Bayesian-Network-Based Algorithm Selection with High Level Representation Feedback for Real-World Intelligent Systems

Martin Lukac
Graduate School of Information Sciences
Tohoku University
Miyagi, Sendai 980-8579, Japan
lukacm@ecei.tohoku.ac.jp

Michitaka Kameyama
Graduate School of Information Sciences
Tohoku University
Miyagi, Sendai 980-8579, Japan
kameyama@ecei.tohoku.ac.jp

***Abstract*—A real-world intelligent system consists of three basic modules: environment recognition, prediction (or estimation), and behavior planning. To obtain high quality results in these modules, high speed processing and real time adaptability on a case by case basis are required. In the environment recognition module many different algorithms and algorithm networks exist with varying performance. Thus, a mechanism that selects the best possible algorithm is required. To solve this problem we are using an algorithm selection approach to the problem of natural image understanding. This selection mechanism is based on machine learning; a bottom-up algorithm selection from real-world image features and a top-down algorithm selection using information obtained from a high level symbolic world description and algorithm suitability. The algorithm selection method iterates for each input image until the high-level description cannot be improved anymore. In this paper we present a method of iterative composition of the high level description. This step by step approach allows us to select the best result for each region of the image by evaluating all the intermediary representations and finally keep only the best one.**

***Keywords*—Natural Image Processing; Algorithm Selection; High Level Representation; Adaptive Rewriting**

I. INTRODUCTION

The complexity and the diversity of the environment make the robust real world information processing at most satisfactory and only for some particular applications. Such low reliability is partially due to the limited suitability of algorithms used for the processing of real world information (Fig. 1): each algorithm provides the best result for a particular set of environmental configurations (conditions). The common approach to solve this problem is an incremental increase of the algorithm's functionality that takes into account the newly encountered environmental configurations. This approach however often results in algorithms with very high complexity, limited scalability and continuously decreasing performance; a single aggregated algorithm cannot always robustly process the real-world information for all available conditions. But such robust performance is crucial in many real world applications such as intelligent cars or service robots.

In this paper we describe an algorithm selection approach to image processing and to the problem of image symbolic segmentation. The described algorithm selection framework uses a bottom-up and a top-down feedback that iteratively constructs an optimal high level image description. That is, given a set of algorithms and a set of distinctive features computed on the input images, the selection paradigm allows using the best algorithm on a case by case basis. From the input image features are extracted and a best algorithm selected. The result is a description of image content by a set of regions and labels. The obtained description is then verified on a symbolic level and a contradiction is generated. The contradiction is used to generate a hypothesis that resolves the contradiction. Then using the hypothesis a new algorithm is selected and a new description is obtained by merging the previous and the new description. This process is iterated until the high level description cannot be improved anymore.

The main contribution of this paper is the introduction and description of a method of merging the various high level descriptions and the contradiction resolution. The merging of the high-level hypotheses is an important part in the algorithm selection platform and a correct hypotheses merging is required to improve the overall result of the platform.

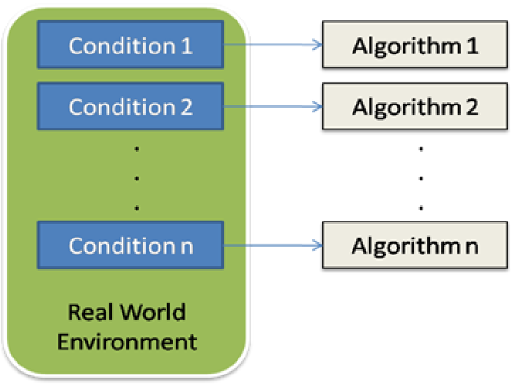

Fig. 1. Many algorithms are suitable for the same problems depending on the environmental conditions.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

This paper is organized as follows. Section II presents the algorithm selection framework. Section III describes the application of the platform to the problem of symbolic segmentation. Section IV briefly explains the algorithm selection mechanism using Bayesian Network and Section V introduces the verification process. Finally Section VI describes the high-level description merging and Section VII concludes the paper and discusses future work.

II. THE ALGORITHM SELECTION PLATFORM

The algorithm selection paradigm was originally introduced by Rice [1] and since various but only a relatively small amount of applications and studies have been made. Previous works related to image processing includes mainly the work of Yong [2] that used algorithm selection for segmentation in noisy artificial images and by Takemoto [3] that used algorithm selection to determine the best edge algorithm for edge detection in biological images. With respect to general robotic processing [4, 5] the concept of algorithm selection was introduced into the middle and high level processing of natural image segmentation and understanding. In particular, in [4] the algorithm selection was used to improve the segmentation of natural images.

The algorithm-selection framework is shown in Fig. 2.

The system operates as follows:

- A first cycle of processing starts with the features of the whole input image are extracted (Box 2) and are used as input to the algorithm selection mechanism (Box 3) that determines what algorithms should be used for processing the image (Box 1). The result of the processing is a high level description and thus the selection process select algorithms in all level of processing. The high level description is then verified (Box 4) for the correctness of the symbolic content and is analyzed for the existence of a logic contradiction. If contradiction was not detected the processing stops.
- A second cycle starts when a contradiction was detected in the high level description. The contradiction is used to extract features from the region where the contradiction is located (Box 2). At the same time, the contradiction and the high level description are used to generate hypothesis resolving the contradiction in the high level description (Box 5). The resulting information: region features, high level description and user specified context information are used as input nodes to a Bayesian Network (Box 3); Bayesian Network is only one of various possible algorithm selection methods such as machine learning, logical induction and so on. The output of the BN is a set of different algorithms that are applied only to the extracted region. The result of processing of the contradiction region is merged back into the whole image (Box 6). The new high level description is verified again and if contradiction occurs the process is repeated. This loop is repeated until there is no more contradiction in any region of the image.

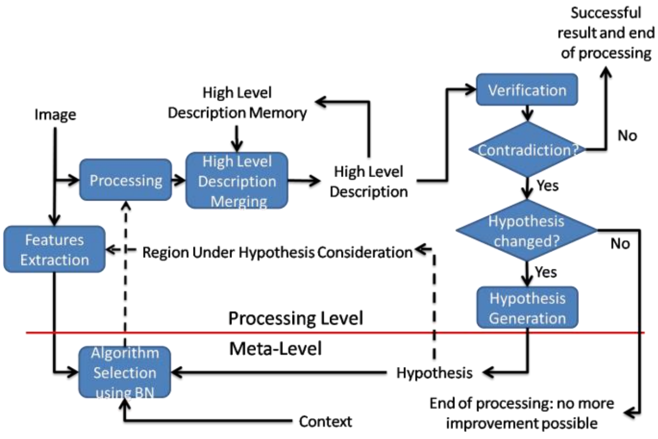

Fig. 2. Algorithm selection based platform for image understanding.

Note that the high-level representation is validated by the correct set of algorithms: only the correct algorithms will generate description without contradiction. Such processing however requires many processing cycles to obtain the desired result. This is due to the fact that each of the mechanism in the loop is in generally inaccurate and various combinations of algorithms needs to be tested to obtain a high level representation without contradiction [5]. Such requirement of high speed processing and many various algorithms is an ideal application for high speed reconfigurable VLSI implementation [6]. Consequently, in each iteration the algorithm must be selected carefully in order not to accumulate error but to ensure the overall convergence.

The algorithms used in this platform are highly heterogeneous. In the preprocessing level edge detection, noise removal, smoothing and color transformations are available. Some examples of algorithms are Canny or Prewitt edge detection. The segmentation contains various algorithms and various features extraction that these algorithms require. Examples of features required for segmentation are salience, histogram of oriented gradients [7] or simply brightness. Segmentation algorithms such as Global probability of boundary [8], maximally stable extremal regions, Normalized Cut [9], Salience Based Segmentation [10] can be used. The recognition processing contains only two distinct algorithms because recognition algorithms are a special case of pattern matching. The available algorithms are SVM matching [11] and components matching using SVM [12], and a Bayesian network approach. In the final level of processing - the understanding - algorithms in general can be divided into two categories: logic and probabilistic. However in this work our focus is mainly on the diversity of algorithms in the first three level of processing [5].

The most important fact is that the various algorithms in the proposed processing levels are highly heterogeneous and even within a same level the variety is relatively high. Moreover because the processing is adaptively changed not only for every image but also for various regions in the image, a real-time platform requires a high-speed platform permitting both the high speed of processing and of reconfiguration.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

III. Algorithm Selection for Image Symbolic Segmentation

The algorithm selection platform described in this paper is applied to the problem of semantic segmentation of real-world images. The semantic segmentation is a task where an image is segmented to a set of regions and then each region is assigned one label from the available labels. In this paper we are using the VOC 2011 image database that requires a labeling into 21 possible labels. These labels are: airplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, TV monitor and background.

To apply the algorithm selection we are using two well established algorithms from [11, 13-15]. Both of the algorithms take an input image and output a symbolic segmentation. Examples of outputs of both algorithms are shown in Fig. 3.

A. Algorithm Evaluation

The algorithms have been evaluated using the F-measure [16]. It is a common measure to determine the correctness of segmentation by evaluating both the true-negatives as well as the false-positives. Because the algorithms used in this work perform symbolic segmentation, the evaluation can be discretely quantized. The incorrect labeling generates a score of 0; while the correct labeling generates a score of 1. The F-measure of the segmentation results in a score in the range of [0,1] with 0 being no segmentation at all and 1 being perfect segmentation. Perfect segmentation means that the obtained segments match completely with the reality. Consequently the best possible result (correct labeling and perfect segmentation) results in a score value of 2.

In semantic segmentation the extreme failures can be categorized in three different categories, Fig. 4. The top row in Fig. 4(a) shows a nil result because the algorithm did not find any suitable segmentation or correct labeling. Fig. 4(b) shows good labeling but bad segmentation; and Fig. 4(c) shows good segmentation with bad labeling.

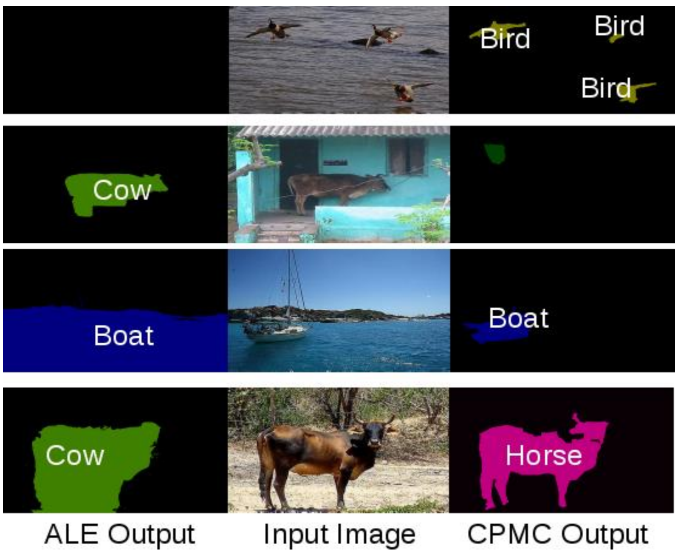

Fig. 3. Example outputs of the used algorithms for semantic segmentation.

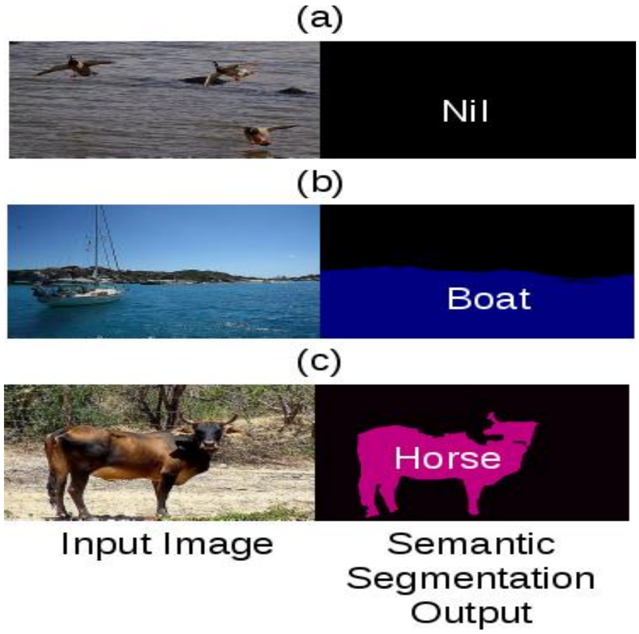

Fig. 4. Examples of extreme cases of failed semantic segmentation.

IV. Bayesian Network Algorithm Selection

The algorithms are selected based on their advantages and disadvantages. The advantages and disadvantages are extracted by the analysis of the training and validation data. In particular, for each algorithm we extract advantages by searching the common properties (image features, region attributes, and labels) for all output results where a particular algorithm is the best. The advantages and disadvantages are represented by a set of properties that are quantified into discrete levels and are used as inputs to the Bayesian Network.

The Bayesian Network used is depicted in Fig. 5. The input nodes and the structure of the algorithm were obtained after a minimization procedure described in [5].

The obtained input nodes represent the advantages and disadvantages of each algorithm used and thus depending on the algorithm the nodes will be changed accordingly. Here the advantages and disadvantages have been obtained by careful study of result cases such as those shown in Fig. 3. The analysis of the advantages and disadvantages is not the focus of this paper but for clarity we describe the main steps. From the initial results of the available algorithms, only the so called Min-Max data set is extracted. In a Min-Max dataset, one of the algorithms' results is of very high quality while all other algorithms' results are very bad. Such input images describe the

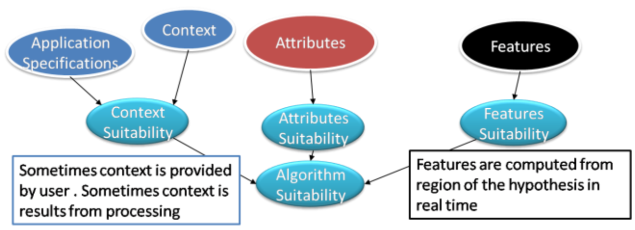

Fig. 5. Bayesian network used for algorithm selection.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

outstanding advantages of each algorithm. Each image is then analyzed using a set of predetermined features until the feature ranges are identified. The identified features and their ranges are then used to design algorithm specific input nodes to the BN.

V. High-Level Verification

The processing starts by extracting features from the input image, selecting the best algorithm and generating a high-level description, Fig. 6. Once a high-level description is obtained it is verified by a set of verification tests. These tests include objects proximity test, objects size test and objects co-occurrence tests. Each test is performed using a co-occurrence matrix obtained from training data. The whole verification system is depicted in Fig. 7. The verification starts from the result of symbolic segmentation or from the result of recognition with localization. First the overlap is estimated by comparing the various detected objects to a set of patterns for estimating the overlap. There are only two generic patterns and they are depicted in Fig. 8. The pattern form Fig. 8(a) is based on the fact that if a region is being partially surrounded by another larger region there is a probability that the partially surrounded region is overlapping (white rectangle in Fig. 8(a)) with the neighboring region. The next pattern is shown in Fig. 8(b), where the white region is splitting the dashed region into to separate parts. This type of overlap is detected by looking for two regions of the same type separated by a region of another type. Finally the last type of overlap detects a smaller object being completely surrounded by a larger region in the background as shown in Fig.8(c). The estimated overlap information is used to determine the relative position of the neighboring regions in the depth of the image (Fig. 7). The estimated depth then allows to adjust the various sizes of the regions in the image. This adjustment is necessary to allow objects of various sizes be scaled proportionally to their distance from the focus point. Such scaled objects then can be evaluated for their relative size table. The second component of the verification is based on relative position. Each region's center of gravity is extracted and relative position for each two regions is recorded. The obtained relative size and relative position data is entered into co-occurrence matrices generated form training data. There are four co-occurrence matrices for the relative position, one for each direction in a 2D space. If the

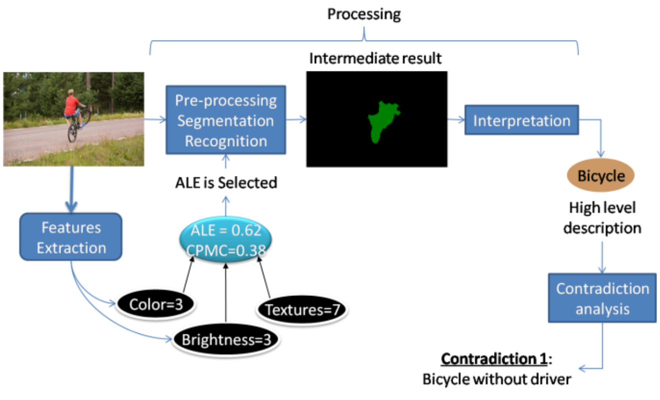

Fig. 6. First pass of processing.

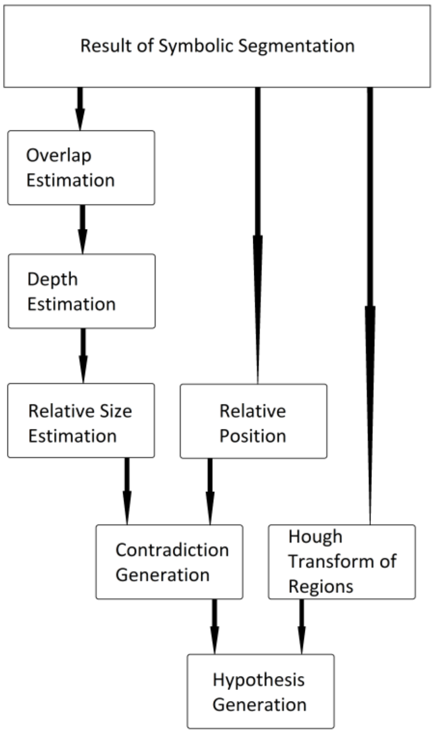

Fig. 7. The general flow of the verification and hypothesis generation.

scaled output from the four relative position and the one relative size co-occurrence matrices is larger than 0.5 we conclude that the current region under consideration contains a contradiction.

If a contradiction is detected, the region under consideration is used to extract lines representing the object in the region. The image from which the lines are extracted is a black and white result of symbolic segmentation and thus no textures or complex colors exists in the image. The Hough transform is applied only to the border of the region containing the contradiction and lines are classified in a histogram that is used to estimate the object that best matches the histogram. The estimated object label is the hypothesis used in later steps of processing. Fig. 9 shows the hypothesis generation and the region selected for re-processing. This time features are extracted only from the region under consideration and together with the hypothesis (that should remove the contradiction) constitutes new inputs to the Bayesian Network.

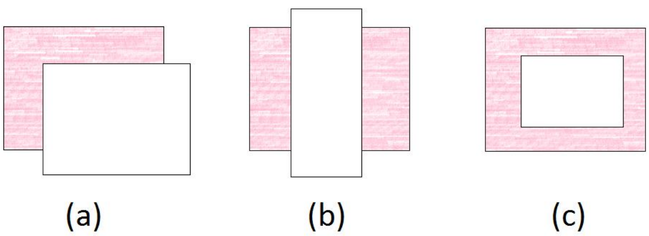

Fig. 8. The patterns used to estimate overlap of neighboring objects.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

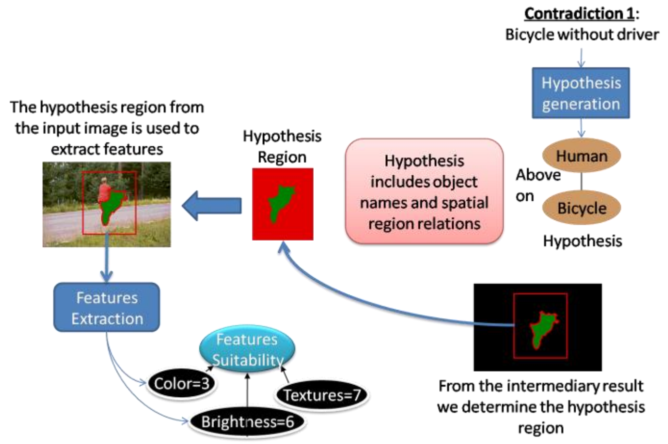

Fig. 9. The contradiction from the verification is used to indicate which regions should be reprocessed.

VI. High-Level Representation Rewriting

The obtained hypothesis candidate from the verification process is now used for the top-down feedback. First the original region containing the label to be changed is used to extract features. The features extracted are: brightness, Gabor edges, FFT, wavelets, contrast, accutance, gist [17], RGB colors, region properties (of black-and-white and gray scale image obtained using the *regionprops* function in Matlab) and 64 different textures (obtained from the source code from [16]).

The principle of selecting regions and replacing old one is a priority replacement. The priority of each result whether it is a full replacement or region based replacement depends on whether the new combined representation removes the contradiction under consideration. Our example shows that the algorithm selection replaces a part of the background by a labeled region indicating the person.

Notice that in this case the contradiction was resolved by simple substitution: the obtained result was in accord with the generated hypothesis. However this is not always the case because the hypothesis that would remove the contradiction might never be obtained by any available algorithms.

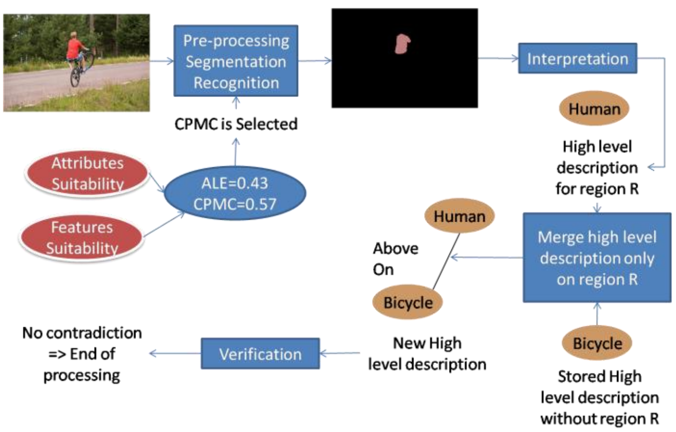

Fig. 10. Example of merging new hypothesis with original high level description and contradiction resolution.

In such case an alternate hypothesis can be selected. In the current status, the verification and hypothesis generation generates 20 hypotheses each with a probability with correctness. The algorithm stops if the new result description contains a hypothesis that has a probability of correctness smaller than the current one. Otherwise all hypotheses above the current probability of correctness are tested.

VII. Conclusion

In this paper we presented an algorithm selection platform for the symbolic image segmentation. The platform iteratively rewrites the final result until no more improvements are possible. This is either because no new algorithms can be selected (all have been tested) or when there is no more contradiction.

The future work includes the usage of more algorithms and to analyze the selection mechanism for larger number of labels. Additionally a more powerful high-level description verifier such that is not using co-occurrences is to be investigated.

As a future extension of this work, the algorithm selection requires improvements in the quality of the algorithm selection. Moreover, it is important to develop the packet generation mechanism such as an address generation unit which generates a series of packets with regular destination addresses for further reduction of the configuration memory size.

References

[1] J. R. Rice, “The algorithm selection problem,” *Advances in Computers*, vol. 15, pp. 65-118, 1976.

[2] X. Yong, D. Feng, and Z. Rongchun, “Optimal selection of image segmentation algorithms based on performance prediction,” in *Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing (VIP2003)*, 2003.

[3] S. Takemoto and H. Yokota, “Algorithm selection for intracellular image segmentation based on region similarity,” in *Proceedings of the 9th International Conference on Intelligent Systems Design and Applications*, 2009.

[4] M. Lukac, R. Tanizawa, and M. Kameyama, “Machine learning based adaptive contour detection using algorithm selection and image splitting,” *Interdisciplinary Information Sciences*, vol. 18, no. 2, pp. 123-134, 2012.

[5] M. Lukac, M. Kameyama, and K. Hiura, “Natural image understanding using algorithm selection and high level feedback,” in *Proceedings of the SPIE, Intelligent Robots and Computer Vision XXX: Algorithms and Techniques*, 2013.

[6] M. Lukac, M. Kameyama, and Y. Fujioka, “VLSI platform for real-world intelligent integrated systems based on algorithm selection,” in *IADIS TPMC*, 2013.

[7] J. Malik, S. Belongie, J. Shi, and T. Leung, “Textons, contours and regions: Cue combination in image segmentation,” in *Proceedings of the International Conference on Computer Vision*, 1999.

[8] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik, “Using contours to detect and localize junctions in natural images,” in *Proceedings of the IEEE Conference on vision and Pattern Recognition, CVPR*, 2008.

[9] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, 2000.

[10] M. Donoser, M. Urschler, and H. Bischof, “Saliency driven total variational segmentation,” in *Proceedings of the 12th International Conference on Computer Vision*, IEEE, 2009, pp. 817-824.

[11] J. Carreira, F. Li, and C. Sminchisescu, “Object recognition by sequential figure-ground ranking,” *International Journal of Computer Vision*, vol. 98, no. 3, pp. 243-262, 2012.

[12] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part based models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, 2010.

[13] L. Ladicky, C. Russell, P. Kohli, and P.H.S. Torr, “Graph cut based inference with co-occurrence statistics,” in *Proceedings of the 11th European Conference on Computer Vision*, ECCV'10, 2010, pp. 239-253.

[14] L. Ladicky, P. Sturgess, K. Alahari, C. Russell, and P. H. S. Torr, “What, where and how many? combining object detectors and crfs,” in *Proceedings of the 11th European Conference on Computer Vision*, ECCV'10, 2010. pp. 424-437.

[15] A. Ion, J. Carreira, and C. Sminchisescu, “Probabilistic joint image segmentation and labeling,” in *Proceedings of the Conference Neural Information Processing Systems*, 2011, pp. 1827-1835.

[16] M. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proceedings of the 8th International Conference on Computer Vision*, July 2001, vol. 2, pp. 416-423.

[17] A. Oliva and A. Torralba, “Modeling the shape of the scene: a holistic representation of the spatial envelope,” *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145-175, 2001.