling stock by mounting cameras on the train. The captured videos with a train mounted camera induces multiple levels of misalignments which increases the computation power for processing.

Kazanskiy and Popov [15], proposed a vision based computer technology for rolling stock monitoring. This work integrates glare free lighting system, video compression models and structured lighting module to detect the presence of train on tracks to monitor rolling stock. The setup was close to making rolling stock automation a reality, despite the fact there is little mention on evaluating each bogie part for inspection. Freid *et al.* [16] provides an undercarriage arrangement with lighting and a camera for rolling stock video capturing. The algorithm uses simple edge detection models for extracting axle rod and measuring its temperature using a thermal camera. The model provides a good insight into the importance of the problem in automating rolling stock examination. Jarzebowicz and Judek [17] and Zhang [18] proposed a 3D reconstruction for monitoring contact strips and rolling stock wheel surfaces. These methods are effective as the 3D models perfectly reconstructs the surface defects in moving parts. These methods use a lot of computation time and energy for processing, as it is difficult to model all defective surfaces.

The literature shows quite a few state-of-the-art computer vision models for remote monitoring a train rolling stock. However, they all give models that does not include the mechanism for inspecting individual bogie parts. The aim of train rolling stock monitoring is to identify a non-performing part or a defective part that can be replaced in time to avoid accidents. Previously proposed models show little impact in this area. Hence, considering this gap in computer vision based rolling stock automation, we propose to use a high-speed video camera to capture moving bogies along with advanced state-of-the-art level set formulations to identify discrete bogie parts and their induced defects. The camera sensor records at 240fps in a wide-angle mode with a resolution of 640×480 pixels covering the entire bogie in one frame as shown in Fig.2. The bogie parts segmentation is being formulated as a convex curve fitting problem with the prior knowledge of the shape, region and boundary of the bogie part.

Active contours and its versions such as level sets were used extensively in segmentation of images and video

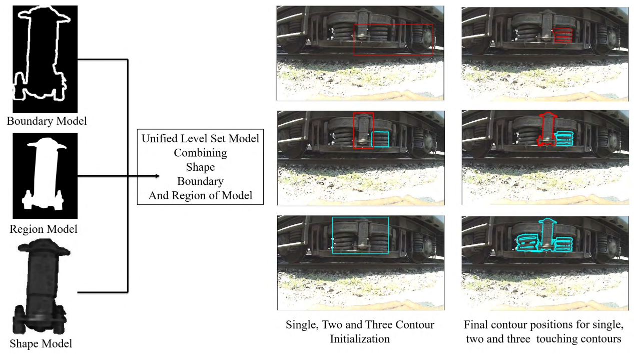


FIGURE 2. **Flow chart showing various processes in the proposed level set model for segmentation of train rolling stock with single and multiple touching contours with multiple and single initial contour.**

objects [19], [20]. The active contours are either region based or boundary based models that mobilize a predefined contour towards the object boundaries. The boundary based approaches are forged by Geodesic as active contours based on gradient vector flow (GVF). These are reliable and fast [21], [22]. However, their performance was limited by their ability to detect gradient edges of the objects in the image. In videos, object movement is being subjected to variations in speed, lighting, camera movements, touching object boundaries and occlusions due to moving objects. The strength of this model is susceptible for detecting and segmenting moving video objects. Most often the touching or overlapping video objects are segmented as a single object [3]. On the other hand, region based models depend on the color or texture attributes of an entire object region to form an energy function for the contour [23]. However, the Chan Vese model [24], uses a level set energy minimization function based on Mumfordshah distance function [25] to force the contour towards the object region. However, the model suffers in the presence of background clutter, touching objects and occlusions. Moreover, both these models are computationally limited on high resolution images or video frames. They are also limited by initialization of contour on high resolution images and videos.

Combining shape, color and texture information of an object with the level set function as prior model has enhanced the model's suitability to segment objects in a wide variety of image applications [26], [27]. Current generation of level sets can segment only one set of objects that are touching or overlapping on to one another accurately [3]. The train bogie as seen in Fig.2, is a densely packed object space with touching and overlapping objects. It is required to detect connected bogie objects, as they must be in their positions during train movement. At the same time, each independent object segmentation determines its structural integrity. From the Indian railway rolling stock operational manual [28], there are around 10 important and crucial things to be checked during rolling examination.

The basis for this work comes from two phase active contours developed by Chan Vese [29], [30] based on difference between mean gray level intensities of the foreground and background. Kernel density estimation [31] and statistical shape model based [32] prior shapes were being used on a set of parametric curve evolutions to represent specific shapes.

This work proposes to segment single video objects and multiple touching objects separately, using the boundary, region, shape based level sets. The level set is evolved by minimizing the unified variational energy function. The novelty of this method is to segment touching and independent rolling stock objects with in a high-speed video frame simultaneously. Unlike our previous works [3] or related works [31], [33], the proposed framework will handle simultaneous segmentation of multiple video objects. Fig.2 gives the flow chart of the overall segmentation process. We show the usefulness of the proposed model in automating the train rolling stock examination process for better, faster condition monitoring to improve rail passenger safety. The rest of the paper is organized as: section 3 describes the proposed Level Sets. Results and discussion showing comparisons with other state-of-the-art models on our train rolling stock video datasets are being presented in section 4. Finally, conclusions are drawn in section 5.

III. UNIFIED LEVEL SET FORMULATION

A video frame $v(x, y, t) \in R^{N\times 3}$ is a set of numbers in RGB color plane representing pixels at locations (x, y) in a frame $'t'$ and $'v'$ giving the intensity values of the pixels in the three planes. Here $'N'$ is the size of the video frame. Each high-speed video frame v is represented as $F_v \in R^2$. The region of interest object in the video frame is modelled as a shape and boundary prior, which is then integrated with the region term to formulate the proposed level set model. In this section, we provide a description for each of these models, which are then unified to form an evolving level set energy functional.

Active Contour models were first introduced by Terzopoulos *et al.* [34] to model shape for image segmentation. Evolution of the equation in an active contour model is labelled as a *snake* and was familiarized by Kass *et al.* [35]. Let $F_{XY}AC \rightarrow R^2$ is observable curve which is controlled by a set of positive real numbers in a space holding 2D shapes. The subspace target is $S : AC \rightarrow R^2$, here $S \subset F$ is a subset of an image. The energy function formulated for an image segmentation problem is defined as:

$$E^S = \int_0^1 \{E^I(V(s)) + E^{F_{XY}}(V(s))\}ds \tag{1}$$

Here the snake energy is E^S. The snake's internal energy is E^I and the image energy is $E^{F_{XY}}$. The snake is positioned on the image frame at:

$$V(s) = (X(s), Y(s)) \tag{2}$$

The internal energy E^I, describes its twisting on the image and the $E^{F_{XY}}$, pushes the snake onto image boundaries in the twisted curve. The E^I is defined as:

$$E^I = \frac{\left(\alpha(s)\left|\omega'(s)\right|^2 + \beta(s)\left|\omega''(s)\right|^2\right)}{2} \tag{3}$$

where $\omega'(s)$ is the first order derivative of $\omega(s)$, which tracks curve length variations and $\alpha(s)$ provides degree of tightening in all directions. Correspondingly, $\omega''(s)$ is the second order derivation of $\omega(s)$ with respect to s and $\beta(s)$ normalizes movement of snake's boundary in the direction of the snake curvature. External image force is modelled as a magnitude of squared gradient function:

$$E^{F_{xy}} = -\left|\nabla F(x, y)\right|^2 \tag{4}$$

A. SHAPE MODEL BASED LEVEL SET FORMULATION

In the first video frame, training shape is aligned at the same position, orientation and scale of the object in the frame. A shape prior model for each of the individual shapes is constructed using region based model. To align the object shape in the video frame, the training shapes are subjected to translation, rotation and scaling. A computationally simple shape training model $'P'$, which defines the Euclidean similarity transformation [36] (EST) as,

$$P = aRP + T \tag{5}$$

Where $'a', 'R'$ and $'T'$ are scaling, rotational and translation factors enabling the shape prior term to align with the object in the frame. These factors can be obtained as

$$a = \frac{\left|A_{ref}\right|}{\left|A_{tar}\right|} \tag{6}$$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{7}$$

$$T = \sqrt{\left(x - x^{ref}\right)^2 + \left(y - y^{ref}\right)^2} \tag{8}$$

where, A_{ref} and A_{tar} are areas of the shape prior and the target shape respectively. $'\theta'$ is the rotational angle between the shape prior and the target shape on its principle axis. $'T'$ is the translation of centroid coordinates (x, y) of each pixel in the target shape and reference shape $\left(x^{ref}, y^{ref}\right)$.

The shape term for the level set function is a zero-level set of the same dimension as the original video frame. A signed distance function (SDF) encodes the shape contour distance to the nearest distance in shape space. Let E^{Shape} defines the energy functional of the shape perimeter information for the active contour. This shape energy term computes the difference between the evolving level set ϕ and the zero-shape prior level set ϕ_S. The prior shape level set is formulated as

$$E^{Shape}(\Theta, \phi_S^{(I)}, \phi_S^{(E)}) = \int_\Theta \left(H\left(\phi\left(x, y\right)\right) - H\left(\phi_S\left(x, y\right)\right)\right)^2 \delta\left(\phi\right) dxdy \tag{9}$$

where, Θ is the initial contour space. Here, H gives the heaviside function and δ is dirac- delta function. To create a unique connection between its adjacent level set ϕ and a pre-defined shape model ϕ_S, it is assumed that $\phi < 0$, *inside* ϕ_S, $\phi > 0$, *outside* ϕ_S and $|\phi| = 1$ everywhere else. There are many ways to define this signed distance function [31], [32], out of which we use the most commonly used with constrains on rotation, scaling and translational properties. In this work we propose to use ϕ as the initial contour and ϕ_S is the shape prior contour to compute level set area difference as in [36]. The above model recovers shape approximated to shape prior object by masking all non-familiar objects. This model recovers only one shape from multiple shapes of similar kind present in the frame.

B. BOUNDARY MODEL LEVEL SET

Boundary is formed by a closed set of points around the object of interest. The object boundary is represented as an edge function $'g'$ calculated on the gradient of the video frame F_x. The energy functional representing the object boundaries in the video frame is formulated as

$$E^{Boundary}(\Theta, \phi) = \int_\Theta \left(g\left(\left|\nabla F_x\right|\right)\left|\nabla\phi\left(x, y\right)\right|\right)\delta\left(\phi\right) dxdy \tag{10}$$

where Θ is the initial contour space. ∇F_x gives the gradient matrix of the video frame. $\delta\left(\phi\right)$ is the contour measured on $\phi = 0$.

C. REGION MODEL LEVEL SET

The Chan Vese Level Set function [24] is the state-of-the-art region based level set model expressed arithmetically on a minimizing energy function defined by

$$\begin{aligned} &E^{region}\left(\Theta, \phi^I, \phi^E\right) \\ &= \omega_1 \int_\Theta ds + \omega_2 \left[\frac{1}{2}\iint_{\Theta^I}\left(F(x, y) - \Phi^I\right)^2 dxdy \right. \\ &\quad \left. + \frac{1}{2}\iint_{\Theta^E}\left(F(x, y) - \Phi^E\right)^2 dxdy\right] \end{aligned} \tag{11}$$

In this above equation, the initial term points to the arc length $\underset{\min(\Theta,\Phi)}{\arg}\left(\omega_1 \times l(\Theta)\right)$, which provides the consistency of Θ throughout the curve evolution and $l(\Theta)$ is contour perimeter. The next term in (11) is a combination of integrals. The first integral force pushes the contour Θ to the image objects and the next integral guarantees the differentiability of the contour Θ. The integral is evaluated on the internal and external contour regions defined by Θ^I *and* Θ^E respectively.

D. UNIFIED LEVEL SET MODEL

In (11) the weights are the positive real numbers ω_1, $\omega_2 \geq 0$. The level set is formulated using Mumford-shah distance function [24] as:

$$\begin{aligned} E^{SBR} &= \alpha E^{shape}(\Theta, \phi_S^{(I)}, \phi_S^{(E)}) + \beta E^{boundary}(\Theta, \phi) \\ &\quad + \gamma E^{region}\left(\Theta, \phi^I, \phi^E\right) \end{aligned} \tag{12}$$

The Level Sets model considers area of the pixels inside and outside the contour given by

$$\Theta = \begin{cases} \phi^I, & where (x, y) lies\ inside \\ & \Theta = \frac{1}{|\Theta^I|}\int\int_{\Theta^I} F_x dx \\ \phi^E, & where\ (x, y) lies\ outside \\ & \Theta = \frac{1}{|\Theta^E|}\int\int_{\Theta^E} F_x dx \end{cases} \tag{13}$$

By using the level set models in [24], we solve the minimization problems in (12) and the unified shape, boundary and region equation of the level set function as

$$\begin{aligned} &E(\Theta, \Phi^{(I)}, \Phi^{(E)}) \\ &= \min_{\Theta, \Phi^{(I)}, \Phi^{(E)}} \alpha \int_{\Theta} (H(\phi(x, y)) - H(\phi_S(x, y)))^2 \delta(\phi)\, dxdy \\ &\quad + \beta \int_{\Theta} (g(|\nabla F_x|)\, |\nabla \phi(x, y)|)\, \delta(\phi)\, dxdy \\ &\quad + \alpha \int_{\Theta^I} (F_x - \phi^I)^2 H(\Theta) \\ &\quad + \int_{\Theta^E} (F_x - \phi^E)^2 (1 - H(\Theta(x, y))) dxdy \\ &\quad + \mu \int_{\Theta} |\nabla H(\Theta(x, y))|\, dxdy \end{aligned} \tag{14}$$

Here heaviside function is $H(\Theta)$. The unified level set in (14) is updated iteratively by minimizing the gradient descent model, defined to arrive at a minimum value for $\Theta(x, y)$

$$\begin{aligned} \Theta^t &= \alpha (H(\phi(x, y)) - H(\phi_S(x, y)))^2 \delta(\phi) \\ &\quad + \beta (g(|\nabla F_x|)\, |\nabla \phi(x, y)|)\, \delta(\phi) \\ &\quad + \gamma \left(((F_x - \phi^I)^2 + (F_x - \phi^E)^2 \right) \delta(\phi) \\ &\quad + \mu \nabla . \frac{\nabla \Theta(x, y)}{|\nabla \Theta(x, y)|} \delta(\phi) \end{aligned} \tag{15}$$

Here the pixel locations in the image are denoted as x and y. The delta function is $\delta(\Theta)$ and iterative adaptations are initiated using the equations of $\Phi^{(I)}$ and $\Phi^{(E)}$ as

$$\phi^I = \frac{\int\int_{\Theta} F_x H(\Theta(x, y)) dxdy}{\int\int_{\Theta} H(\Theta(x, y)) dxdy} \tag{16}$$

$$\phi^E = \frac{\int\int_{\Theta} F_x (1 - H(\Theta(x, y))) dxdy}{\int\int_{\Theta} (1 - H(\Theta(x, y))) dxdy} \tag{17}$$

E. MULTIPLE OBJECT SEGMENTATION WITH TOUCHING BOUNDARIES AND OCCLUSIONS

The level set equation in (15) can segment single object per frame. Its functionality is limited for multi object segmentation if there are inter object touching's, overlapping's and occlusions. To extend the proposed model into a multi object segmentation level set, we apply the method in [37]. Ali and Madabhushi [37] propose to segment multiple objects of same dimension and shape. However, in this work we propose to segment objects of various shapes found on the train bogie having different common boundaries among the shapes. Let the video frame F_x is gives consisting of multiple objects $\{\psi_1, \psi_2, \ldots.., \psi_k\}$ of different shapes that are touching each other. Each object is associated with a level set function $\{\Theta_1, \Theta_2, \ldots.., \Theta_k\}$ which are made to interact among themselves at the boundaries. Unlike the model in [41], we try to model the intersections of the evolving level sets instead of trying to find a characteristic function to model the level set. We model the multiple contours as union $\left\{\Theta_1 \bigcup \Theta_2 \bigcup \ldots.. \bigcup \Theta_k\right\}$and mutually dependent events $\left\{\Theta_1 \bigcap \Theta_2 \bigcap \ldots.. \bigcap \Theta_k\right\} = M \in R^+$. The expression for simultaneous segmentation of 2 objects with touching boundaries with respect to the given shape and boundary prior is solved by minimizing the energy functional

$$\begin{aligned} \bigcup_{k=1}^{2} \Theta_k^t &= \sum_{k=1}^{2} \alpha \left(H(\phi_k(x, y)) - H\left(\phi_{S_k}(x, y)\right)\right)^2 \delta(\phi) \\ &\quad + \beta \sum_{k=1}^{2} \left(g\left(\left|\nabla F_x^k\right|\right) |\nabla \phi_k(x, y)| \right) \delta(\phi) \\ &\quad + \gamma \sum_{k=1}^{2} \left(((F_x^k - \phi_k^I)^2 + (F_x^k - \phi_k^E)^2 \right) \delta(\phi) \\ &\quad + \mu \bigcup_{k=1}^{2} \nabla . \frac{\nabla \Theta(x, y)}{|\nabla \Theta(x, y)|} \delta(\phi) \\ &\quad + \bigcap_{k=1}^{2} H(\phi_k(x, y)) \end{aligned} \tag{18}$$

Where $'\bigcup'$ and $'\bigcap'$ denote the union and intersection of contours respectively. The last term will find the intersecting or touching boundaries and merge the intersecting region into the individual contours belonging to the object. This model can be extended to simultaneous segmentation of multiple objects with different shapes, provided shape and boundaries of the objects as prior information. For multiple objects the expression in (18) can be modified by increasing the value of 'k'.

IV. EXPERIMENTATION AND PERFORMANCE ESTIMATION

A. EXPERIMENTAL SETUP AND DATASETS

The level set build is tested for segmenting train rolling stock moving at 30Kmph in real time conditions. To capture entire bogie of the train, we employed a high speed 240 frames per second (fps) wide angle 52° sports action camera. The camera module is installed near a railway station in India and 4 videos were recorded at different times of the day. This is done to test the effect of ambient lighting on the proposed segmentation model. Each train video is used in designing an experiment. In the final experiment, a defect was introduced deliberately using photo editing software and the defective frames were reinserted into the video sequence. The goal of 5^{th} experiment is to test the ability of the proposed segmentation model to segment the defective part without using a defective shape prior. This procedure is used to test

TABLE 1. Video Datasets considered in for testing the proposed model.

	Name	Number of Frames
Dataset-1	Train video capture at 6.30AM	80×36=2880
Dataset-2	Train video capture at 12.30PM	80×40=3200
Dataset-3	Train video capture at 4.30PM	80×32=2560
Dataset-4	Train video capture at 7.00PM	80×26=2080
Dataset-5	Defective part at 12.30PM Train	40×40×4=2400

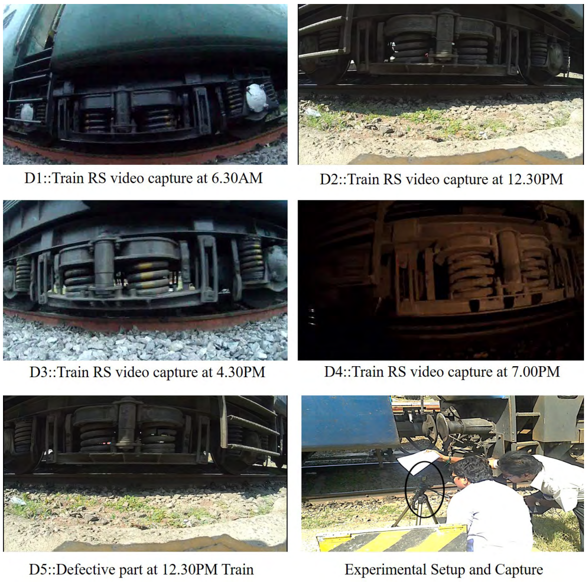


FIGURE 3. Capture process and Datasets used for testing the proposed multi object segmentation model.

the robustness of the proposed algorithm in identifying or segmenting defective pieces using unifying shape and boundary priors of a healthy part. Table 1. shows the five experimental evaluations formulated on the following databases.

B. PERFORMANCE ESTIMATION PARAMETERS

We qualitatively and quantitatively compared our proposed model against previously proposed models in [3], [4], [32], and [33]. All these models were shape prior level set models. However, [3] and [4] consider rolling stock examination on the same dataset, whereas [32] and [33] are the state-of-the-art two phase and multi-phase level set models with shape prior terms in their energy function. We solved all the models using Mumfordshah distance minimization framework used by Chan Vese model [24]. Experiments naming 1 to 5 are formed using the datasets 1 to 5 on the level set models proposed in this work and the works in [3], [4], [32], and [33].

C. BOGIE PART DETECTION: DETECTION SENSITIVITY (DS)

Detection sensitivity (DS) evaluates the strength of the proposed level set model to correctly detect the part from the cluttered bogie parts. This measure is computed from the true positive (TP) and false negative (FN). TP gives the number of times a bogie part is detected correctly provided the equivalent shape and boundary prior terms. FN is the number of times in all frames the bogie part is un – detected or misclassified as other part. DS is an important measure as we are using one shape and boundary prior model on all trains having ICF bogies in all ambient conditions. DS is given as

$$DS = \frac{TP}{TP + FN} \tag{19}$$

The DS value lies between 0 and 1.

D. TOUCHING BOUNDARY SENSITIVITY (TBS)

TBS quantity is calculated for measuring the model's ability to segment multiple objects of interest having touching boundaries or inter object occlusions. TBS is formulated as

$$TBS = \frac{NTBD}{TTB} \tag{20}$$

NTBD is the number of touching boundaries detected correctly and TTB is the total number of touching boundaries. This measure characterizes the model's capability to correctly resolve touching object boundaries.

E. MEAN ABSOLUTE DISTANCE (MAD)

MAD measures the error between the segmented video frame and ground truth shape prior model with

$$MAD = \left| \frac{\left|S^O - GT\right| - \sum_{i=1}^{n} \left|S^O - GT\right|}{n} \right| \tag{21}$$

Where, S^O is the segmented output and GT is the ground truth obtained from an expert train engineer. Here $'n'$ denotes the number of pixels in the segmented frame. This value measures the closeness of the segmented part with the absolute real part.

F. NUMBER OF ITERATIONS – MODEL SPEED (MS)

As condition monitoring happens in real time, it is advisable to test the level set model for its speed in producing the output. The MS is a measure giving the number of iterations or updates of ϕ_0, the initial contour, during the program execution. This measure depends on CPU timing and memory constraints. However, the entire simulations were carried out on an 8GB RAM with GPU support without reducing the resolution of the frames to maintain quality.

V. RESULTS AND ANALYSIS

Each 20-coach train consists of 40 bogies per train. However, the dataset is made of different trains with varying number of coaches. Hence, each train rolling capture produces a different number of bogie frames. For analysis, the bogie frames are separated from non-bogie frames by a simple density matching algorithm which extract bogie frames based on the pixel density in the image. For each bogie, the minimum number of frames obtained are 80. Total number of frames evaluated per train is given by, number of frames/bogie × number of bogies per train. Hence, each experiment used around 2000 bogie frames.

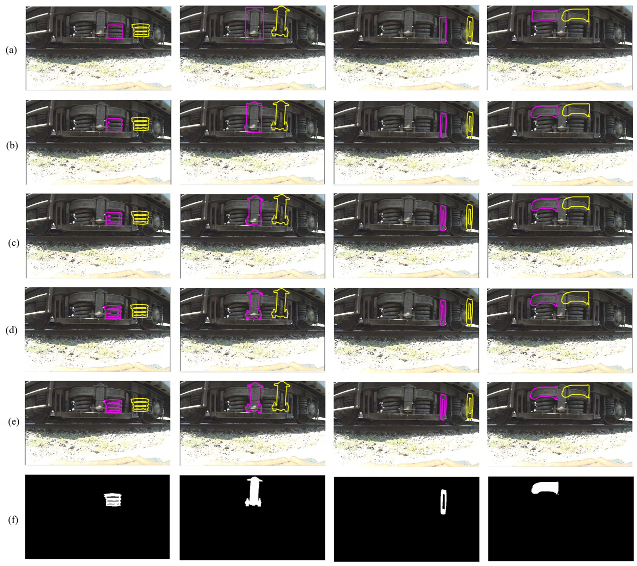


FIGURE 4. (a) Initial contour, (b) Contour Evolution at 10 iterations, (c) 20 iterations, (d) 30 iterations, (e) 40 iterations and (f) the segmented bogie parts using the proposed Shape, Boundary and Region Unified Level sets.

For datasets 1 to 3 the contour model is initialized with $\alpha = 0.18$,$\beta = 0.35$,$\gamma = 0.47$ and $\mu = 0.2$. For dataset-4, $\alpha = 0.38$, $\beta = 0.35$,$\gamma = 0.27$ and $\mu = 0.2$. This is due to less ambient lighting which is affecting the contour propagation. Hence, more of the shape and boundary priors were used compared to the region parameters.

A. QUALITATIVE ANALYSIS

The results for the dataset-2 on 4 different parts is shown in Fig.4. The models superiority can be judged using the results of previous shape prior models reported in [3] and [4] shown in Fig.5.1 and Fig.5.2 respectively. The results displayed in Fig.4, Fig.5.1 and Fig.5.2 aim to substantiate the strength of the proposed model in terms of detection and segmentation of a single bogie part. All the frames (910×512 resolution) used at different time stamps produced similar results with the proposed model. From the figures, the proposed model segments bogie parts exceptionally close to their ground truth models, where the boundaries have a close resemblance with actual object boundaries. Here the shape restriction avoids local edge minimums and the boundary constraints limits the contour from over segmentation. Fig.6 shows the 2 touching objects of interest in a train bogie frame. We demonstrate the simulation of (18) and its ability to segment multiple touching bogie parts in Fig.6.

The number of iterations for the proposed Multi bogie part segmentation model is 40% lower compared to other models in literature and our proposed single object case. CPU time per iteration on 8GB RAM is around 0.78s for the multi bogie part segmentation and 1.6s for single part segmentation. For 10 bogie parts in the frame, the model is quite robust to iterations producing highly qualitative segments. Fig.7 shows the contour evolution along with the final 10 object segments extracted from the proposed multi object segmentation model. For clarity, each local contour is shown in different color. All these contours are united to form the initial zero level set. Each segment can be extracted from thresholding the individual evolved level sets.

TABLE 2. Average Dataset (DS) for each dataset with the proposed model(S) and the state of the art level set models in [4], [3], [32], and [33].

	DS-1	DS-2	DS-3	DS-4	DS-5
SP_LS [4]	0.532	0.587	0.535	0.421	0.425
ISP_LS [3]	0.681	0.713	0.679	0.645	0.625
MPDM [33]	0.861	0.887	0.852	0.803	0.825
JOSP_LS [32]	0.798	0.825	0.804	0.769	0.775
Proposed Model(S)	0.904	0.937	0.911	0.887	0.901

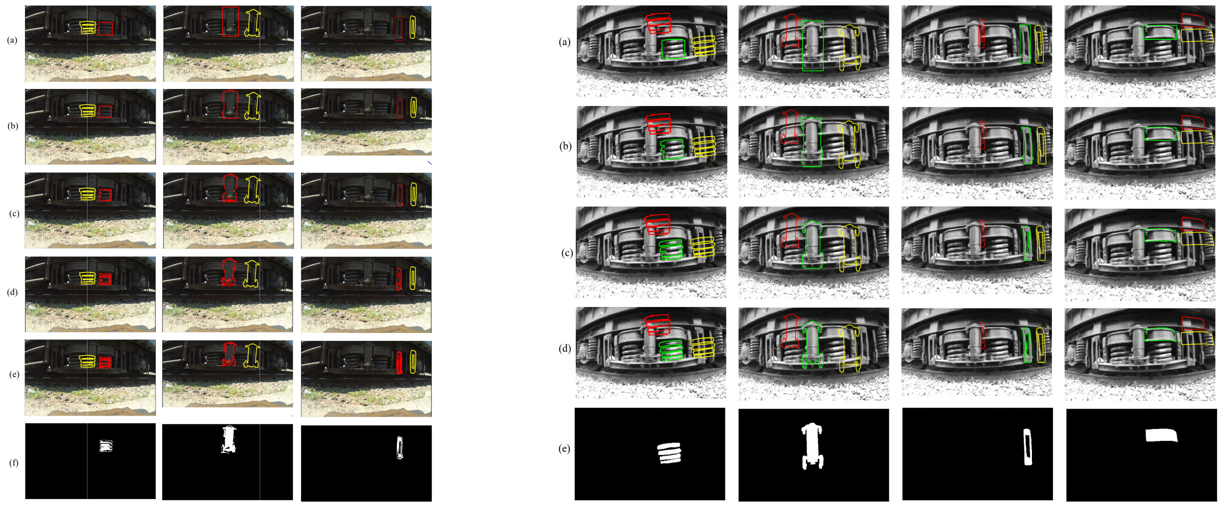
TABLE 3. Average TBS for each dataset (DS) with the proposed model and the state of the art level set models in [3], [4], [32], and [33].

	DS-1	DS-2	DS-3	DS-4	DS-5
SP_LS [4]	0.05	0.05	0.05	0.04	0.04
ISP_LS [3]	0.06	0.07	0.06	0.06	0.02
MPDM [33]	0.1	0.1	0.1	0.08	0.05
JOSP_LS [32]	0.1	0.1	0.1	0.07	0.06
Proposed Model(M)	0.9	0.9	0.9	0.8	0.7

The structural condition of each extracted bogie object can be tested against the GT models for bogie part health detection. Each extracted level set becomes a shape and boundary prior model to the next corresponding frame. This model is advantageous as only one set of parts per train are required during evaluation of the proposed model on the entire train. Fig.8 shows the proposed model on a defective right spring suspension. The broken spring is modelled with a crack created in photoshop as we could not find one during data acquisition. Single contour model is used to check its robustness in segmenting defective spring bogie part with the healthy shape and boundary priors. Multiple object segmentation model with different shapes is limited by its ability to segment parts whose shape prior models are available. However, for defective part segmentation with non-defective shape priors is an over segmentation problem. In the multiple case, each contour is regularized by using $'\alpha'$ from (6) to stop the segmentation process. The number of iterations to attract contour towards the finer cracks has augmented, based on the dimension of the cracks. In Fig.8, each defective part has taken around 80 iterations to reach the critical segmentation output.

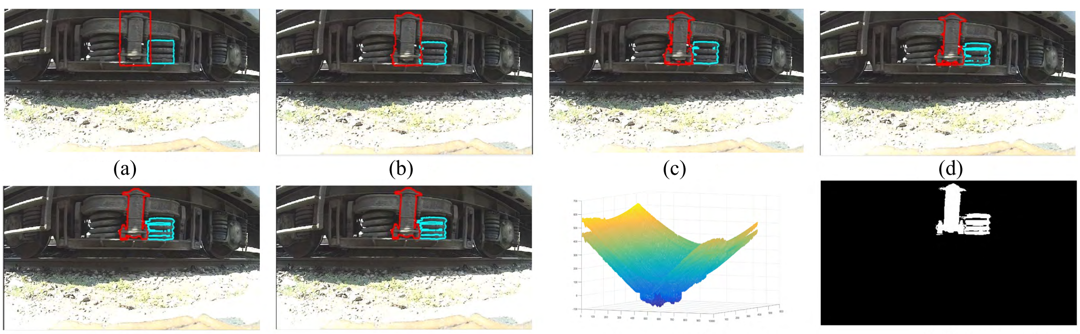
B. QUANTITATIVE ANALYSIS

Quantitative results present a performance evaluation of the proposed level set model against the previous models in [3], [4], [32], and [33]. These results show an improved performance of the proposed region, shape and boundary level set model against the shape prior models in [3], [4], [32], and [33]. Table 2 gives the average detection sensitivity (ADS) of the segmentation algorithms on a set of 2000 frames of one bogie. The values listed in Table 2 shows the accuracy of the proposed model against other models for single contour case. Proposed model(S) is a single object segmentation model.



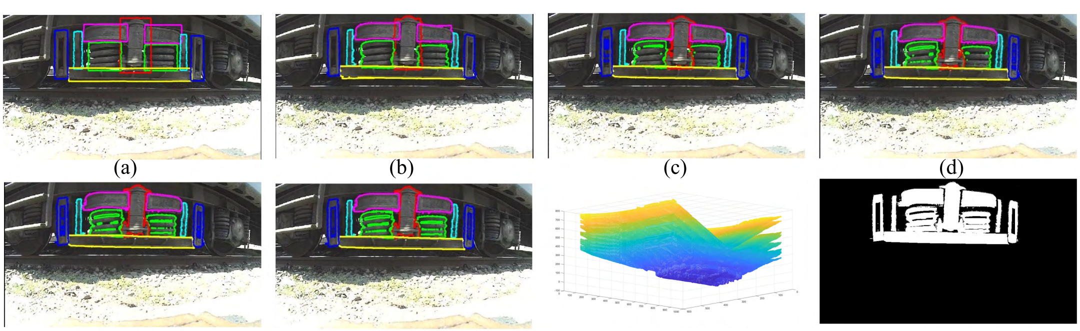
(a) (b)

FIGURE 5. (a) 5.1: Shape prior CV level set model (a) at 3 itr,(b) 20 itr, (c) 40 itr, (d) 60 itr, (e) 80itr and (f) their segmented outputs. (b)5.2: Shape invariance level sets using translation, rotation and scale models, Red (original shape prior), yellow (Transformed shape prior) and Green (Evolving level set), (a) at 5 iterations, (b) 15 iterations, (c) 40 iterations, (d) 70 iterations and (e) segmented bogie parts. (a) 5.1. (b) 5.2.



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 6. Two touching contour evolution with 3D final level set and its object segments.



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 7. Two touching contour evolution with 3D final level set and its object segments.

Table 3, gives the touching boundary sensitivity (TBS) of the level set functional in segmenting multiple objects at a time. The results show robustness of the proposed multi object segmentation level set with multiple shape areas with touching and occluding boundaries over the previous models. Proposed model(M) is the segmentation of multiple objects of interest.

Mean Absolute Distance (MAD) values for 2000 dataset sample video frames for each of the segmentation models is computed. The MAD values of our model are less than

TABLE 4. Average MAD computed with the proposed model and the state of the art level set models in [3], [4], [33] and [32].

	Dataset-1	Dataset-2	Dataset-3	Dataset-4	Dataset-5
SP_LS [4]	$9.98X10^{-2}$	$7.08X10^{-2}$	$9.84X10^{-2}$	$7.01X10^{-1}$	$4.92X10^{-1}$
ISP_LS [3]	$5.82X10^{-2}$	$1.37X10^{-2}$	$4.96X10^{-2}$	$9.88X10^{-2}$	$9.16X10^{-1}$
MPDM [33]	$9.61X10^{-3}$	$3.57X10^{-3}$	$9.93X10^{-3}$	$8.17X10^{-2}$	$6.37X10^{-2}$
JOSP_LS [32]	$9.97X10^{-3}$	$5.11X10^{-3}$	$9.72X10^{-3}$	$7.82X10^{-2}$	$5.57X10^{-2}$
Proposed Model(M)	$2.24X10^{-3}$	$4.18X10^{-4}$	$1.96X10^{-3}$	$9.18X10^{-3}$	$5.61X10^{-3}$

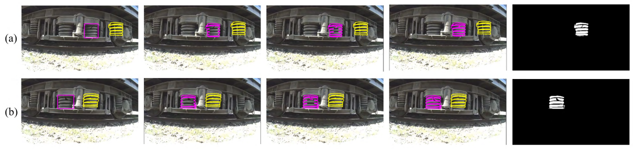


FIGURE 8. Spring defective crack detection and segmentation with the proposed model given non-defective shape and boundary priors.

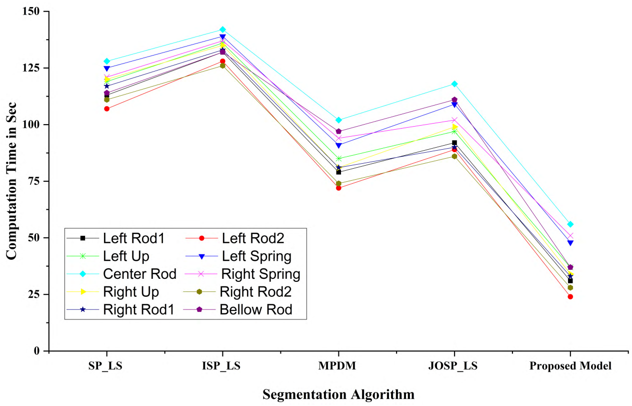


FIGURE 9. CPU times in sec for each object segmentation over 80 frames.

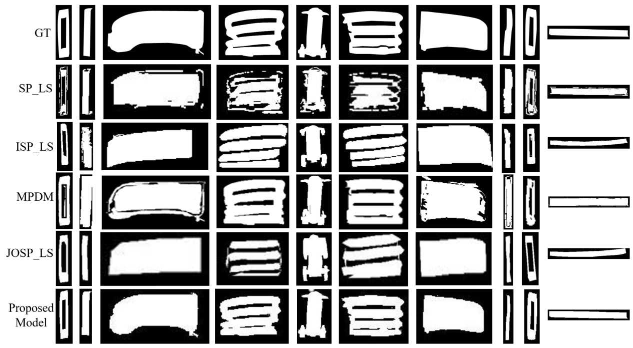


FIGURE 10. Segmented bogie objects from Ground Truth (GT), SP_LS [4], ISP_LS [3], MPDM [33], JOSP_LS [32] and the proposed model level set.

10 pixels in difference for all the samples in the dataset. All the MAD values per dataset per method are averaged over the entire length of the train and are reflected in Table 4. The proposed model (M) gives segmentation of bogie parts closer to GT model compared to other similar class models.

Since the described application in this work is time sensitive, we intend to test the proposed model's speed with the CPU time consumed for segmenting the objects of interest. This is calculated based on the number of iterations per object segmentation model. Compared to single object segmentation models, multiple object segmentations registered 40% less number of iterations to reach the desired segmentation. Fig.9 shows the plot of averaged segmentation times for 2000 frames per bogie part, i.e. 40 bogies of a train with the tested algorithms. The proposed multiple object level set model has worked faster in segmenting the desired bogie parts compared to the other 4 models.

A final comparison of extracted bogie parts in [3], [4], [32], and [33] with the proposed segmentation models is presented in Fig.10. The results reflect the improvement in robustness of the proposed model against the remaining models. The proposed region, boundary and shape based multi object level sets have shown in significant improvement in all performace parameters.

VI. CONCLUSION

This work presents a novel segmentation model based on region, boundary and shape prior level sets with application to real time moving object detection and segmentation. We further improve the model to detect and segment touching and occluding object boundaries having different shapes at a time. The over segmentation in this case is handled effectively by stopping the evolving level set using the area ration between the object boundaries. The results reflect that our proposed model is accurate compared to traditional state of the art shape prior models. The ability of our algorithm in detecting and segmenting multiple bogie parts accurately may be useful in developing automated conditioning monitoring models for train rolling stock examination. The algorithm can also be used to define the relationship between bogie parts and their binding force.

REFERENCES

[1] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

[2] Y. Li, Y. Li, H. Kim, and S. Serikawa, "Active contour model-based segmentation algorithm for medical robots recognition," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10485–10500, 2018.

[3] P. V. V. Kishore and C. R. Prasad, "Computer vision based train rolling stock examination," *Optik*, vol. 132, pp. 427–444, Mar. 2017.

[4] P. V. V. Kishore and C. R. Prasad, "Shape prior active contours for computerized vision based train rolling stock parts segmentation," *Int. Rev. Comput. Softw.*, vol. 10, no. 12, pp. 1233–1243, Dec. 2015, doi: 10.15866%2Firecos.v10i12.8110.

[5] D. Kosmopoulos and T. Varvarigou, "Automated inspection of gaps on the automobile production line through stereo vision and specular reflection," *Comput. Ind.*, vol. 46, no. 1, pp. 49–63, 2001.
[6] V. Milanés *et al.*, "Intelligent automatic overtaking system using vision for vehicle detection," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3362–3373, 2012.
[7] H. Fathi, F. Dai, and M. Lourakis, "Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges," *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 149–161, 2015.
[8] H. Zhang and D. Li, "Applications of computer vision techniques to cotton foreign matter inspection: A review," *Comput. Electron. Agricult.*, vol. 109, pp. 59–70, Nov. 2014.
[9] Y. Yang, Z.-J. Zha, M. Gao, and Z. He, "A robust vision inspection system for detecting surface defects of film capacitors," *Signal Process.*, vol. 124, pp. 54–62, Jul. 2016.
[10] S. Narayanaswami, "Urban transportation: Innovations in infrastructure planning and development," *Int. J. Logistics Manage.*, vol. 28, no. 1, pp. 150–171, 2017.
[11] A. Sabato and C. Niezrecki, "Feasibility of digital image correlation for railroad tie inspection and ballast support assessment," *Measurement*, vol. 103, pp. 93–105, Jun. 2017.
[12] J. M. Hart, E. Resendiz, B. Freid, S. Sawadisavi, C. P. L. Barkan, and N. Ahuja, "Machine vision using multi-spectral imaging for undercarriage inspection of railroad equipment," in *Proc. 8th World Congr. Railway Res.*, Seoul, South Korea, 2008, pp. 1–8.
[13] H. Kim and W.-Y. Kim, "Automated inspection system for rolling stock brake shoes," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 8, pp. 2835–2847, Aug. 2011.
[14] A. L. Sanchez-Revuelta and C.-J. G. Gomez, "Installation and process for measuring rolling parameters by means of artificial vision on wheels of railway vehicles," U.S. Patent 5 808 906 A, Sep. 15, 1998.
[15] N. L. Kazanskiy and S. B. Popov, "Integrated design technology for computer vision systems in railway transportation," *Pattern Recognit. Image Anal.*, vol. 25, no. 2, pp. 215–219, 2015.
[16] B. Freid, C. P. Barkan, N. Ahuja, J. M. Hart, S. Todorvic, and N. Kocher, "Multispectral machine vision for improved undercarriage inspection of railroad rolling stock," in *Proc. 9th Int. Heavy Haul Conf. Spec. Tech. Session-High Tech Heavy Haul*, Kiruna, Sweden, 2007, pp. 11–13.
[17] L. Jarzebowicz and S. Judek, "3D machine vision system for inspection of contact strips in railway vehicle current collectors," in *Proc. Int. Conf. Appl. Electron. (AE)*, Sep. 2014, pp. 139–144.
[18] Y. Zhang, J.-Y. Hu, J.-L. Li, J.-L. Wu, and H.-Q. Wang, "The application of WTP in 3-D reconstruction of train wheel surface and tread defect," *Optik*, vol. 131, pp. 749–753, Feb. 2017.
[19] K. Zhang, L. Zhang, H. Song, and W. Zhou, "Active contours with selective local or global segmentation: A new formulation and level set method," *Image Vis. Comput.*, vol. 28, no. 4, pp. 668–676, 2010.
[20] K. Zhang, L. Zhang, K.-M. Lam, and D. Zhang, "A level set approach to image segmentation with intensity inhomogeneity," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 546–557, Feb. 2016.
[21] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.
[22] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 3, pp. 266–280, Mar. 2000.
[23] R. Ronfard, "Region-based strategies for active contour models," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 229–251, 1994.
[24] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[25] X. Huang, H. Bai, and S. Li, "Automatic aerial image segmentation using a modified chan-vese algorithm," in *Proc. IEEE 9th Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2014, pp. 1091–1094.
[26] P. V. V. Kishore and C. R. Prasad, "Train rolling stock segmentation with morphological differential gradient active contours," in *Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Aug. 2015, pp. 1174–1178.
[27] G. Charpiat, O. Faugeras, R. Keriven, and P. Maurel, "Approximations of shape metrics and application to shape warping and empirical shape statistics," in *Statistics and Analysis of Shapes*, H. Krim, and A. Yezzi, Jr., Eds. Basel, Switzerland: Birkhäuser, 2006, pp. 363–395.
[28] A. Amitabh, "Rail accidents due to human errors-indian railways experience," in *Proc. Int. Railway Saf. Conf.*, Cape Town, South Africa, 2005, pp. 1–14.
[29] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 271–293, Dec. 2002.
[30] M. Sussman, P. Smereka, and S. Osher, "A level set approach for computing solutions to incompressible two-phase flow," *J. Comput. Phys.*, vol. 114, no. 1, pp. 146–159, 1994.
[31] D. Cremers, S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *Int. J. Comput. Vis.*, vol. 69, no. 3, pp. 335–351, 2006.
[32] A. Saito, S. Nawano, and A. Shimizu, "Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs," *Med. Image Anal.*, vol. 28, pp. 46–65, Feb. 2016.
[33] D. Cremers, N. Sochen, and C. Schnörr, "A multiphase dynamic labeling model for variational recognition-driven image segmentation," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 67–81, 2006.
[34] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, "Elastically deformable models," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 205–214, Jul. 1987.
[35] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
[36] M. Werman and D. Weinshall, "Similarity and affine invariant distances between 2D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 810–814, Aug. 1995.
[37] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1448–1460, Jul. 2012.

Authors' photographs and biographies not available at the time of publication.