VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MORPHOLOGICAL DIFFERENTIAL GRADIENT ACTIVE CONTOURS FOR ROLLING STOCK SEGMENTATION IN TRAIN BOGIES

Ch. Raghava Prasad and P. V. V. Kishore
Department of Electronics and Communication Engineering, K L University, Andhra Pradesh, India
E-Mail: chrp@kluniversity.in

ABSTRACT

This paper focuses on Chan vese active contour (CV) model for segmenting the rolling stock. We present a modified version of Chan vese using morphological differential gradient (CVMDG) to segment rolling stock. The rolling stock videos are captured under four different lighting conditions near Guntur railway station in INDIA. Rolling examination as it is called by railway maintenance staff of Indian railways is visual and auditory examination of moving bogies of a train for defects. The undercarriage moving parts of the train are called rolling stock. This paper makes an attempt to segment the rolling stock from video frames of the rolling stock for further analysis. For better segmentation of rolling stock, video frames are contrast enhanced with virtual exposure wavelet image fusion. The segmented rolling stock is compared with ground truth model to assess the usability of the proposed method for rolling stock segmentation.

Keywords: rolling stock examination, active contours, video segmentation, wavelet transform, morphological differential gradient.

1. INTRODUCTION

This paper is the starting point for applying image processing applications to rolling stock video segmentation. Rolling stock examination as popularly known among the people of Indian railways is visual and auditory examination of moving train near railway stations for identifying defects. The rolling defect identification is possible only during the movement of the train.

The under part of the train is called bogie. Bogie of a train is displayed in Figure-1. It consists of wheels, suspension, axle box, break panels, hanging parts and springs. The idea behind rolling stock examination is to identify how these parts are behaving under movement of train. A total of six personal are required to examine the defects near every railway station. They use their trained eyes and ears to extract the defects while the train is moving at 30km/h.

Some of the visual defects are falling objects; parts drag and pull; hanging suspension; missing parts; break functioning and axle box condition. Auditory defects include sounds produced during movement that can identify mostly flat tyres. Any unwanted object attached to the wheel base is termed flat tyre in trains.

Figure-1. Train Rolling Stock components of an Indian Train moving at 30kmph at Guntur railway station.

The segmentation [1, 2] of rolling stock video falls under the category of natural image segmentation. Since all parts are so closely bound it becomes very difficult for the thresholding algorithms [3], edge detection methods [4] and gradient operators [5] to segment the rolling frames. Hence active contour models are tested for segmentation. Chan vase [6] active contours algorithm is used initially.

Active contours are measurable curves that extract image object boundaries. Active contours come under a category of model based segmentation methods. Active contours formulate [7] the movement of a predefined contour within the image domain. The boundaries and regions of image objects defines image domain. Energy function controls the contour movement in the image domain. Terzopoulos introduced the first active contours model [8]. Earlier models of active contours are prone to topological disturbances and are extremely susceptible to initial conditions. However with the growth of level sets, addressing topological changes in the objects in the image has become manageable. Nevertheless all active contours depend on the gradient of the image for ending the growth of the curve.

Due to large number of frames and the complexity of the rolling frame the regular CV model [9] is rendered unsuitable for segmentation. Hence a modified CV model based on morphological differential gradient CV-MDGS based segmentation is proposed in this work. Due to huge brightness variations during image capture uniform contrast enhancement is performed using virtual exposure wavelet image fusion. Unlike histogram equalization, this method does not induce white patches on frames and preserves color and texture. Section 2 gives virtual exposure wavelet image fusion on video frames. CV-MDGS procedure is explained in section 3. Section 4 discusses results. Section 5 concludes the paper with references.

2. CONTRAST ENHANCEMENT WITH VIRTUAL IMAGE FUSION

Contrast in videos depends on shutter speed and F-stop of the digital SLR camera. Shutter speed controls the amount of reflected light reaching the image sensor in the camera. F-Stop controls the hole aperture that passes light to the sensor. Generally the F-Stop approximates to geometric sequence corresponding to powers of root 2 [11]. This concept is used to generate virtual video frames from non-uniform contrast frames.

Let $I^{xy}: D \to \Re^2$ be the low contrast RGB (red, Green and Blue) video image frames of bogies. Converting the RGB image I^{rgb} into YCBr image space to extract the luminance image $Y(x, y)$. (x, y) gives the location of pixels in the image $Y(x, y)$. For this application x and y have a dynamic range of $[0,511]$ pixels each. For every pixel in the low contrast luminance image $Y(x, y)$, N^{th} order virtual image is generated using the expression

$$Y^N(x, y) = \begin{cases} Y(x, y) \times \left(\sqrt{2}\right)^{-N}, Y(x, y) \times \left(\sqrt{2}\right)^{-N} < 511 \\ \max(Y(x, y)) \quad , \text{otherwise} \end{cases} \quad (1)$$

To generate brighter images N is positive and takes values $N = 0,1,2...$ and for darker side N is negative $N = -1,-2,-3...$. In this research the range of N is restricted to [-3 – 0 – +3]. Figure 3 shows the all the 7 images for a differently contrast frame of fig.2 (b). The set of 7 virtual exposed images $Y^{(-3)...0...(+3)}$ are transformed into wavelet domain [11]. The basic filter bank approach used in transforming the 2D images form spatial domain to wavelet domain is shown in Figure-3.

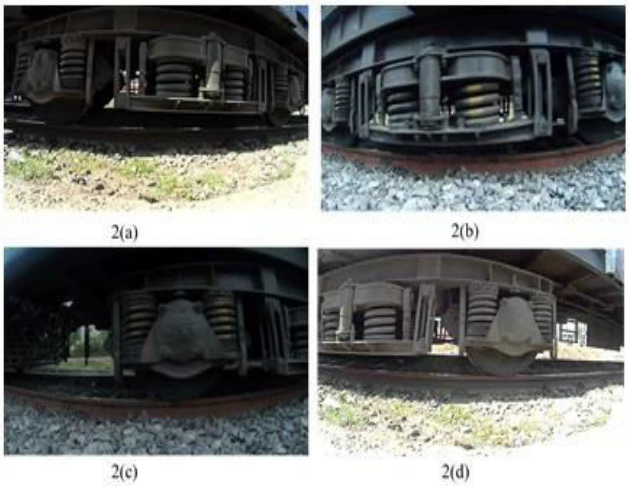


Figure-2. Light intensity variations at different times of the day.(a) at 11.30AM,(b)4.30 PM, (c) 5.45 PM, (d) 2.00 PM.

The DWT approximate coefficients [12] for a 2D signal I(x, y) is formulated as

$$A^L = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I(x, y) \psi^L{}_{ab}(x, y) \quad (2)$$

And the detailed coefficients are formulated as

$$D^L = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I(x, y) \psi^L{}_{a,b}(x, y) \quad (3)$$

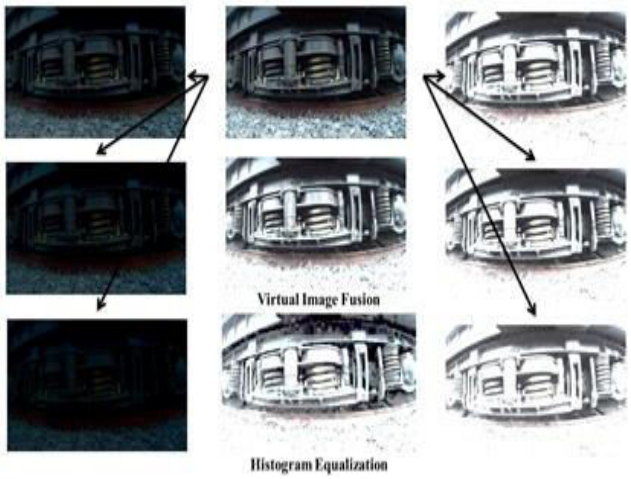


Figure-3. Virtual images varying in the order of N[-3,3]. On the left, from top to bottom,N=-1 to -3, middle,N=0 and on the right,from top to bottom,N=1 to 3, In the middle 2nd column middle image frame is due to virtual image fusion and last is from histogram equalization[10].

Wavelet decomposition level L can be iteratively used to purify the image into various frequency planes. The Haar wavelet is defined by: ψ(x) = 1 if x ∈ [0, 0.5]; ψ(x) = −1 if x ∈ [0.5, 1]; and 0 if not. The associated scaling function is the function: φ(x) = 1 if x ∈ [0, 1]; and 0 if not. This family contains the Haar wavelet, db1, which is discontinuous, resembling a square form.

Image fusion processes bring together two different sets of virtual images by extracting information that is distinctive to a particular image, there by producing a good contrast image. In wavelet based fusion two images having unique properties are transformed using time frequency scaling of wavelet transform individually. Each image transformation produces four coefficients at assumed level 1, known as approximate coefficients, horizontal coefficients, vertical coefficients and diagonal coefficients.

Different fusion rules are applied on these transform coefficients such as max-min, max-max etc. For example in max-min rule, maximum of approximate coefficients and minimum of remaining detailed coefficients are preserved and 2D transformation model is created. Finally by applying 2D inverse transformation in wavelet domain fabricates into an enhanced fused image.

This experimentation uses first level (L=1) debauches (db-2) wavelet decomposition with mean-mean fusion rule. Mean-Mean fusion rule evaluates the average value at each position of approximate and detailed

coefficients for N=-1 image and N=+1 image respectively. Fusion of all sets of virtual images is done to produce a one single image in wavelet domain. Mean-Mean fusion rule is computed using the expression

$$W_{\psi}^{F} = \begin{cases} W_M{}^{\psi^A} + W_S{}^{\psi^A}/2 \\ W_M{}^{\psi^H} + W_S{}^{\psi^H}/2 \\ W_M{}^{\psi^V} + W_S{}^{\psi^V}/2 \\ W_M{}^{\psi^D} + W_S{}^{\psi^D}/2 \end{cases} \quad (4)$$

Inverse wavelet transform brings the image from wavelet domain to spatial domain. The spatial domain image is a good contrast image. The two step algorithm is tested against three sets of frames at different light intensities. The virtual frame fusion is compared with histogram equalized [10] HE image in Figure-3.

3. CV-MDGS

Chan-Vese (CV) [6] active contour model discovers a closed contour $O: D \to \Re^2$ defined on image space D consisting of positive real numbers. The discovered contour optimally estimates the object edges in a gray scale image $I^{xy}: D \to \Re^2$. A single real gray value is selected $\Phi^{(I)}(x, y, t) > 0$, if the gray pixel is within the contour Θ and another single gray level value selected $\Phi^{(E)}(x, y, t) < 0$ for values outside the contour Θ. 't' is the step size. $\Phi^{(E)}$ *or* $\Phi^{(I)}\{(x, y, t)\} = 0$ for gray values on the contour O.The basic idea of CV Active model is to find an optimal contour that fits the object boundaries. Alongside the best contour, the solution should also find a pair of optimal gray scale values $\Phi^F = \left(\Phi^{(I)}, \Phi^{(E)}\right)$ that discriminates object pixels from background pixels.

Mathematically the Chan-Vese active contour is formulated as an energy minimization problem

$$E^{cv}\left(O^F, \Phi^F\right) = \min_{\Phi} E^{cv}(O, \Phi) \quad (5)$$

where, O^F is the final contour shape to be discovered with O as its chosen initial contour. Piecewise linear Mumford-Shah [6] minimizes the energy function or force function CV active contour model E^{cv}. This Energy minimization estimates the pixel values of a gray scale image I^{xy} by a linear piecewise smooth contour O.

The minimization problem is solved using the level set model [6] and this formulates in terms of level set function Θ^{xy} as

$$E^{cv}(\Theta, \Phi^{(I)}, \Phi^{(E)}) = \min_{\Theta, \Phi^{(I)}, \Phi^{(E)}} \chi_2 \left[\iint_{\text{int}(\Theta)} (I^{xy} - \Phi^{(I)})^2 \hbar\left(\Theta^{xy}\right) + \iint_{ext(\Theta)} (I^{xy} - \Phi^{(E)})^2 (1 - \hbar(\Theta^{xy}))dxdy \right] + \chi_1 \int_{\Theta} \left|\nabla \hbar(\Theta^{xy})\right| dxdy \quad (6)$$

where, $\hbar(\Theta)$ is Heaviside function. χ_1 and χ_2 are constants. Euler-Lagrange [6] equations solve minimization problems which uses gradient descent methods in updating level set function $O(x, y, t)$. This is formulates to the following function

$$\Theta^t = -\delta(\Theta)((I^{xy} - \Phi^{(I)})^2 - (I^{xy} - \Phi^{(E)})^2 - \chi_1 \nabla . \frac{\nabla \Theta^{xy}}{\left|\nabla \Theta^{xy}\right|} \quad (7)$$

where x and y stand for the locations of pixels in the image. $\delta(\Theta)$ is the delta function and $\Phi^{(I)}$ and $\Phi^{(E)}$ are iteratively updated using the equations

$$\Phi^{(I)} = \frac{\iint_{\Theta} I^{xy} \hbar(\Theta^{xy}) dxdy}{\iint_{\Theta} H(\Theta^{xy}) dxdy} \quad (8)$$

$$\Phi^{(E)} = \frac{\iint_{\Theta} I^{xy} (1 - H(\Theta^{xy})) dxdy}{\iint_{\Theta} (1 - H(\Theta^{xy})) dxdy} \quad (9)$$

Generally, active contours models use image gradient to identify object boundaries. The CV method performs region based edge detection for frames from most imaging sensors. Considering the model for segmenting of on field rolling stock image frames from various sensors, where the frame conditions changes continuously, one may find the difficulties in edge identification using CV model. As pointed out in [7], the computation time for getting an optimal curve for on field rolling stock frame segmentation is a considerable challenge.

Although CV method is a promising model for segmentation of images, we propose to adjust the CV model's image gradient (IG) i.e. the last term in eq'n (7) to morphological gradient difference (MGD) term. Image energy function for on field video sensor produced rolling stock frames cannot represent edge map effectively using CV model's IG. Therefore for edge mapping rolling field frames image gradient (IG) transforms into MSD using difference of morphological operators. For the RS video frame $I^{xy}: D \to \Re^2$ in space D, the morphological operator's dilation and erosion for gray scale image for a line structuring element with m rows and n columns $L^{mn}: D \to \Re^2, m, n = \{-M, ..., M\}$ are defined as

$$I_d^{xy} = I^{xy} \oplus L^{mn} = \{\max(I^{x-m, y-n} - l^{mn} \mid m, n = -N \text{ to } N\} \quad (10)$$

$$I_e^{xy} = I^{xy} \ominus L^{mn} = \{\min(I^{x-m, y-n} - l^{mn} \mid m, n = -N \text{ to } N\} \quad (11)$$

where M means size of the line structuring element L^{mn}. A set of four different line orientations $\{\pi/4,\ \pi/2,\ 3\pi/4,\ \pi\}$ with single neighborhood overlapping pixels forms structuring element L^{mn}. The

morphological gradient operator for dilation is ∇^d and ∇^e for erosion. The term in equation (5), $\nabla.\frac{\nabla O^{xy}}{|\nabla O^{xy}|}$ adjusts into morphological gradient difference term as $(\nabla^d).\frac{\nabla^d O^{xy}}{|\nabla^d O^{xy}|}-(\nabla^e).\frac{\nabla^e O^{xy}}{|\nabla^e O^{xy}|}$. This small change in calculating image energy is clearly observable in Figure-4.

The arrow points to the region in Figure-4(a) and (b) for comparison between the IG in regular CV and MGD in adjusted CV having a clear edge. Also there is a 44% drop in number of iterations, making the adjusted CV algorithm faster compared to traditional algorithm.

4. RESULTS AND DISCUSSIONS

Manual rolling stock examination is identifying defects of bogies under train movement. This paper models to accomplish the same task using image processing. The first task in the defect identification is to segment various parts of a moving train. To capture video images of bogies, a high speed camera was setup at the edge of railway station. The first step in the segmentation of rolling stock involves contrast enhancement of the video frames. Histogram equalization is a good option if a uniform enhancement is required.

But the captured video frames of moving bogies possess unequal distribution of pixel intensities. Hence, virtual image fusion contrast enhancement is used to preserve video object texture and color information. This can be observed in Figure-3.

The Chan vese active contours without edges [13] are given the task of segmentation of enhanced video frames. It's a gradient based method. Active contour models are iterative methods that use image gradient as a stopping criterion. There must be a minimum number of iterations for videos. In this case the video frames are very complex and there is huge demand for the processing power. Hence in this paper we proposed a new method with morphological differential gradient instead of image gradient. Our results show an increase in speed and segmentation efficiency. Figure-4 shows a particular frame segmented using CV and CV-MDGS method. Figure-4 (a) is segmented with CV model using image gradient and Figure-4(b) with the same CV model with modified image gradient.

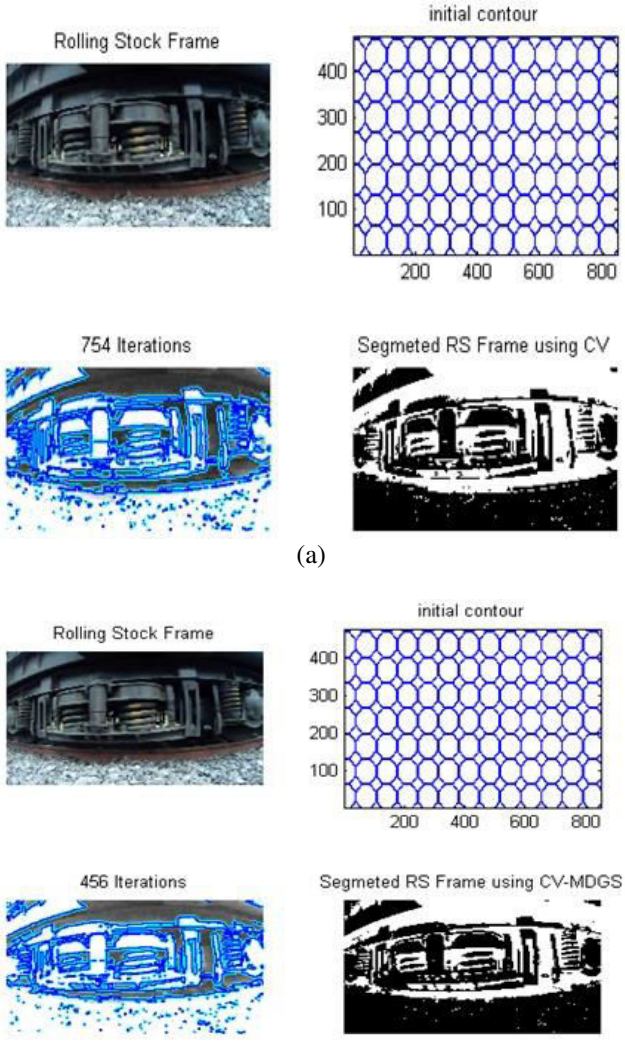
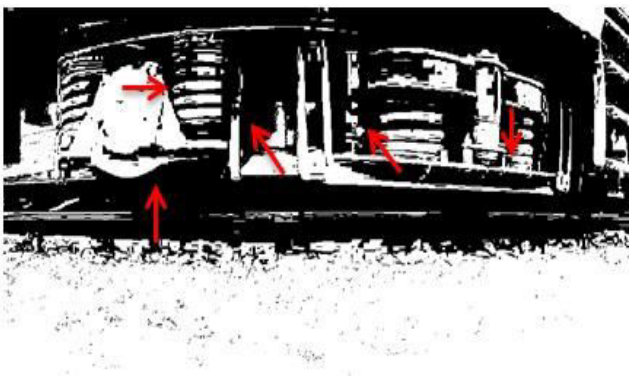
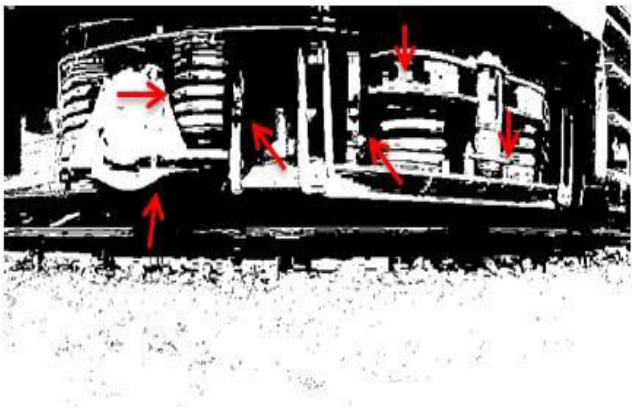


Figure-4. (a) Chan Vese Active contours without edges [6]. (b) CV-MDGS.



(a)



(b)

Figure-5. (a) Segmentation with CV Model [6] (b) segmentation with CV Morphological differential gradient.

The arrows mark the points of difference between the two models as shown in Figure-5. Moreover the proposed CV-MDGS model is around 44% faster compared to original CV. Figure-6 shows frames of rolling stock segmented under various lighting conditions from the proposed CV-MDGS method. Both frames are from different trains at two different time steps.

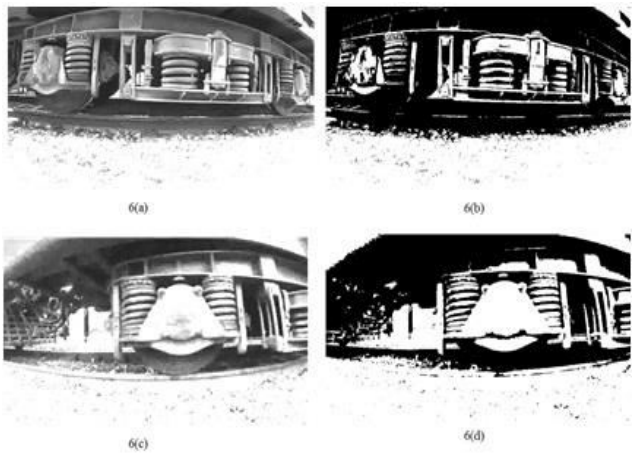


Figure-6. Segmentation results from the proposed algorithm CV-MDGS (a) Contrast enhanced at 11.30AM (b) It's Segment (c) Contrast enhanced at 2.00PM, and (d) segmented frame.

Figure-7 gives the reason for enhancement of video frames. Figure-7(a) shows original image. Figure-7(b) is the segmented frame with CV-MDGS with contrast enhancement for a k value of +3 to -3 and 7(c) with contrast enhancement with k value between +5 and -2. Figures-7(d) and 7(e) are segmentations from the two contrasted frames respectively. Observations on the Figures-7(d) and 7(e) highlight the need for contrast enhancement as a preprocessing step for this application.

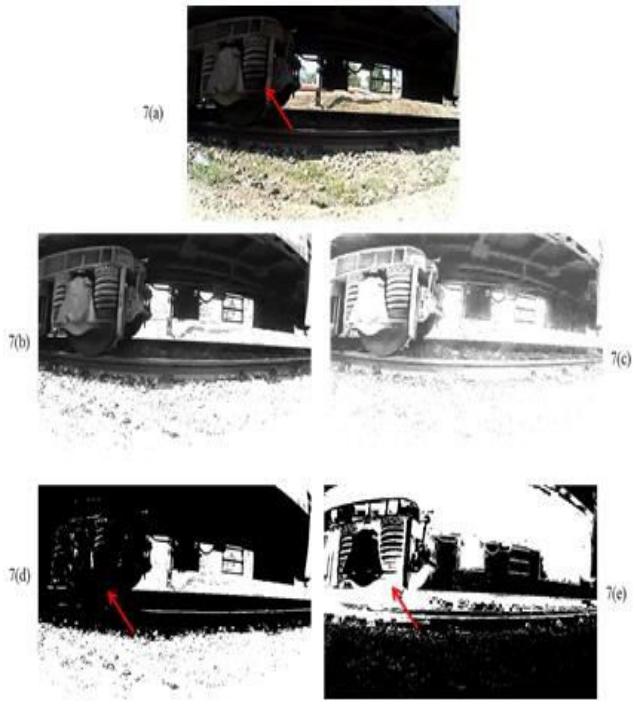


Figure-7. (a) Original frame (b) Contrast enhanced with k=-3 to 3(b) Contrast enhanced with k=5 to -2 (c) Segmented frame with 7(b), and (d) Segmented frame with 7(c).

5. CONCLUSIONS

This paper looks at the automation of rolling stock examination during train movement. Here a brief description of the problems associated with rolling examination was given. Automation is proposed with a camera model. The video frames are captured under different time steps of the day near a railway station when the train is moving at 30kmph. To compensate for brightness and contrast variation a uniform contrast enhancer is used. The enhancer computes virtual images and performs wavelet based fusion to enhance contrast. After testing with a few segmentation models, active contours Chan vese model is considered best for segmenting complex images. But for promising results the CV model is improvised into CV-MDGS model. This improvised model provides better and faster segmentation. There is a 44% speed increase and a 24% increase in segmentation when compared to ground truth.

REFERENCES

[1] Bali Akanksha Singh and Shailendra Narayan. 2015. "A Review on the Strategies and Techniques of Image Segmentation," Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on, pp.113, 120, 21-22 February.

[2] Zhang Shaoming He, Fang Zhang, Yunling Wang, Jianmei Mei and Xiao Feng Tiantian. . 2015 "Geometric active contour based approach for segmentation of high-resolution spaceborne SAR

images," Systems Engineering and Electronics, Journal of Vol.26, No.1, pp.69-76, February.

[3] Guang Yang, Kexiong Chen, Maiyu Zhou, Zhonglin Xu and Yongtian Chen. 2007. "Study on Statistics Iterative Thresholding Segmentation Based on Aviation Image," Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on, Vol.2, pp.187-188, July 30 2007-August 1.

[4] Yongjian Yu and Acton S.T. 2004. "Edge detection in ultrasound imagery using the instantaneous coefficient of variation," Image Processing, IEEE Transactions on, Vol. 13, No. 12, pp. 1640-1655, December.

[5] Patankar S.S., Mone A.R. and Kulkarni J.V. 2013. "Gradient features and optimal thresholding for retinal blood vessel segmentation," Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on, Vol., No., pp.1-5, 26-28 December.

[6] Chan T. and Vese L.A. 2001. “Active contours without edges,” IEEE Transactions on Image Processing, Vol. 10, No. 2, pp. 266–277.

[7] Xincai Huang, Huang Bai and Sheng Li. 2014. "Automatic aerial image segmentation using a modified Chan-Vese algorithm," Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on, pp.1091-1094, 9-11 June.

[8] M. Kass, A. Witkin and D. Terzopoulos. 1987. “Snakes:active Contour Models”, International. Journal of Computer Vision, Vol. 1, pp. 321-331

[9] Kishore P.V.V., Sastry A.S.C.S. and Kartheek A. 2014. "Visual-verbal machine interpreter for sign language recognition under versatile video backgrounds," Networks & Soft Computing (ICNSC), 2014 First International Conference on, pp.135-140, 19-20 Aug.

[10] Noor N. M., Khalid N.E.A., Ali M.H. and Numpang A.D.A. 2010. "Fish Bone Impaction Using Adaptive Histogram Equalization (AHE)," Computer Research and Development, 2010 Second International Conference on, pp.163-167, 7-10 May.

[11] Chang-Hsing Lee, Ling-Hwei Chen and Wei-Kang Wang. 2012. "Image contrast enhancement using classified virtual exposure image fusion," Consumer Electronics, IEEE Transactions on, Vol.58, No.4, pp.1253-1261, November.

[12] Kishore P. V. V., P. Rajesh Kumar, E. Kiran Kumar and S. R. C. Kishore. 2011. "Video Audio Interface for Recognizing Gestures of Indian Sign." International Journal of Image Processing (IJIP), Vol.5, No. 4, pp. 479-503, September.

[13] Kishore P. V. V. and P. Rajesh Kumar. 2012. "A Model For Real Time Sign Language recognition System." International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, No. 6, pp.29-35, June.