Carvana Image Masking Challenge

Bhavya Patwa (1401063), Deval Shah (1401060), Karan Patel (1401113), Parth Shah (1401054)

***Abstract*—Carvana through kaggle has posed an image segmentation challenge, which is tackled by using classical computer vision approaches. Company has provided images of different type of cars captured at various angles. 16 different angle shot images of a same automobile is given. As problem can also be solved using supervised learning techniques, image masks for all the images are also given. Through this project a comparison of various unsupervised segmentation techniques like segmentation through multiple cues, water-shed algorithm and global threshold techniques is done. A Kernelized structural SVM learning based approach is also visited. A comparison between accuracy and efficiency of various techniques is also presented.**

***Keywords*—Integration of multiple cues, minimum descriptor length(MDL), watershed, Kernelized Structural Support Vector Machines (SVM).**

I. INTRODUCTION

Image Segmentation is a revered computer vision problem. Carvana, a successful start-up dealing with sales and purchase of used cars posed a problem of segmenting their given data-set of images. Deep learning approach discussed in [1][2] is producing state-of-the-art results. Deep neural nets implementation are not cost efficient which has motivated us to look at classical computer vision approaches to tackle the problem. We have compared water-shedding techniques, global threshold approach, multiple cues based segmentation and Kernelized SVM in terms of accurately giving the foreground pixel points and computational time efficiency.

II. ALGORITHMS FOR SEGMENTATION

A. Watershed

Watershed is a transformation defined on a gray scale image. The watershed algorithm implemented is a highly revered Meyer's algorithm [3]. It uses a topological graph, and dark and light properties of image for segmenting an image in foreground and background. Results for a watershed algorithm are quite accurate for car images not having the same color map as of background. This is an highly constrained situation and hence other methods needs to be explored.

B. Global Threshold Technique

We implemented several global threshold techniques like OTSU, Yen, Isodata, Triangle, Li, Minimum, Mean. As the images taken are in highly lit area a global thresholding approach is not suitable where intensities of foreground and background are almost equal. Also threshold techniques like minimum gives shaky results producing good results for angle 3,7,9,12,15. For all the future analysis OTSU global thresholding techniques are considered as OTSU generates a good segmented image for almost every angle.

Mehul Raval is with the Department of Information and Communication Technology, School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, India e-mail: mehul.raval@ahduni.edu.in

C. Image Segmentation by multiple cues

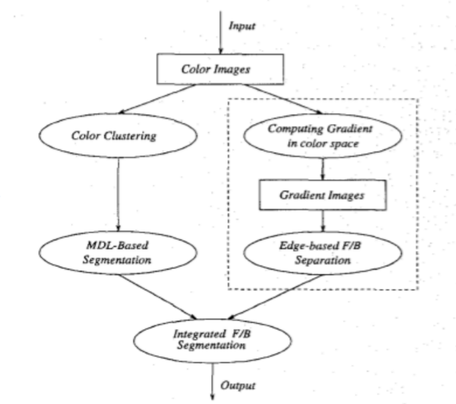

Fig. 1: Multiple Cues Algorithm

Figure 1 shows the model for image segmentation followed in [5]. The only assumption for this analysis is that background is slow varying and foreground is centered, which is satisfied in our given kaggle data-set. The algorithm for fitting a polynomial to classify a given image as foreground or background is presented and coded in [4].

D. Kernelized Structural SVM

In this paper, they have proposed a supervised segmentation approach that tightly integrates object-level top down information with low-level image cues. The information from the two levels is fused under a kernelized structural SVM learning framework. A novel nonlinear kernel is defined for comparing two image-segmentation masks. This kernel combines four different kernels: the object similarity kernel, the object shape kernel, the per-image color distribution kernel, and the global color distribution kernel. The result is a segmentation algorithm that not only knows what good segmentations are, but also learns potential segmentation mistakes and tries to avoid them.
The idea behind Structured SVM is to discriminatively learn a scoring function $F : X * Y \to R$ over input/output pairs (i.e. over image/mask pairs). Once this function is learned, the prediction function f can be obtained by maximizing F over all possible $y \in Y$, for a given image input:

$$y^* = f(x) = \underset{y \in Y}{argmax} F(x, y)$$

[6] claims that Kernelized SVM works for almost all datasets, but implementation is not shown here to verify the fact. However as SVM falls under supervised category one can be assured of a decently fair accuracy by training 16 different models.

III. Results

A. Watershed

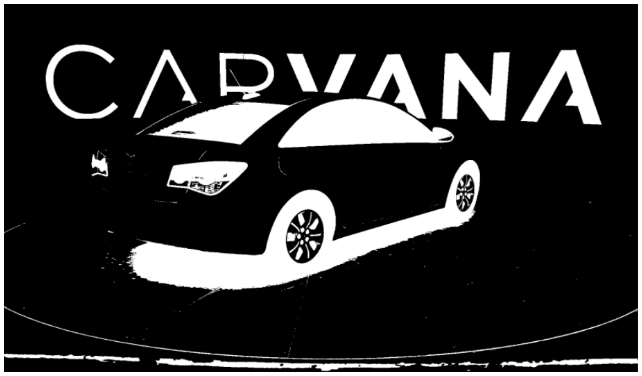

Fig. 2: Output of Watershed Algorithm

B. Global Threshold Technique

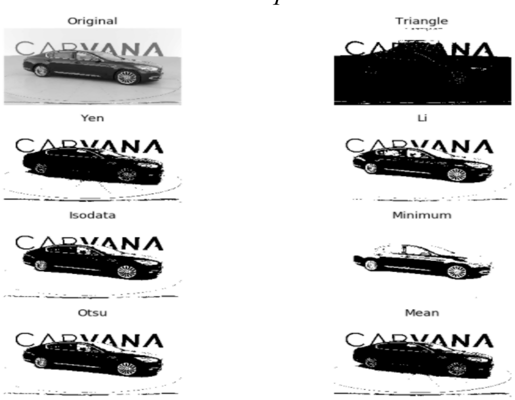

Fig. 3: Output of Global Threshold

C. Image Segmentation by multiple cues

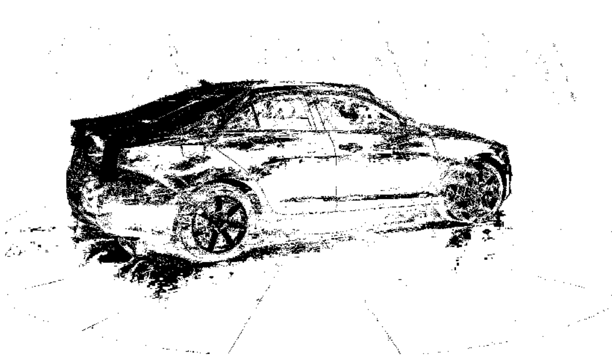

Fig. 4: Output by Multile Cues

IV. Discussing Accuracy and Efficiency of Unsupervised Techniques

A. Accuracy

Table-1 compares average FB1 score on 320 different images which comprises of all 16 angles for different varieties of cars ranging from Sedan , SUV, to hatchback.

Algorithm	**Precision**	**Recall**	**FB1-Score**
Watershed	79.344%	98.886%	88.044%
Global Thresholding	57.843%	88.121%	69.839%
IBM's approach	86.652%	99.445%	92.607%

TABLE I: Accuracy Comparison

B. Efficiency

Efficiency

Algorithm	**Time Taken(mins)**
Watershed	≈ 15
Global Thresholding	≈ 5
IBM's approach	≈ 20

TABLE II: Efficiency Comparison

V. Conclusion

By discussing segmentation through classical computer vision approach we observed that in a slightly varying background with foreground centered multiple cues performs more efficiently and accurately than all other unsupervised techniques. Kernelized SVM being a supervised approach incurs a certain training cost, but can efficiently perform on test dataset once trained. Thus we conclude by commenting that image segmentation in a constrained scenario like Carvana Image Mask Challenge does not need to suffer high computational cost by applying deep learning techniques, as results generated by the classical Computer Vision techniques are almost as accurate as deep learning methods.

References

[1] Carvana-image-masking-challenge

[2] Implementation of deep learning framework – Unet, using Keras

[3] Meyer's Watersheding Algorithm

[4] Minimum Description Length Binning

[5] Foreground/background segmentation of color images by integration of multiple cues

[6] Kernelized structural SVM learning for supervised object segmentation