ENSEMBLE EVALUATION FOR IMAGE SEGMENTATION USING A SEMI-SUPERVISED SVM

Sang Jun Lee and Sang-Gyu Ryu
Department of Electrical Engineering
POSTECH
pohang, gyeongsangbuk-do, Korea
email: {lsj4u0208,sgryu}@postech.ac.kr

Yong-Ju Jeon and Doo-chul Choi
Department of Electrical Engineering
POSTECH
pohang, gyeongsangbuk-do, Korea
email: {jjeon83,dooshe}@postech.ac.kr

Sang Woo Kim
Department of Creative IT Excellence Engineering
POSTECH
pohang, gyeongsangbuk-do, Korea
email: swkim@postech.ac.kr

ABSTRACT
A new unsupervised ensemble evaluation algorithm is proposed for image segmentation. A semi-supervised support vector machine is used to combine the existing unsupervised evaluators. We also proposed feature extraction and data selection procedures to enhance the overall performance. We experimentally demonstrated that our proposed algorithm is superior to existing segmentation evaluation measures.

KEY WORDS
Segmentation and Representation, Segmentation Evaluation, Pattern Recognition, Support Vector Machine

1 Introduction
Image segmentation is a process that partitions an image into meaningful segments. Because image segmentation is often used as a preprocessing step in the construction of larger systems, the result of this task critically affects subsequent processes. Therefore, it is one of the most important areas in the field of image analysis, and its applications include object recognition[1][2], medical image processing[3], and video surveillance systems[4]. Many segmentation methods have been developed thus far, including EDISON[5], CTM segmentor[6], and UCM segmentor[7]. However, comparing the performances of existing evaluation methods continues to be a difficult task. Thus, more accurate algorithms are needed for measuring and comparing existing segmentation algorithms and for parameterizing segmentation algorithms for specific applications.

Segmentation evaluation algorithms are classified into two approaches based on the necessity of a benchmark. One is the supervised segmentation evaluation that scores a segmented image by comparing the image with a benchmark. Because it needs benchmark images which are manually partitioned by human, this method is known as a human-added method. The other approach is an unsupervised evaluation method, which does not rely on a benchmark. Because it evaluates a segmented image using only its own statistics, it is also known as a stand-alone evaluation method.

In research, supervised segmentation evaluation methods have received more attention, and they generally yield superior results. However, the supervised approach requires tedious work for the generation of a set of benchmark images, so this approach is not feasible in some applications. In contrast, unsupervised approaches do not require benchmark images and can be applied in real-time systems[8].

Most segmentation evaluation algorithms are application-specific[9]. In other words, an evaluation method yields a good result in some cases, but yields insufficient results in the other cases. To solve this problem, we propose an unsupervised ensemble evaluation algorithm that combines the existing segmentation evaluation measures using machine learning. We used a semi-supervised Support Vector Machine (SVM) to use the information of unlabeled data. Furthermore, we introduced a 16-dimensional feature vector and data selection methods to improve the training process. Several experiments were conducted using the Berkeley-segmentation database, which is a well-known database of 500 images, and our program code is available on the Internet1.

This paper discusses three main concepts: (1) the application of a semi-supervised SVM, (2) feature extraction, and (3) the procedure for data selection. In Section 2, we review previous works related to image segmentation evaluation and SVM. Section 3 describes the proposed algorithm and provides a detailed explanation of the feature extraction method and data selection procedure. Experimental results and analysis are presented in Section 4, and finally, we give a conclusion in Section 5.

1\141.223.87.129\SourceCode (ID: paper, PW: 0000)
2 Previous Work

2.1 Image Segmentation Evaluation

Several unsupervised segmentation evaluation algorithms have been proposed, such as Zeboudj’s contrast, Levine and Nazif’s inter-region contrast and intra-region uniformity, Rosenberger’s criterion, Borsotti’s criterion, and Roman-Roldan’s criterion. Most evaluation methods have been formulated by following three fundamentals: intra-class uniformity, inter-class disparity, and semantic cue.

Intra-class uniformity measures the similarity of pixel values in a segment. This measure can be constructed using the standard variation or the entropy of a segment. Inter-class disparity, on the other hand, captures the differences between segments in a segmented image, e.g., average color difference between segments. Further, semantic cue measures the shape of a segment. Elongation or circulation can be analyzed to obtain semantic information, but this measure is known to be highly dependent on the types of images used.

Trials to combine the existing evaluation methods have been conducted in the past. Zhang proposed several co-evaluation frames using weighted sums, a Bayesian strategy, and SVM. Furthermore, Chabrier combined three evaluation methods with several parameters using SVM.

2.2 Support Vector Machine

The field of machine learning is divided into three subfields: supervised learning, unsupervised learning, and reinforcement learning. In brief, supervised learning estimates a black box using a labeled training set \( \chi = \{(x_t, y_t)\}_{t=1}^N \), and regression and classification problems are examples of supervised learning. On the other hand, unsupervised learning methods observe and analyze the unlabeled data \( \{x_t\}_{t=1}^N \). The applications of unsupervised learning include clustering and dimensionality reduction. Semi-supervised learning lies between supervised and unsupervised learning and takes advantage of both learning methodologies.

SVMs, introduced by Vapnik, estimate a hypothesis that classifies the given training data within one of two classes. SVM finds a decision function (hypothesis), which maximizes the margin between two classes with minimum empirical error. Theoretically, this approach is based on the principle of statistical risk minimization and works well when sufficient training data is available.

Later, Bennett introduced semi-supervised SVM, which finds a decision function according to the principle of overall risk minimization with both labeled and unlabeled data. He showed that semi-supervised SVM improves the overall performance and performs well especially when training data are insufficient.

3 Proposed Algorithm

The proposed algorithm gives visually better segmented image among two segmented images corresponding to an original image. The inputs of proposed algorithm are an original image and its pair of segmented images. These pairs are generated by two different segmentation algorithms (or one segmentation algorithm with two different parameters). The output is a label, and this label signifies which is visually better between two segmented images. The structure for the proposed algorithm is given below.

![Algorithm](Figure 1. Algorithm)

\( OI \) is an original image, and \( (S1, S2) \) implies a pair of segmented images. \( SEa(\text{Levine and Nazif’s intra-region uniformity}) \), \( SEb(\text{Levine and Nazif’s inter-region contrast}) \), and \( SEc(\text{Roman-Roldan’s criterion}) \) are the base evaluators (existing evaluation algorithms) we used. The output of \( \text{Feature Extractor}(f) \) is a 16-dimensional vector, which characterizes original and segmented images. Using scores of base evaluators and feature vector, good segmented image is decided by a semi-supervised SVM. This proposed algorithm is called unsupervised evaluation algorithm, because this algorithm needs not benchmark images for evaluating process. A detailed explanation will be explained in the following subsections.

3.1 Semi-Supervised SVM

We adopted a semi-supervised learning strategy for combining the existing evaluation criteria with considerations for unlabeled data. Evaluating a segmentation criterion is achieved by measuring the similarity between the result of an evaluation algorithm and the decisions of human evaluators. This process can be viewed as a classification problem, which classifies each pair of segmented images into good (+1) and bad (-1) classes. Here, pairs of segmented images with their labels are needed for the training process. However, labeling all the data is not easy, and it is difficult to decide which segmented image is better in the case of some pairs of images, because both images may possess similar qualities or both may be low-quality images. Moreover, the labels of ambiguous pairs may include personal bias, and therefore these labels are not reliable. On the other hand, neglecting these ambiguous data may cause...
The gradient image is obtained using the following formula:

\[ t = \text{intensity means of the original image and its gradient image.} \]

six features are used for each segmented image.

First, we convert the red, green, blue (RGB) color space into the hue, saturation, value (HSV) color space, and construct the 16-dimensional feature vector. Two features are used for the hue and value space of the original image, and six features are used for each segmented image.

The first two features for an original image are intensity means of the original image and its gradient image. The gradient image is obtained using the following formula:

\[
g(s_x, s_y) = \{(f(s_x, s_y) - f(s_x, s_y - 1))^2 \\
+ (f(s_x, s_y) - f(s_x - 1, s_y))^2\}^{\frac{1}{2}}
\]

Here, an original image is used in the hue and value spaces, and \((s_x, s_y)\) is a point on the image.

Features for two segmented images are based on the following six measures: the under segmentation features \((US_H, US_V)\), number of segments \((N_s)\), total length of the boundary \((L)\), sum of the standard deviations obtained from the pixel values in the corresponding segmented regions \((\sigma_R)\), and sum of the standard deviations for the pixel numbers of each segments \((\sigma_N)\). The detailed descriptions of these measures are as follows.

1. \(US_H\)

\[
US_H = \sum_{s \in f} \{1 - d(s)\} \times g(s),
\]

where \(d(s) = \min_{b \in B} \|s - b\|_1\), and \(d(s) = \frac{d(s)}{\max_{t \in f} d(t)}\).

Here, \(s\) and \(b\) are points on the \(xy\)-coordinate, \((s_x, s_y)\) and \((b_x, b_y)\), respectively.

\(US_H\) is measured using the hue space of the original image. First, a segmented image is converted into a boundary image(B), whose pixel values are zero except at the boundary points, and then \(d(s)\), which is a normalized version of the minimum distance between the boundary points and given point \(s\), is computed.

A segmented image with many segments(boundary pixels) has a large value for \(\{1 - d(s)\}\). Therefore, \(\{1 - d(s)\}\) of an over-segmented image has larger than that of an under-segmented image. Moreover, by multiplying the value of the gradient image, if the boundary of a segmented image is well-matched to the boundary of an object in the original image then \(US_H\) becomes bigger than \(US_H\) of an under-segmented image or that of a low-quality segmented image. Therefore a low-value for \(\{1 - d(s)\}\) implies that the input segmented image is either under-segmented or of low-quality.

2. \(US_V\) is measured using the value space of the original image, and the other computations are the same as that for \(US_H\).

3. \(N_s\) represents the number of segments(regions), and it detects severely under- or over-segmented images.

4. \(L\) represents the total length of the boundary in a segmented image, and it is obtained from the boundary image(B).

\[
L = \text{card}(\{b \mid b = 1, b \in B\})
\]

Here, \(\text{card}\) denotes a cardinal number. (i.e., \(L\) = the number of boundary points)

5. \(\sigma_R\)

\[
\sigma_R = \sum_{i=1}^{K} \text{std}(R_i),
\]

where \(\text{std}(R_i)\) is the standard deviation of the pixel values in the \(i^{th}\) segment(region). Here, we used the components of V-space in the HSV space for the pixel values.

6. \(\sigma_N\)

\[
N_i = \text{card}(\{s \mid s \in R_i\})
\]

\[
\sigma_N = \text{std}(N_i),
\]

3.2 Feature Extraction

First, we convert the red, green, blue (RGB) color space into the hue, saturation, value (HSV) color space, and construct the 16-dimensional feature vector. Two features are used for the hue and value space of the original image, and six features are used for each segmented image.

The first two features for an original image are intensity means of the original image and its gradient image. The gradient image is obtained using the following formula:

\[
g(s_x, s_y) = \{(f(s_x, s_y) - f(s_x, s_y - 1))^2 \\
+ (f(s_x, s_y) - f(s_x - 1, s_y))^2\}^{\frac{1}{2}}
\]

Here, an original image is used in the hue and value spaces, and \((s_x, s_y)\) is a point on the image.

Features for two segmented images are based on the following six measures: the under segmentation features \((US_H, US_V)\), number of segments \((N_s)\), total length of the boundary \((L)\), sum of the standard deviations obtained from the pixel values in the corresponding segmented regions \((\sigma_R)\), and sum of the standard deviations for the pixel numbers of each segments \((\sigma_N)\). The detailed descriptions of these measures are as follows.

1. \(US_H\)

\[
US_H = \sum_{s \in f} \{1 - d(s)\} \times g(s),
\]

where \(d(s) = \min_{b \in B} \|s - b\|_1\), and \(d(s) = \frac{d(s)}{\max_{t \in f} d(t)}\).

In the proposed algorithm, the labeling process is conducted for the training set, except for the ambiguous data. Here, ambiguous data implies pairs of segmented images wherein it is difficult to decide the better segmentation result. Two major examples of ambiguous data are pairs of
two similar images and low-quality images. Labeling ambiguous data may generate unreliable training data. However, excluding these data may severely affect the result of the classifier because SVM uses only support vectors to find a decision function. Our proposed algorithm does not label ambiguous data and these data are used as unlabeled data.

Data selection is a filtering process to find unreliably labeled data. The algorithmic process for finding ambiguous data is as follows.

1. Train a semi-supervised SVM several times, and compute the standard deviation of the differences between the validation result and the pre-assigned label.
2. Select labeled data that have standard deviations higher than the threshold value.
3. Erase the labels of the selected labeled data, and use them as unlabeled data.

Here, high standard deviations imply that such data are not classified into one class in most iterations, and hence, their labels may be unreliable. Therefore, these data are used as unlabeled data. The above process is iterated with different training and validation sets, and the number of iterations are sufficient to allow for the use of all the data as a validation set.

4 Experimental Result

One approach for evaluating an evaluation algorithm is comparing the results of algorithms to a human segmentor’s decision. A detailed evaluating procedure is as follows.

1. Pairs of segmented images are evaluated using an unsupervised segmentation measure.
2. The result of (1) is compared with a human’s decision.
3. The accuracy of this comparison is measured.

We conducted experiments in two manners. In the first experiment, benchmark images, which were segmented manually, are used. These benchmark images are considered as ground-truth. The second experiment was conducted with two sets of segmented images, which were generated from two segmentation algorithms or a segmentation algorithm with two sets of parameters. In addition, we conducted an experiment with several test ratios. This experiment shows the overall performance of the proposed algorithm with insufficient training sets.

The results of our algorithm are compared to the results of base evaluators and previous ensemble frameworks that use SVM for combining three evaluation criteria with several parameters[15]. The last experiment shows the accuracy of previous and proposed algorithms as the test ratio is varied.

4.1 Experimental Environment

All the images and benchmarks used in these experiments are from the Berkeley Segmentation Dataset and Benchmark, which is available for free on the Internet. ² We generated four sets of segmented images using the CTM segmentor³, EDISON(with two sets of parameters)⁴, and

---

²http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bbsd/
³http://www.eecs.berkeley.edu/ yang/software/lossy segmentation/
⁴http://coewww.rutgers.edu/riul/research/code/EDISON/index.html
the UCM segmentor\(^5\).

We used three segmentation evaluation algorithms, called base evaluators, including Levine and Nazif’s inter-region contrast, Levin and Nazif’s intra-region uniformity, and Roman-Roldan’s criterion. A convex optimization tool CVX, which was implemented by Boyd, was used to implement the semi-supervised SVM[18].

The performance of the evaluation criterion is measured on the basis of balanced accuracy, and this measure is defined as follows:\(^6\)

\[
\text{Balanced accuracy} = \frac{1}{2} \left( \text{sensitivity} + \text{specificity} \right) = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)
\]

We divided our data randomly: 30% into the training data set and 70% as the test data. Each experiment was iterated 200 times with different training and test sets.

### 4.2 Benchmark versus segmented image

In the first experiment, we evaluated the segmentation criteria by comparing results of the benchmarks and segmented images. This experiment was conducted with only the original SVM with the proposed features, because comparing the benchmarks and their corresponding segmented images is trivial. The result was compared to the results of base evaluators and previous ensemble frameworks.

Three experiments were conducted using benchmark images and the result of three types of images segmentations: the CTM segmentation, the EDISON, and the UCM segmentation. (Figure 3)

![Figure 3. Comparing Segmented Image with its Benchmark](image)

4.3 Segmented image versus segmented image

In the second experiment, two sets of segmented images were used. These segmented images were generated by two different segmentation algorithms or a segmentation algorithm with two different parameters. We present the performance of the SVM with a feature vector, the semi-supervised SVM, and the semi-supervised SVM with data selection.

Four experiments were conducted using the following pairs of segmented images: CTM segmentor versus EDISON, CTM segmentor versus UCM, EDISON versus EDISON2, and EDISON versus UCM. Here, EDISON and EDISON2 imply a segmentation algorithm with two parameters. (Figure 4)

![Figure 4. Compareion of two different segmented images](image)

4.4 Test ratio

The last experiment was conducted with several test ratios. We used two pairs of segmented images: CTM segmentor versus EDISON and CTM segmentor versus UCM. The following two graphs show that our proposed algorithm generally yields a better accuracy than the previous algorithms which use a SVM with the result of base evaluators.

In particular, the gap between the accuracy of a semi-supervised SVM with data selection and general SVM is large with a relatively large test set. A large test ratio implies that the SVM was trained with a small amount of training data. This means that the proposed algorithm is

---

\(^5\)http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html

\(^6\)TP: true positive, TN: true negative, FP: false positive, FN: false negative
more reliable than previous algorithms in the case of insufficient training data.

Figure 5. Accuracy with varying test ratio (top: CTM segmentor vs. EDISON, bottom: CTM segmentor vs. UCM)

5 Conclusion

In this paper, we proposed an unsupervised ensemble evaluation algorithm for image segmentation. The key idea of this study is to combine base evaluators using labeled and unlabeled data. Labeled data are generated for strictly different pairs of segmented images, while unlabeled data are generated for ambiguous pairs. We viewed the unsupervised evaluation as a classification problem and adopted a semi-supervised SVM. This paper also describes a feature extraction method and a data selection process.

The proposed algorithm evaluates the goodness of the two segmentation algorithms by combining three unsupervised algorithms. This algorithm can be expanded easily for comparing more base segmentation algorithms by extracting their feature values and evaluation scores for additional segmented images. Also, this algorithm can be applied to combine more evaluation algorithms by considering additional score values as additional components of feature vector. These generalization of proposed algorithm can be used to set the parameters automatically by combining an evaluation algorithm with several parameters. Although this generalization yields high dimensional input of SVM, it dose not severely affect the computation cost of SVM because SVM decides a decision function using just inner product of input.

This algorithm has two limitations. First, the proposed algorithm is a comparative evaluation algorithm in contrast to general quantitative supervised evaluation algorithm. However, we can choose the best segmentation algorithm by comparing algorithms available for a specific application. The second limitation is the computation cost for data selection process. However, this training computation time is not critical for the offline system.

In a real work environment, choosing an efficient segmentation evaluation measure is not easy because most algorithms are application-specific. Furthermore, in a small workspace, some resultant labels may not be reliable because of the small number of human evaluators. From this point of view, using a semi-supervised learning methodology is a more reasonable strategy to deal with ambiguous data. Experimental results showed that our proposed algorithm is superior to previous evaluation algorithms.

References


