

Evaluation of Image Segmentation Based on Histograms

Andrej FOGELTON*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
fogelton@fiit.stuba.sk*

Abstract. The presented paper evaluates an experiment on image class segmentation based on Hue-Saturation histograms. Training is based on histogram calculation for every object class separately. Sliding window is performed to segment (label) individual pixels of the evaluation image. Sliding window around the given pixel encloses the local appearance from which the histogram is calculated. Local appearance histogram is subsequently used to be compared with the precomputed class histograms. Given pixel is labeled according to the best match using the intersection method. We show how this method depends on data character and window size. Unfortunately, this algorithm suffers from both quality and speed performance.

1 Introduction

Object recognition is the dreamed future of computer vision algorithms. The state of the art algorithms performance is not sufficient neither in speed nor in precision performance. The idea is to develop an algorithm which would be similar to the human vision system. Human vision combines different recognition cues such as object color, shape (contour, 3D model), semantics and object's abilities with possible appearance changes. It works with scale and time precision. Firstly, object class is recognized (e.g. tree) and then the exact type of the class (lime tree) by other discriminative features (the shape of the leaves).

We believe that the important step is proper object class recognition algorithm in order to imitate human vision system. There are several algorithms using different cues and classification methods. Object class recognition based on contour fragments can also provide reliable results [5]. *TextonBoost* presents a new kind of feature, which stores appearance, shape and context cue data [7] with boosting classification algorithm. Semantic Texton Forest shows a very good performance of randomized forest classifier using *Texton* features (color based) for image pixel segmentation [6]. A very good review of different object recognition methods can be found in [4]. Conditional Random Field (CRF) is often used for precision improvements of the classification methods for image pixel labeling [3].

* Doctoral study programme in field: Applied Informatics

Supervisor: Dr. Wanda Benešová, Institute of Applied Informatics, Faculty of Informatics and Information Technologies STU in Bratislava

2 Image Segmentation

The idea is to experiment on solving object class recognition problem using only one cue – color represented by histogram. Histograms are experimentally used for object recognition in [2], while tested with different color models on luminance changes. Our idea is to use manually segmented images to create object class histograms. We will calculate the histogram from the local appearance on the evaluation image and compare it with precomputed object class histograms to get the most probable label (object class of the given pixel).

The HSV color model is chosen to achieve partial luminance invariance by dropping the *Value* channel [1]. Histogram is calculated using only *Hue* and *Saturation* channels with the common bin configuration, 30 bins to represent Hue and 32 bins to Saturation [1].

MSCR-21 dataset

MSCR-21 dataset¹ is very often used to refer and compare segmentation results [6, 7]. It consists of 591 weakly labeled images of 21 object categories (tree, car, horse, flower, etc.). In Figure 1 we can see that weakly segmented images can also consist of not specified regions (void class).



Figure 1. Image and his ground truth (weakly labeled segmentation) from MSCR-21 dataset. Class labels are added manually for better understanding of different gray levels. White color represents void class (not specified).

Algorithm

Before performing the learning procedure (training), the dataset needs to be split into training and evaluation set. This is done along the interested class occurrence over the images. Half of the images of given class (e.g. flowers) are used for training and the other half for evaluation. Learning consists of histogram calculations for every object class separately. For each pixel in every training image, algorithm checks the assigned label from the ground truth image to distinguish in which histogram will the bin value be given by the pixel values of Hue and Saturation incremented. Image pixels segmented in the ground truth images as void class are not used for histogram calculations.

Normalization of these histograms is needed, because the values in the histograms strongly depend on the amount of training images (pixels) of given class occurrence. To be able to get proper comparison results, we normalize the sum of histogram bins to the number of pixels in the sliding window (e.g. for window size 15×15 it is 225). Sliding window of the given size is performed on the evaluation image. The evaluating pixel is the middle one of the sliding window. Sliding window method results in a smaller resulted image than the original, exactly half of the window size from each image side.

¹ <http://research.microsoft.com/en-us/projects/ObjectClassRecognition/>

Intersection histogram comparison

For every pixel a histogram from the window around the given pixel is calculated. Intersection comparison method is subsequently used to compare the sliding window histogram with class histograms. Equation 1 represents the intersection method² as the sum of the minimum bin values over the compared histograms H_1, H_2 . Higher distance d means better histogram match. Pixel label is assigned based on the maximum value of histogram comparisons.

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I)) \quad (1)$$

The overview of the entire algorithm is listed in Listing 1.

Listing 1. Algorithm overview: Image segmentation based on object class.

```

1 Compute object class histograms from images using their ground truth
  information
2 Normalize histograms
3 For each image to segment do
4   For each pixel of given image do
5     Compute histogram around the given pixel within a sliding window
6     Compare computed histogram with the object class histograms
      using the intersection method
7     Pixel label (color segmentation) is chosen based on the best
      histogram match

```

3 Evaluation

As already mentioned, half of the MSCR-21 dataset images are used for the evaluation of the given method. Table 1 presents achieved results. The average correct pixel segmentation value represents the percentage of correctly labeled pixels on evaluation images in total. The segmentation percentage per class represents the number of correctly labeled pixels of the given class to all pixels belonging to the given class. Both values are important to get a sense of accuracy of the presented system, because we can distinguish from the influence of classes occurrences. In our case, high percentage of grass occurrence increases the average correct pixel percentage.

The results suffer in quality, but few classes achieve good performance, mostly sky, grass, road and face. From this we can conclude that the given algorithm is good for classes of objects which are in monotone color. There is no red sky, only blue or the grass is only green. Their colors do not differ as much as between the other classes (e.g. there are different colors of cars). If the color is not the most specific cue of the given object, it is proven that it is not suitable to be used as the main cue for object recognition. The different size of the sliding window does not influence the segmentation results much. From the values presented in Table 1 and the analyze of the images we can deduce, that the window size influences mostly the percentage of smaller objects, classes like book, body, sheep, cow, etc., which decreases with larger window size.

We can also see in Figure 2, that the image segmentation quality – precision performance is very low. It is interesting to see how the sliding window influences pixel labeling around the object borders. Pixels around the border are assigned by labels of the most dominant histograms as we can see on the flower in the presented image. The flower leaves are thinner in the segmented image compared to the original (Figure 2). This is because the grass histogram is much stronger at green colors than the flower histogram at pink colors (the presented flower is pink in color).

² <http://opencv.itseez.com>

Table 1. Percentage of the correctly labeled pixels of given categories using different sliding window size.

object class	11 × 11	15 × 15	19 × 19
Building	0.44%	0.29%	2.46%
Grass	52.69%	53.01%	55.23%
Tree	9.00%	8.39%	0.88%
Cow	2.24%	2.20%	1.97%
Sheep	2.21%	2.44%	0.88%
Sky	71.77%	59.34%	75.19%
Airplane	2.23%	1.90%	1.39%
Water	8.73%	8.55%	5.66%
Face	24.47%	27.15%	23.40%
Car	1.01%	1.72%	0.14%
Bicycle	0.88%	1.31%	0.05%
Flower	6.68%	6.35%	2.60%
Sign	4.35%	12.69%	4.58%
Bird	0.07%	0.12%	0.00%
Book	4.58%	3.53%	1.84%
Chair	1.99%	2.00%	0.32%
Road	31.72%	36.48%	30.63%
Cat	0.23%	0.37%	0.00%
Dog	5.35%	4.75%	5.11%
Body	2.47%	1.86%	0.00%
Boat	3.75%	4.11%	0.31%
Average per category	11.28%	11.36%	11.13%
Correctly labeled pixels in total	17.53%	16.49%	17.66%

Training performance is excellent, it takes only about 5 seconds to train 21 class histograms from approximately 300 images. On the other side, evaluation takes much more, about 2 minutes for every image. Because of the processing time and segmentation quality performance, this method is not suitable for processing large datasets. It can be used for specific purpose, when the object colors within a class do not differ much and when the processing time is not very relevant.

To compare, one of the state of the art method [6] performs recognition and image segmentation in 0.1 second but with learning time about 15 minutes, achieving average precision about 64% while using multiple cues. These results are much above ours.

4 Conclusions

This paper presents an experiment of object recognition and segmentation on MSCR-21 dataset. It uses just one cue – color represented by the Hue-Saturation histogram. Precomputed histograms of object classes are compared to the histogram of local region (sliding window) in the given image using the intersection method. To the center pixel of the given region, label of the most probable class is assigned. Results are good for classes of objects which do not differ in color much like grass, sky, road. For other classes the results are quite insufficient. The average correct recognition rate per class is about 11% and the average per pixel labeling rate in total about 17%.

Acknowledgement: This work was supported by KEGA 068UK-4/2011.

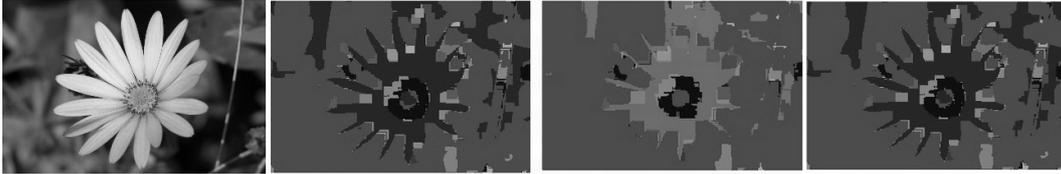


Figure 2. Original image, segmentation results using 11×11 , 15×15 , 19×19 sliding window.

References

- [1] Bradski, D.G.R., Kaehler, A.: *Learning opencv, 1st edition*. O'Reilly Media, Inc., 2008.
- [2] Gevers, T., Smeulders, A.: Color based object recognition. In Del Bimbo, A., ed.: *Image Analysis and Processing*. Volume 1310 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 1997, pp. 319–326.
- [3] Ladicky and, L., Russell, C., Kohli, P., Torr, P.: Associative hierarchical CRFs for object class image segmentation. In: *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 739–746.
- [4] Roberto Cipolla, S.B., Farinella, G.M., eds.: *Computer Vision - Detection, Recognition and Reconstruction*. Volume 285 of *Studies in Computational Intelligence*. Springer Berlin / Heidelberg, 2010.
- [5] Shotton, J., Blake, A., Cipolla, R.: Multiscale Categorical Object Recognition Using Contour Fragments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2008, vol. 30, no. 7, pp. 1270–1281.
- [6] Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008.
- [7] Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation. In Leonardis, A., Bischof, H., Pinz, A., eds.: *Computer Vision – ECCV 2006*. Volume 3951 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2006, pp. 1–15.