Context by Region Ancestry

Joseph J. Lim, Pablo Arbeláez, Chunhui Gu, and Jitendra Malik
University of California, Berkeley - Berkeley, CA 94720
{lim,arbelaez,chunhui,malik}@eecs.berkeley.edu

Abstract

*In this paper, we introduce a new approach for modeling visual context. For this purpose, we consider the leaves of a hierarchical segmentation tree as elementary units. Each leaf is described by features of its ancestral set, the regions on the path linking the leaf to the root. We construct region trees by using a high-performance segmentation method. We then learn the importance of different descriptors (*e.g. *color, texture, shape) of the ancestors for classification. We report competitive results on the MSRC segmentation dataset and the MIT scene dataset, showing that region ancestry efficiently encodes information about discriminative parts, objects and scenes.*

1. Introduction

The role of context in visual perception has been studied for a long time in psychology [20, 3, 11]. Visual context can be defined by the scene embedding a particular object [20], and by the semantic relations among different objects [3]. Thus, contextual information for a table is provided by the dining room where it lies, but also by the presence of chairs around it and dishes on its top.

In the computer vision community, despite early attempts as in [26], the importance of context has only recently been widely acknowledged. Its operationalization often relies on the definition of a holistic descriptor of the image. For instance, the seminal work of Oliva and Torralba [19] aimed at capturing the "gist" of the scene. In the case of multi-class segmentation, many recent approaches express relations among objects as pairwise potentials on a probabilistic model [9, 14, 25, 23, 30, 2, 8, 13, 22].

However, contextual cues are naturally encoded through a "partonomy" of the image, the hierarchical representation relating parts to objects and to the scene. In this paper, we argue in favor of using the hierarchy of regions produced by a generic segmentation method in order to model context.

Concretely, given an image, we first construct a region tree using the publicly available hierarchical segmentation algorithm of [1]. The leaves of the tree are the regions of the finest segmentation considered; the root is the entire image and the nodes represent regions ordered by inclusion.

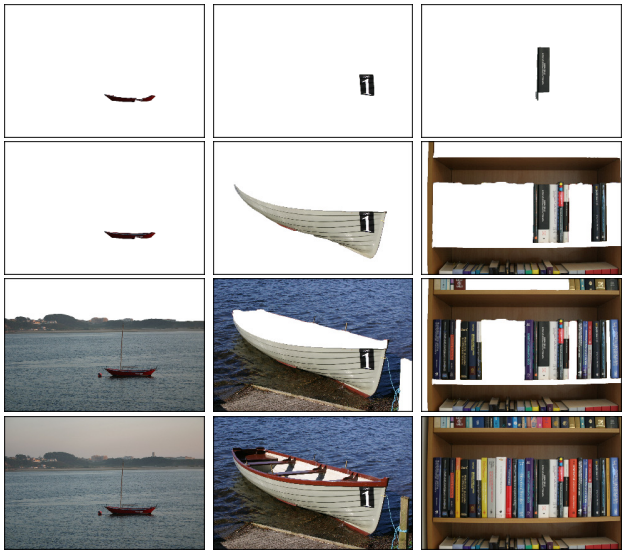

Figure 1. Region ancestry encodes discriminative parts, objects and scenes. Each column shows one leaf of the segmentation tree (top) and three of its ancestors. We measure similarity between leaves by comparing their ancestors and learn their importance on a discriminative framework.

We consider the leaves of the tree as elementary units. Each leaf is represented by features on the set of regions on the path linking it to the root. Borrowing a metaphor from genealogy, we call this set of regions the *ancestral set* of the leaf, with the noteworthy difference that ancestors are in our case *spatial*, rather than temporal.

By progressively enlarging the window of analysis, the ancestral set of a leaf region encodes information at all scales, ranging from local parts of objects, to the whole scene. When making a decision about the category of a leaf or the scene, we expect discriminative features to be different at each level. For instance, in the example of Fig. 1, shape information may be informative for the immediate ancestors of the leaf, but color and texture may be more important at the scene level. In order to take into account such

differences, we define the dissimilarity between leaves as a weighted sum of distances among ancestral features and learn their importance in a discriminative framework.

In order to illustrate the power of region ancestries as a model of context, we address two different tasks: class-specific segmentation, where a category label is assigned to each pixel in the image, and scene classification, where the whole image is labeled according to the type of scene depicted. We report results on the MSRC dataset for the former and on the MIT scene dataset for the latter, obtaining competitive performance on both tasks. Our experiments show that including contextual information modeled by region ancestry provides a significant improvement of 20% in performance on the MSRC dataset, as shown in Table 1.

The rest of the paper is organized as follows. Section 2 reviews previous work. Section 3 introduces the building blocks of our approach. Section 4 describes the learning frameworks considered. Section 5 presents our method for classifying leaves based on ancestry. Section 6 is devoted to experiments. Finally, Section 7 contains some concluding remarks.

2. Related Work

Many recent approaches to multi-class segmentation address the problem using a probabilistic framework and, more specifically, conditional Markov random fields (CRFs) [9, 14, 25, 23, 30, 2, 8, 13, 22]. These methods reason on a graph, where the nodes are usually entities extracted from the image, *e.g.* pixels or patches [9, 25], superpixels [8, 2], or regions from multiple segmentations [13]. Context is in this case understood as relations between entities and modeled by a pairwise potential between nodes. This term is often defined between adjacent entities and is used to enforce spatial consistency on the labels [9, 25]. Other approaches consider a fully connected graph and can therefore model larger-range connections [14, 22, 23]. Some recent methods apply CRFs after an initial estimation of the labels and are therefore able to express more semantic relations. For example, in [23], the pairwise potential captures the co-occurrence of categories in the same image, which is learned, either from the training data, or from external resources. Gould *et al.* [8] introduce a local feature that encodes relative location among categories. Similarly, [14] proposes a two-layer CRF, the first one acting on pixels and the second one modeling relations between categories.

A second type of methods introduces contextual cues by considering a holistic image descriptor or a prior on the categories present in the image. In [24], this prior is provided by a global image classifier. In [21], regions are obtained from multiple segmentations, and contextual information is included by describing each region with features computed on the region mask and on the whole image.

Hierarchical approaches are much less common in the literature. In [28], the image classification task is addressed by matching transitive closures of segmentation trees. Zhu *et al.* [31] do inference on a structure composed by a fixed hierarchy (a quad-tree) and a set of segmentation templates. In [4], features for boosting are constructed from the lower levels of a bottom-up segmentation hierarchy. However, by restricting regions to belong to a single category, contextual information is not fully exploited in this reference.

Context has also been used to improve object detection. In [18], a holistic descriptor is used to narrow the search space of an object detector to a set of likely locations. Heitz and Koller [10] improve the detection of rigid objects by learning spatial relations with amorphous categories. A different approach for modeling context is the work of Hoeim *et al.* [12], who estimate the three dimensional layout of a scene by labeling pixels according to surface orientations. A similar problem is addressed by Sudderth *et al.* [27], by using a hierarchical version of Dirichlet porcesses. Graphical models are also used in [16] in order to infer scene categories.

3. Comparing Leaves by Ancestry

In this section, we first define and describe how we obtain ancestral sets from a segmentation tree and the method to compare them.

We use the segmentation algorithm of [1] that constructs a region tree starting from a contour detector. This method is a fast and parameter-free generic grouping engine. Furthermore, when applied on the output of the high-quality contour detector gPb [17], it significantly outperforms other segmentation approaches, occupying the first place in the Berkeley Segmentation Dataset and Benchmark [6].

The output of the low-level segmenter is an image of weighted boundaries (see Fig. 2). When thresholded, the weighted boundary image produces the segmentation corresponding to a uniform cut in the tree.

We define the **ancestral set**, or **ancestry**, of a region R in a tree $\mathcal{T}$ as the set of regions on the path linking R to the root:

$$A(R) = \{R' \in \mathcal{T} \,|\, R \subseteq R'\} \tag{1}$$

The elementary units of our approach are the leaves of a segmentation tree, often referred as *superpixels*. They are the elements of the finest partition of the image considered. Figure 2 presents an example. Note that the finest partition can be any level in the hierarchy, depending on the task.

We describe each region node in the tree using various features (*e.g.* color, texture, shape) and represent each leaf r by the descriptors of its ancestral set, $F(r)$,

$$F(r) = \{f_1, \ldots, f_M\}, \tag{2}$$

where $M = |A(r)| \cdot Q$, and Q is the total number of features per region.

Figure 2. Example of ancestral set. **From left to right:** • Schematic representation of segmentation tree and example of ancestral set; the leaves are in this case the nodes $\{A, B, C, D, E, F, G, H\}$ and the ancestral set of leaf D is $A(D) = \{D, J, M, O\}$. • Example of ancestral set on a real image. • Original image. • Hierarchical boundaries produced by the segmentation algorithm. • Finest segmentation, containing the leaves of the tree.

In order to measure the similarity between two leaves, we first consider elementary distances between region descriptors of the same type. For simplicity, we use a single notation $d(\cdot, \cdot)$, although they can be different depending on the type of descriptor.

We define the **dissimilarity vector** of a leaf s with respect to a fixed leaf r as:

$$\mathbf{d_r}(s) = [d_r^1(s), \ldots, d_r^M(s)], \tag{3}$$

where:

$$d_r^i(s) = \min_{f_j \in F(s)} d(f_i, f_j), \; i = 1, \ldots, M. \tag{4}$$

and the comparison is done only among features of the same type.

We then define the **dissimilarity** of s with respect to r as a weighted sum of elementary distances:

$$D_r(s) = \sum_{i=1}^{M} w_i \, d_r^i(s) = \langle \mathbf{w}, \mathbf{d_r}(s) \rangle, \tag{5}$$

where $\mathbf{w}$ is the weight vector learned using the framework of the next section. Note that, in general, D is not symmetric, *i.e.*, $D_r(s) \neq D_s(r)$.

4. Learning the Importance of Ancestors

As motivated in the introduction, in an ancestry, descriptors of ancestors in various scales contribute differently to the dissimilarity measures. For instance, the shape descriptor could be the most predominant cue for the body of an aeroplane, whereas its ancestor, the scene of the plane in the sky, could be best described by color. We address this issue by learning a set of weights for the leaf as well as its ancestors from the training data. It is worth noting that the learning framework assigns high weights to the most repeatable regions, making our approach robust to possible errors from the low-level segmenter.

We adopt an exemplar-based matching framework for our purpose. Precisely, given an exemplar leaf r of class $C(r)$, and a set of leaves $\{s_1, \ldots, s_N\}$ of classes $\{C(s_1), \ldots, C(s_N)\}$, all from the training set, first we compute the dissimilarity vectors of s_i with respect to r, $\{\mathbf{d_r}(s_1), \ldots, \mathbf{d_r}(s_m)\}$. We also denote $D_r(s_i)$, the dissimilarity of s_i with respect to r, by $\mathbf{w}^T \mathbf{d_r}(s_i)$.

We introduce and compare three learning approaches in the following, all of which attempt to decrease $D_r(s_i)$ and increase $D_r(s_j)$ for $C(r) = C(s_i)$ and $C(r) \neq C(s_j)$.

4.1. Logistic Regression

The first method to find the weight vector $\mathbf{w}$ is to use a binary logistic classifier to learn directly a mapping from the input vector $\mathbf{d_r}(s)$ to a probabilistic output that measures the probabilities of two leaves having the same class:

$$P(C(s) = c | C(r) = c, \mathbf{w}) = \frac{1}{1 + \exp\{-\mathbf{w^T d_r}(s)\}} \tag{6}$$

Given $\mathbf{d_r}(s_i), i = 1, 2, \ldots, N$, the L_2-regularized large-margin optimization is formulated as follows:

$$\min_{\mathbf{w}} \; \frac{1}{2}\mathbf{w}^T\mathbf{w} + C \sum_{i=1}^{N} \log(1 + \exp(-y_i \mathbf{w^T d_r}(s_i))), \tag{7}$$

where

$$y_i = 2 \cdot \mathbf{1}_{[C(s_i) = C(r)]} - 1, i = 1, 2, \ldots, N \tag{8}$$

Logistic regression is also used in the two other methods below but as a post-processing, in order to normalize a raw

dissimilarity to the range $[0, 1]$. More details will be stated in Section 5.

4.2. Support Vector Machines

A second option is to consider a binary linear support vector machine, using $\mathbf{d_r}(s_i), i = 1, 2, \ldots, N$ as feature vectors and y_i's defined in Equation (8) as binary labels. Compared to the logistic regression classifier, we replace in this case the loss function in (7) for a hinge loss:

$$\min_{\mathbf{w},\xi} \ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \qquad (9)$$

$$s.t.: \ y_i\mathbf{w}^T\mathbf{d_r}(s_i) \geq 1 - \xi_i, \ \xi_i \geq 0, \ i = 1, \ldots, N \qquad (10)$$

4.3. Rank Learning

A third approach for learning the importance of ancestors is the framework of [7]. Unlike the linear-SVM approach that enforces a maximum margin between two sets of data, this learning algorithm enforces maximum margins between pairs of data points from two sets.

Again, given the exemplar r, we consider a second leaf s with the same category as r, and a third leaf t belonging to a different category, we have:

$$D_r(t) > D_r(s) \qquad (11)$$

$$\Rightarrow \quad \langle \mathbf{w}, \mathbf{d_r}(t) \rangle \ > \ \langle \mathbf{w}, \mathbf{d_r}(s) \rangle \qquad (12)$$

$$\Rightarrow \quad \langle \mathbf{w}, \mathbf{x} \rangle \ > \ 0, \qquad (13)$$

where $\mathbf{x} = \mathbf{d_r}(t) - \mathbf{d_r}(s)$.

A set of T such triplets, $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$, is constructed from $\{s_1, \ldots, s_N\}$. The large-margin optimization is then formulated as follows:

$$\min_{\mathbf{w},\xi} \ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{T}\xi_i \qquad (14)$$

$$s.t.: \ \mathbf{w}^T\mathbf{x_i} \geq 1 - \xi_i, \ \xi_i \geq 0, \ \forall i = 1, \ldots, T \qquad (15)$$

$$\mathbf{w} \succeq 0. \qquad (16)$$

5. Leaf Classification

We use the learned dissimilarities between exemplars and test data for leaf classification. When a test leaf is compared to each one of the exemplar leaves, the associated dissimilarities are not directly comparable because they are derived from different learned weights and thus their values may have different ranges. To address this issue, we do a second round of training for each exemplar by fitting a logistic classifier to the binary training labels and dissimilarities, so that dissimilarities are converted to probabilities. We omit this procedure when the weights are learned using the logistic regression because the optimization automatically returns probability outputs that are directly comparable.

To predict the category label for a leaf, we define the confidence score of the leaf s to a category c as the average of the probabilities from the exemplar leaves of that category. That is:

$$Score(s|c) = \frac{1}{|C|}\sum_{j:C(j)=c} p_j(s), \qquad (17)$$

where $p_j(s)$ is the probability of s with respect to exemplar j from the logistic classifier.

The test leaf is assigned to the category label with the largest confidence score. The final segmentation is then obtained by assigning to each pixel the predicted category label of the leaf where it lies.

In practice, in order to make the confidence score more robust to outliers, we compute the average by using only the top $r\%$ of the probabilities, and multiply them by the weight of each class, $m(c)$. Both r and m are learned on the validation set.

In Figure 3, we show the confidence maps for four categories and the final segmentation. Each category's confidence map is obtained by assigning the leaf region with the confidence score.

6. Experiments

6.1. Implementation Details

Since our purpose is to explore the power of our context model, we describe each region in the segmentation tree with standard features.

We represent color by concatenating the marginal histograms in the CIELAB space. Texture is encoded by following the texton approach [15], where filter-bank responses are clustered with the k-means algorithm. In the experiments presented, we consider a codebook of 250 universal textons and 30 bins per color channel.

Shape is encoded by placing an $n \times n$ grid on the bounding box of the region and measuring oriented contour energy on each grid cell. We use both gPb and Sobel filters for this purpose. We then concatenate the responses in each cell to obtain a single shape descriptor. In the results below, we consider 8 orientations and $n = 3$, for a total of 144 dimensions.

Additionally, for the regions on an ancestral set, we consider the normalized coordinates of the leaf's centroid as absolute location features. We compare histogram features using χ^2 as elementary distance and location features using Euclidean distance.

We use the linear SVM and logistic regression implementations of LibLinear [5].

Figure 3. Example of merging the per category confidence maps into a single segmentation. **Top:** Original image, initial partition, ground truth and final segmentation. **Bottom:** Confidence maps for the categories grass, tree, sky and aeroplane. The confidence maps of all the categories are used to produce the final segmentation.

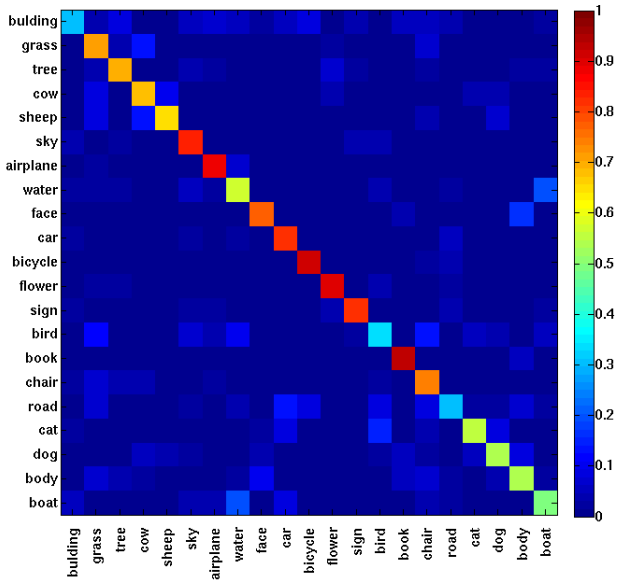

Figure 4. Confusion matrix for MSRC dataset. Average performance is **67%**

6.2. Class-specific Segmentation

We conduct experiments on the MSRC database [25], containing 591 images, with objects from 21 categories. In order to compare our results with other methods, we use the standard split of the database and measure performance by computing the average pixelwise classification accuracy across categories. This metric is preferable to the overall classification accuracy because it does not favor the more common categories.

Figure 4 presents the confusion matrix of our method. Figure 5 presents some qualitative results.

	Leaf Only	Ancestral Set	Improvement
LR	47%	**67%**	**+20%**
SVM	**48%**	62%	+14%
RL	45%	57%	+12%

Table 1. **Results on MSRC**. Comparison of the performance of our method with the different learning approaches considered in Section 4 and by using only the leaves or the full ancestral set. Contextual information provided by the ancestry significantly improves performance in all cases. The best result is obtained with the simple logistic classifier. (LR: Logistic Regression, SVM: Support Vector Machines, RL: Rank Learning)

Table 1 shows the improvement obtained by using the ancestral set instead of the leaves for the learning frameworks considered. The improvement in performance is considerable in all the cases, regardless of the learning method. Figure 5 presents some qualitative results.

Tables 2 and 3 compare our method against recent approaches. It's worth noting that in amorphous and common categories such as grass, tree, sky or road, the inclusion context does not necessarily increase performance with respect to the leaf only model. However, our approach provides a significant improvement over the state-of-the-art in 10 structured object categories where the context is relatively more important (aeroplane, car, bicycle, flower, sign, book, chair, dog, boat).

6.3. Scene classification

We test our system on the MIT scene dataset for the scene classification task. The dataset is composed by 2688 images belonging to 8 scene categories. It is divided in 800

*These results were obtained on different splits of the dataset and are therefore not directly comparable.

Method	[25]	[2]	[21]	[30]	[8]	[24]	[29]	[31]	Ours
Performance	58	55	60	64*	64*	67	68	74	67

Table 2. Comparison of our average performance with other recent methods on MSRC. In MSRC, there are two different evaluation methods and we use the average classification per category.

Method	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat
[25]	62	**98**	86	58	50	83	60	53	74	63	75	63	35	19	92	15	**86**	54	19	62	7
[2]	68	94	84	37	55	68	52	71	47	52	85	69	54	5	85	21	66	16	49	44	32
[21]	68	92	81	58	65	**95**	85	**81**	75	65	68	53	35	23	85	16	83	48	29	48	15
[24]	49	88	79	**97**	**97**	78	82	54	**87**	74	72	74	36	24	**93**	51	78	**75**	35	66	18
[29]	**69**	96	**87**	78	80	**95**	83	67	84	70	79	47	61	30	80	45	78	68	52	**67**	27
LO	34	74	66	49	46	83	56	49	85	34	55	50	44	37	28	16	61	33	35	28	16
AS	30	71	69	68	64	84	**88**	58	77	**82**	**91**	**90**	**82**	**34**	**93**	**74**	31	56	**54**	54	**49**

Table 3. Comparison of our results on each category with other recent methods in MSRC. We obtain the best performance in $10/21$ categories, all of them objects. Results reported for our method correspond to weight learning with logistic regression, using only the leaves (LO) and the ancestral set (AS). [31] is not included in this table, because their confusion matrix is not available.

Method	Building	City	Street	Highway	Coast	Country	Mountain	Forest	Average (%)
[19]	82	**90**	**89**	**87**	79	71	**81**	91	**84**
Ours (LO)	89	27	83	44	**80**	67	48	96	67
Ours (AS)	**93**	81	88	64	77	**79**	80	**96**	82

Table 4. Results on MIT scene dataset. Using the ancestral set provides a significant boost in performance of 15% for this task. Results reported for our method correspond to weight learning with logistic regression, using only the leaves (LO) and the ancestral set (AS).

images for training, with 100 per category, and 1888 images for testing. In this case, instead of assigning a label to each pixel, we assign to the image the label of the confidence map with highest average value.

Table 4 presents the results. We perform comparably to [19]. Also, it is shown that by using ancestral sets, our performance significantly improves.

7. Conclusions

We introduced a simple yet effective approach for modeling contextual information by considering the ancestral sets of leaves in a segmentation tree. Our approach obtains competitive results on both multi-class segmentation and scene classification tasks.

While the main emphasis of this paper is to understand the power of inclusion relationship among regions, other relationships, such as co-occurrence and adjacency, are also important cues for modeling context [23, 8]. Extending our current framework to incorporate these relationships would be an interesting direction for future work.

Acknowledgement

This work was supported by ONR MURI N00014-06-1-0734, and the Haas Scholars Fellowship.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. In *CVPR*, 2009.

[2] D. Batra, R. Sukthankar, and T. Chen. Learning class-specific affinities for image labelling. In *CVPR*, 2008.

[3] I. Biederman, R. Mezzanotte, and J. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14:143–177, 1982.

[4] J. Corso. Discriminative Modeling by Boosting on Multilevel Aggregates. In *CVPR*, 2008.

[5] R. Fan, K. Chang, and X. L. C. Hsieh, C. Wang. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[6] C. Fowlkes, D. Martin, and J. Malik. The Berkeley Segmentation Dataset and Benchmark (BSDS). www.cs.berkeley.edu/projects/vision/grouping/segbench/.

[7] A. Frome, Y. Singer, and J. Malik. Image retrieval and classification using local distance functions. In *NIPS*, 2006.

[8] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. *International Journal of Computer Vision*, 80(3), December 2008.

[9] X. He, R. Zemel, and M. Carreira Perpinan. Multiscale conditional random fields for image labeling. In *CVPR*, pages II: 695–702, 2004.

[10] G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. In *ECCV*, 2008.

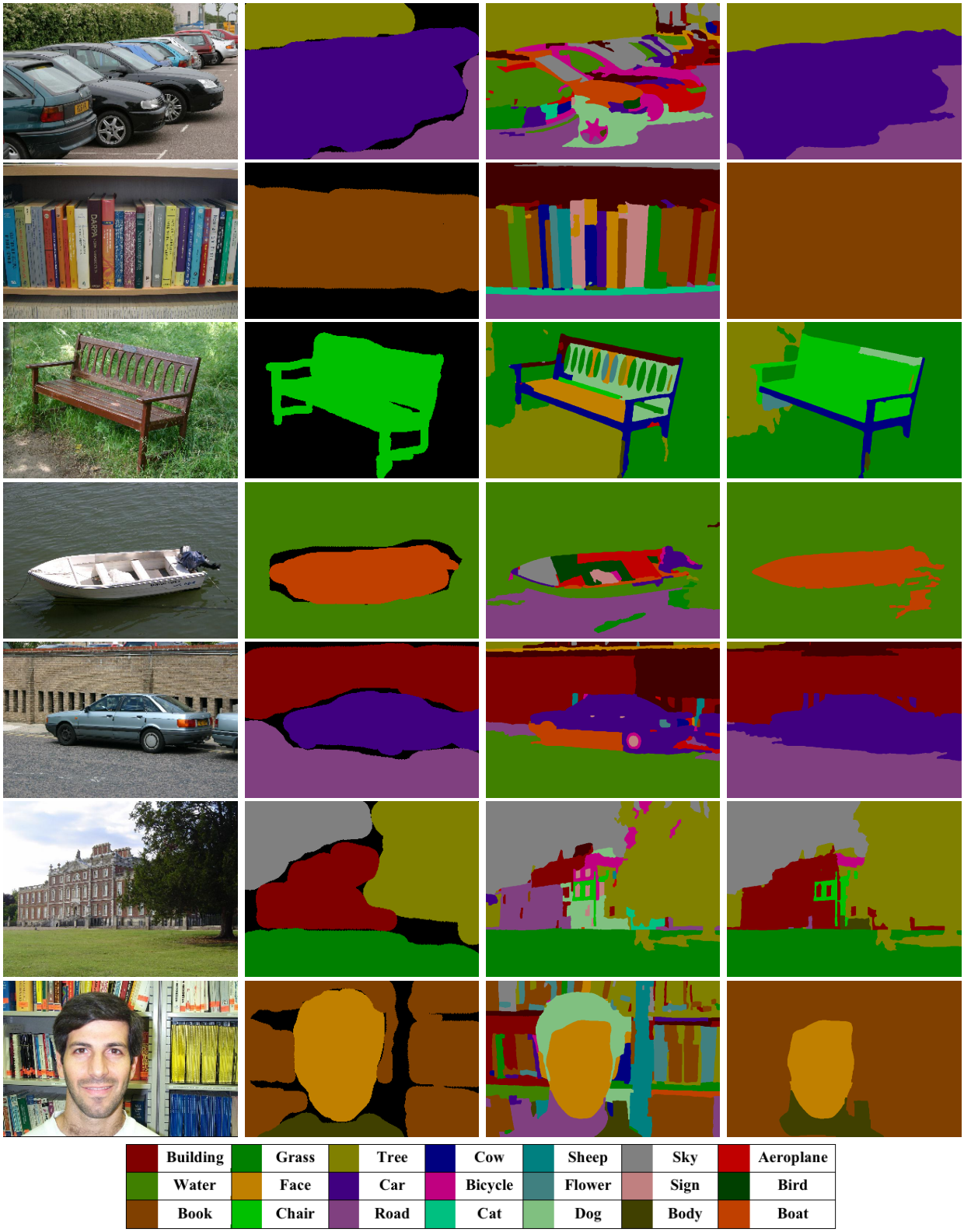

Figure 5. Example of results on MSRC dataset. **From left to right:** • Original image. • Ground truth segmentation. • Results of our method using only the leaves. • Results of our method using the ancestral set.

[11] J. Henderson and A. Hollingworth. High-level scene perception. *Annual Review of Psychology*, 50:243271, 1999.

[12] D. Hoiem, A. Efros, and M. Hebert. Geometric context from a single image. In *ICCV*, pages I: 654–661, 2005.

[13] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.

[14] S. Kumar and M. Hebert. A hierarchical field framework for unified context-based classification. In *ICCV*, pages II: 1284–1291, 2005.

[15] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44, June 2001.

[16] F. Li and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, pages II: 524–531, 2005.

[17] M. Maire, P. Arbelaez, C. Fowlkes, and M. Malik. Using contours to detect and localize junctions in natural images. In *CVPR*, 2008.

[18] K. Murphy, A. Torralba, and W. Freeman. Using the forest to see the trees: A graphical model relating features, objects, and scenes. In *NIPS*, 2003.

[19] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, May 2001.

[20] S. Palmer. The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 3:519–526, 1975.

[21] C. Pantofaru, C. Schmid, and M. Hebert. Object recognition by integrating multiple image segmentations. In *ECCV*, pages III: 481–494, 2008.

[22] D. Parikh, L. Zitnick, and T. Chen. From appearance to context-based recognition: Dense labeling in small images. In *CVPR*, 2008.

[23] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *ICCV*, 2007.

[24] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR*, 2008.

[25] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, 2006.

[26] T. Strat and M. Fischler. Context-based vision: Recognizing objects using information from both 2-d and 3-d imagery. *IEEE Trans. on PAMI*, 13(10):1050–1065, October 1991.

[27] E. Sudderth, A. Torralba, W. Freeman, and A. Willsky. Depth from familiar objects: A hierarchical model for 3d scenes. In *CVPR*, 2006.

[28] S. Todorovic and N. Ahuja Learning subcategory relevances to category recognition. In *CVPR*, 2008.

[29] Z. Tu. Auto-context and Its application to High-level Vision Tasks. In *CVPR*, 2008.

[30] J. Verbeek and B. Triggs. Region classification with markov field aspect models. In *CVPR*, 2007.

[31] L. Zhu, Y. Chen, Y. Lin, C. Lin and A. Yuille. Recursive Segmentation and Recognition Templates for 2D Parsing. In *NIPS*, 2008.