A Combined Method with automatic parameter optimization for Multi-class Image Semantic Segmentation

S.Suganya	J.Nivedha	S.Pavithra	R.K.Ramiya
Asst.prof/IT,	Dept of IT,	Dept of IT,	Dept of IT,
SREC,Coimbatore,	SREC,Coimbatore,	SREC,Coimbatore,	SREC,Coimbatore,
suganyasb07@gmail.com	nivejaya22@gmail.com	pavisrinivas92@gmail.com	rkramiya@gmail.com

ABSTRACT-- Multi-class image semantic segmentation deals with many applications in consumer electronics fields such as image editing and image retrieval. Segmentation is done by combining the top down and bottom-up segmentation. Top-Down Process can be done by Semantic Texton Forest and bottom up- process using JSEG. These two segmentation process can be executed in a combined manner. But this cannot choose the optimal value of JSEG parameter for each interested semantic category. Hence an automatic parameter selection algorithm has been proposed. An automatic parameter selection technique called an automatic multilevel thresholding algorithm using stratified sampling and PSO is used to remedy the limitations.

Keywords- Combined segmentation , Learning to segment, Multi-class image segmentation, Multi-Scale, Semantic Texton Forest.

I.   INTRODUCTION

Image processing is any form of signal processing for which the input is an image or video frame and output of image may be either an image or set of characteristics related to the image. The visual scene requires the ability to recognize objects and their location in the image. These two goals are essentially the problems of recognition and Segmentation. This is considerable computational challenges. The dominant approach to segmentation has been that of a bottom-up (BU) process, primarily involving the incoming image, without using stored object representations. The image is first segmented into regions that are relatively homogeneous in terms of colour , texture, and other image based criteria, and a recognition process is then used to group regions corresponding to a single, familiar, object. According to this approach, segmentation facilitates recognition. Another approach to segmentation is that of a top-down (TD), high-level visual process, in which segmentation is primarily guided by stored object representations. The object is first recognized as belonging to a specific class and then segmented from its background using prior knowledge about its possible appearance and shape. In other words, according to this approach, recognition facilitates segmentation. BU segmentation algorithms provide impressive results in the sense that they can be applied to any given image to detect image discontinuities that are potentially indicative of object boundaries. Their major difficulties, however, include the splitting of object regions and the merging of object parts with their background. These shortcomings are due to prior knowledge of the object class, since most objects are non- homogeneous in terms of colour, texture, etc.

Moreover, object parts do not necessarily contrast with their background. TD segmentation uses prior knowledge of the object class at hand to resolve these BU ambiguities. However, it also has difficulties due primarily to the large variability of objects within a given class, which limits the ability of stored representations to account for the exact shape of novel images. In this work, we introduce a segmentation scheme that addresses the above challenges by combining TD and BU processing to draw on their relative merits. The TD part applies learned "building blocks" representing a class to derive a preliminary segmentation of novel images. This segmentation is then refined using multiscale hierarchical BU processing. Our TD approach was introduced in [1], and later extended to include automatic learning from un-segmented images [2], as well as a preliminary scheme for combining BU processing [3]. The current version formulates the TD, as well as the combination components using a computationally efficient framework. It presents a fragment extraction stage that, unlike previous methods, produces a full cover of the object shape. This improvement is due to a modified mutual information criterion that measures information in terms of pixels, rather than images. This version also refines the automatic figure-ground labelling of the extracted fragments through an iterative procedure relying on TD/BU interactions. Another new aspect is the use of segmentation for improving recognition.

II. RELATED WORK

A. Bottom-up segmentation

Bottom-up segmentation approaches use different image based criteria and search algorithms to find homogenous segments within the image. A common bottom-up approach is to use a graph representation of the image (with the nodes representing pixels) and partition the graph into subsets corresponding to salient image regions. A recent bottom-up segmentation [4] produces a multiscale, hierarchical graph representation of the image. The resulting image takes colour, texture, intensity and boundary properties of image-regions. In this manner, the image is segmented into fewer and fewer segments. This process produces a segmentation weighted, hierarchical graph in which each segment is connected with a relating weight to any other segment at a coarser level, if the first was one of the segments used to define the latter. Neighbouring segments within each level in the hierarchy are connected with appropriate weights that reflect their similarities. This process produces efficient number of pixels. The algorithm also provides a measure of saliency (related to a normalized-cut function) that ranks segments according to their distinctiveness. The saliency is reflected by an energy measure Γi that describes the segment's dissimilarity to its surrounding, divided by its internal homogeneity. Uniform segments that contrast with their surrounding (e.g. a uniform black segment on a white background) will be highly salient, and will therefore have very low energy Γi, whereas any segments against similar background will have low saliency and high energy Γi.

B. Top-down segmentation

The top-down segmentation approaches rely on acquired class-specific information, and can only be applied to images from a specific class. These include deformable templates [5], active shape models (ASM) [6] and active contours [7]. A recent top-down, class-specific segmentation approach deals with the high variability of shape and appearance within a specific class by using image fragments (or patches). These fragments are used as shape primitives for the class. Segmentation is obtained by covering the image with a subset of these fragments, followed by the use of this cover to delineate the figure boundaries. The approach can be divided into two stages – training and segmenting. In the training stage, a set of informative image fragments is constructed from training data to capture the possible variations in the shape and appearance of common object parts within a given class. The figure-ground segmentation of each fragment is then learned automatically [8], or can be given manually, and used for the segmentation stage. A set of classifying fragments are derived from the fragment set. These are used in the segmentation stage to classify a novel input image as well as to detect the approximate location and scaling of the corresponding objects [9, 10, 11, and 12]. The entire fragment set also provides an over-complete representation of the class. For instance, for the class of horses, the set contains a large repertoire of fragments representing different options for the appearance and shape of the legs, tail, body, neck and head. Consequently, in a given class image, detected fragments are overlapping, and together they are likely to completely cover the entire object. The same image area can be covered by a few alternative fragments. The fragments are also free to move with respect to each other as long as they preserve consistency, allowing variety in shape. Each covering fragment applies its figure-ground segmentation to vote for the classification of the pixels it covers. The overall voting defines a figure-ground segmentation map $T(x, y)$ of the image, which classifies each pixel in the image as figure or background. The map can be given in either a deterministic form (a pixel can be either figure, $T(x, y) = 1$, or background, $T(x, y) = -1$) or a probabilistic form (figure with likelihood $T(x, y)$ and background with likelihood $1 - T(x, y)$).

C. Combined Segmentation

The method automatically learns an object representation called Pictorial Structures (PS) from video sequences. The PS is combined with a contrast dependent Markov Random Field (MRF) that biases the segmentation to follow image boundaries. Un-segmented images to learn a global figure/ground mask and a global edge mask that represent the "average" shape and edges of objects in the class. Shape and edge variations are constrained solely by a smoothness constraint. The global shape approach is limited in its ability to address rigid objects whose shape largely deviate from the "average." Additionally, the assumption of different object and background regions may be violated, especially in gray-level images. However, it requires a manually segmented training set. It also assumes simple transformations that can align each object instance with a canonical grid. This assumption makes it hard to handle object classes with high shape variability. The object representation is set manually and the part configurations areas (V1, V2) are influenced by higher level neurons, depending on figure-ground relationships [13], [14]. In particular, many edge-specific units in low level visual areas respond differently to the same edge, depending on the overall figure-ground relationships in the image.

III.PROPOSED WORK

In proposed work, Multi-class image semantic segmentation (MCISS) can be done by using combined approach with automatic parameter selection technique. Image can be segmented by using two types of approaches: Top-down and bottom –up approach. Top –down approach uses Semantic Texton Forests (STF) [15] and bottom up-process can be done by using JSEG. To increase the average and global accuracy the optimized parameter of JSEG can be proposed. An automatic parameter selection technique called an automatic multilevel threshold algorithm using stratified sampling and Particle Swarm Optimization approach [16]. The proposed method automatically determines the appropriate threshold number and values by (1) dividing an image into even strata (blocks) to extract samples; (2) applying a PSO based optimization technique on these samples to maximize the ratios of their means and variances (3) preliminarily determining the threshold number and values based on the optimized samples. Experimental result provides improved average and global accuracy than existing work.

A. *TD Segmentation based on STF*

As a TD method, the STF algorithm contains two separate stages: (i) a learning stage and (ii) a predicting stage. The learning stage of the STF method consists of three steps: (i) A Semantic Texton Forest (STF) is trained based on the semantic textons, which are essentially a kind of local appearance features (ii) A non-linear support vector machine (SVM) is trained based on the bag of semantic textons (BoST), which is computed over the whole image and captures the long-range contextual information (iii) a segmentation forest (SF) consisting of T decision trees is constructed based on rectangle count features, which act on both the semantic texton histograms and BOST region prior to encode texture, layout and semantic context information. At the predicting stage, given a pixel i , a decision tree t in the SF works by recursively branching left or right down the tree until a leaf node $t\,l$ is reached. The forest produces a class distribution by averaging the learned class distribution over the leaf nodes.

B. *BU Segmentation based on JSEG*

BU method used to detect a set of image regions. Each of these regions is homogeneous in color-texture, and thus tends to belong to the same object. Furthermore, a quite pleasing property of these BU segmentation results is that the accurate object boundaries are always lying in the obtained region boundaries. Specifically, the JSEG-based BU segmentation process consists of two independent steps: color quantization and spatial segmentation. In the color quantization step, an input color image is converted to a color-class map by replacing its pixel colors with their corresponding color class labels. A color class is the set of image pixels quantized to the same color. A measure J is then defined on top of the class-map to measure the quality of a given BU segmentation. Further, the local averaged J is proposed as the criterion to be minimized over all possible ways of segmenting the image. In the spatial segmentation step, a J-image is produced by calculating the J values over local windows centered on each pixel. The J image is segmented using a multi-scale region growing method. An agglomerative clustering process is adopted to merge the over-segmented J-image regions based on their color similarity. In this way, the homogeneous BU regions are obtained.

C. *Optimization*

In this module an image is treated as a population that contains the gray values of pixels. This method consists of four main steps: 1) An image is evenly divided into several blocks (strata), and a sample is taken from each stratum.(2) The threshold number and values of sample is optimized by PSO whose fitness function is the ratio of its mean and variance. Stratified Sampling is a method of sampling from a population which often improves the representative of the sample by reducing its sampling error. Stratification always achieves great precision provided that the strata have been chosen so that members of the same stratum are closely similar with regard to the characteristics of interest's .All the samples in the image are arranged from left to right and top to bottom orientations. They are denoted as $\omega_1,\omega_2,...,\omega_{16}$.The mean and variance of ω_i are m_i and s_i respectively, where i=1,2,…,16. The fitness function of each sample is formulated as $f_i = m_i / s_i$.

D. *combining TD and BU segmentation*

Since BU segmentation results are highly dependent on the threshold ☞ the set of regions partitioned from an input image through BU segmentation are denoted as $R(☞)$ and one of such regions is $r(☞)$.

When σ is small, the pixels in a BU region $r(\sigma)$ can be assumed to have the same category distribution. The distribution is formulated as

$$\text{E.} \quad P(c \mid r(\sigma)) = \frac{1}{Z} \sum_{i \in r(\sigma)} P_\theta(c \mid i)$$

Algorithm of TD&BU Combination

Input:

1. The homogeneous region set $R(\sigma)$ in BU,
2. The category probability distribution
$P_\theta(c \mid i)$ $e\ P\ c\ i$, $c \in \{1,...,C\}$ at each location i in TD.

Output:

The category set $\{ c^*_{r(\sigma)} \}$ for each pixel in each region $r(\sigma) \in R(\sigma)$.

    1: **for** each region $r(\sigma) \in R(\sigma)$ **do**

    2: Initialize $P(c \mid r(\sigma)) = 0$.

    3: **for** each location i $\in r(\sigma)$ **do**

    4: Accumulate the $P_\theta(c \mid i)$ of pixel i in the region $r(\sigma)$

    5: **end for**

    6: The category $c^*_{r(\sigma)}$ of all pixels in region $r(\sigma)$ is determined

    7: **end for**

    8: **return** Set $c^*_{r(\sigma)}$, $r(\sigma) \in R(\sigma)$.

Fig 1. Algorithm for TD&BU combination

IV. IMPLEMENTATION

The random image is selected then it is given as input to the system. The input image accuracy is improved by the following techniques. The input is taken from the database for segmentation.
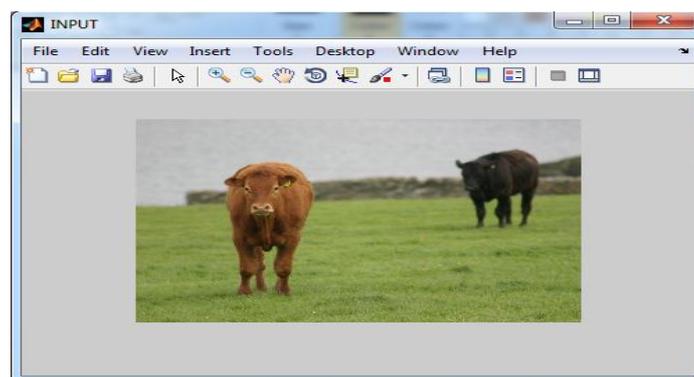

Fig 2. Input image

The above image is allowed to the process of top down segmentation in which Semantic texton forest is implemented. By this method the regions are identified .The threshold values are used to segment regions.
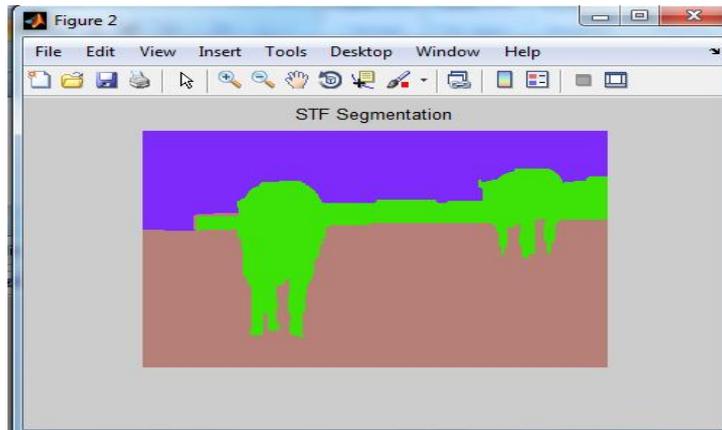
Fig 3. Top down segmentation

The regions are detected by implementing Semantic Texton forest method and the output for top down segmentation is shown above
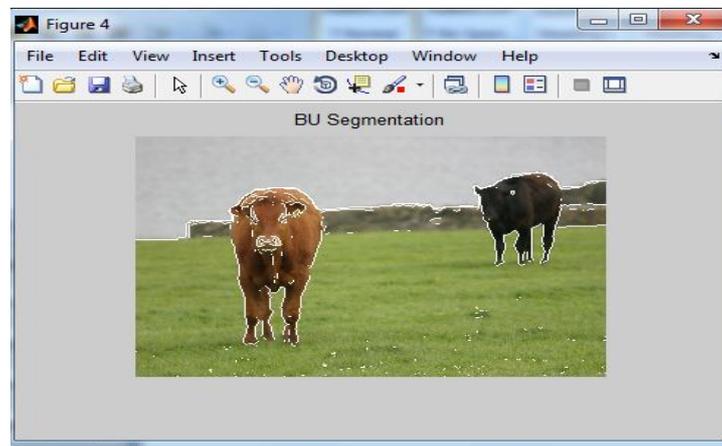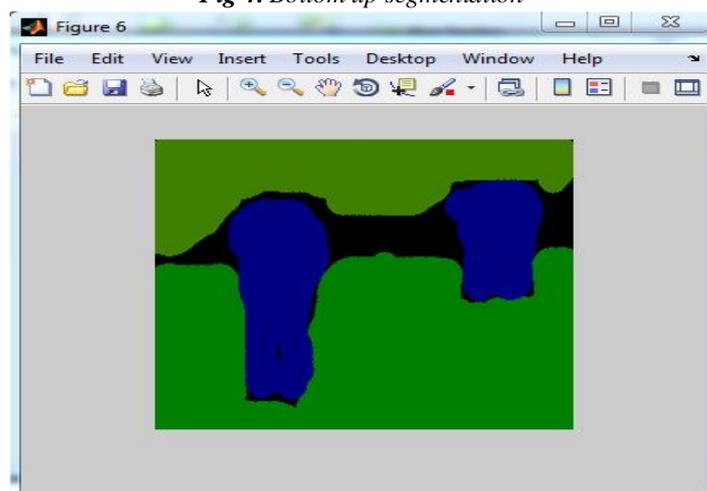

Fig 4. Bottom up segmentation

Fig 5.Combined method

The segmented image is send for further segmentation which is bottom up method .In which spatial segmentation and color quantisation is performed. The resulting image is shown above. If the pixel value increases the image quality increases. The images extracted from the above two methods are then combined to form the final segmented output.

V.CONCLUSION

In this paper, we proposed automatic parameter optimization for segmenting   multi class image. This technique has better parameter for more accuracy. When applied to an input image, the proposed method generates the final semantic segmentation through two stages. In the first stage, the category probability distribution of each pixel is estimated by the TD process meanwhile a set of homogeneous regions are partitioned by the BU segmentation process. In the second stage, these TD and BU segmentation results are combined together to generate the final semantic segmentation. Experimental results reveal that the combined method can achieve better segmentation accuracies without notably prolonging the computation time. It gives more accuracy then the existing.

REFERENCES

[1] Z. Zhu, Y. Wang, and G. Jiang, "Statistical Image Modeling for Semantic Segmentation", ", IEEE Trans. on Consumer Electronics, vol. 56, no. 2, pp. 777-782, 2010.
[2] S. N. Sulaiman and N. A. M. Isa, "Adaptive Fuzzy-K-means Clustering Algorithm for Image Segmentation " , IEEE Trans. on Consumer Electronics, vol. 56, no. 4, pp. 2661-2668, 2010.
[3] G. Csurka and F. Perronnin, "An efficient approach to semantic segmentation," International Journal of Computer Vision, vol. 95, pp.198–212, 2011.
[4] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in CVPR, Anchorage, AK, 2008, pp. 1–8.
[5] Y. Deng, B. S.Manjunath, and H.Shin, "Color image segmentation," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'99, Jun 1999, vol. 2, pp. 446–451.s
[6] T. K. Ho, "Random decision forests," in Proc. Int'l Conf. Document Analysis and Recognition, Montreal, Aug. 1995, pp. 278–282.
[7] X. M. He, R. S. Zemel, and M. A. Carreira-Perpinan, "Multiscale conditional random fields for image labeling," in IEEE Conf. On Computer Vision and Pattern Recognition. (CVPR). 27 June-2 July, 2004, vol. 2, pp. 695–702.
[8] X. He, R. S. Zemel, and D. Ray, "Learning and incorporating top-down cues in image segmentation," in In ECCV. 2006, pp. 338–351, Springer.
[9] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," Int. J. Comput. Vision, vol. 80, pp. 300–316, December 2008.
[10] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in European Conf. on Computer Vision (ECCV). 2006, vol. 3951 of Lecture Notes in Computer Science, pp. 1– 15, Springer Berlin / Heidelberg.
[11] J. Shotton, J. W., C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," Int. Journal of Computer Vision, vol. 81, pp. to appear, January 2009.
[12] N. Plath, M. Toussaint, and S. Nakajima, "Multi-class image segmentation using conditional random fields and global classification," in Intl. Conf. on Machine learning (ICML), Montreal, Quebec, Canada, 2009, pp. 817–824, ACM.
[13] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimalboundary & region segmentation of objects in N-D images," 2001, vol. 1, pp. 105–112. 2
[14] Y. Pan, J. D. Birdwell, and S.M. Djouadi, "An efficient bottom-up image segmentation method based on region growing, region competition and the mumford shah functional," in Multimedia Signal Processing, 2006 IEEE 8th Workshop on, oct. 2006, pp. 344 –349. 2
[15] Z. Tu, X. Chen, A.L. Yuille, and S.-C. Zhu, "Image parsing: unifying segmentation, detection, and recognition," in Proceedings of Ninth IEEE International Conference on Computer Vision, oct. 2003, pp. 18–25 vol.1.
[16] M. P. Kumar, P. H. S. Ton, and A. Zisserman, "Obj cut," CVPR 2005. vol. 1, pp. 18 – 25 vol. 1. 2