

Segmentation Ensemble

A Knowledge Reuse for Model Order Selection using Case-based Reasoning

Pakaket Wattuya¹ and Ekkawut Rojsattarat²

¹*Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand*

²*Department of Computer Science, Faculty of Science at Si-Racha, Kasetsart University Si-Racha Campus, Chonburi, Thailand*

Keywords: Image Segmentation Ensemble, Case-based Reasoning, Model Order Selection.

Abstract: Cluster ensemble has emerged as a powerful technique for improving robustness, stability, and accuracy of clustering solutions, however, automatic estimating the appropriate number of clusters in the final combined results remains unsolved. In this paper we present a new approach based on a case-based reasoning to handle this difficult task. The key success of our approach is a novel use of cluster ensemble in a different role from the past. Each ensemble component is viewed as an expert domain for building a case base. Having benefited from the information extracted from cluster ensemble, a case-based reasoning is able to settle efficiently the appropriate number of clusters underlying a clustering ensemble. Our approach is simple, fast and effective. Three simulations with different state-of-the-art segmentation algorithms are presented to illustrate the efficacy of the proposed approach. We extensively evaluate our approach on a large dataset in comparison with recent approaches for determining the number of regions in segmentation combination framework. Experiments demonstrate that our approach can substantially reduce computational time required by the existing methods, more importantly, without the loss of segmentation combination accuracy. This contribution makes the segmentation ensemble combination concept more feasible in real-world applications.

1 INTRODUCTION

Cluster ensemble has emerged as a powerful technique for dealing with several difficulties in clustering problems. Thus far, it has been used for dealing with 1) instability of clustering algorithms (Topchy et al., 2005; Franek et al., 2012); 2) sensitivity to noise, outliers or sample variations (Nguyen and Caruana, 2007); and 3) inaccuracy of individual clustering algorithms (Vega-pons and Ruiz-shulcloper, 2011; Strehl and Ghosh, 2002). Within this role, cluster ensemble integrates solutions from multiple different clusterings into a single better solution, normally, beyond what a single clustering algorithm can achieve.

In addition to the benefits outlined above, we present a novel use of cluster ensemble to handle another most difficult problem in data clustering, namely, a model order selection - the problem of estimating the optimal number of clusters in data. Over the years, numerous approaches have been proposed to this problem. A comprehensive survey of methods for estimating the number of clusters is given in (Hardy, 1996; Milligan and Cooper, 1985), as well

as references therein. There have also been several researches devoted to this issue known as the cluster validation problem, whose comprehensive survey can be found in (Halkidi et al., 2001; Dubes, 1993). However, many of such statistical and model-based approaches often require high computation which is not suitable for large data (such as images) and real-time applications, for example, a real-time image segmentation problem. The goal of this paper is to provide a learning procedure to deal with model order selection in an image segmentation ensemble problem.

Image segmentation ensemble combination has emerged in a similar manner as cluster problems, namely, in order to improve the robustness, stability, and accuracy of segmentation results. In general, the desired number of regions is not known in advance and is often specified by a human user. Most segmentation ensemble approaches assume a known number of regions (Jiang and Zhou, 2004; Rohlfing et al., 2004; Fischer and Buhmann, 2003). The approaches for automatically estimating the number of regions are mostly based on a validity index where the optimal combination solution with a desired number

of regions is found from a series of candidate combination solutions under a different number of regions. Rabinovich et al. (Rabinovich et al., 2006) proposed a validity index based on the concept of cluster stability and used it to find a shortlist of most stable segmentations from a large number of possible segmentation solutions. Wattuya et al. (Wattuya et al., 2008) proposed a validity index based on the concepts of generalized median and mutual information. Their segmentation validation criterion demonstrated promising results. More details of this method will be given in the experiment section as we will use it in comparison with our proposed method. The main drawback of these approaches is their high computation time required for computing a series of candidate combination solutions and for validating them.

In a general data clustering ensemble, Fred and Jain (Fred and Jain, 2005) proposed a thresholding criterion for estimating the true number of clusters called a lifetime criterion. The k -cluster lifetime is defined as the range of threshold values on the dendrogram that lead to the identification of k clusters. The optimal number of clusters is determined as the one with the highest lifetime value. The important drawback of this criterion is that it does not work well with complex data (i.e. not well separated and a non-spherical cluster shape). It requires a large size of ensemble (e.g. 200 clusterings per ensemble) in order to ensure convergence of the method. This is quite infeasible in practice for an image segmentation combination approach.

In this work, we propose a *case-based reasoning* (CBR) approach to handle this task. CBR is suitable for handling our problem in many aspects (will be discussed in Section 3). A knowledge domain for building a case base is derived from the segmentation ensemble in a framework of knowledge reuse. Once a case base is built, the solution (i.e. the number of regions) of a new problem is derived by adapting the solutions of the previous cases retrieved from the case base. The approach is designed to be applicable to practically any image segmentation method and image segmentation combination framework. The main advantages of our approach over the existing methods are that i) it requires significantly less computation time, i.e. no time required for computing a series of candidate combination solutions and no time required for searching a large set of candidate solutions for the optimal combination solutions; ii) it can work well with a small size of segmentation ensemble. These advantages make the concept of image segmentation combination more feasible in real-world applications.

In fact, the idea of the use of segmentation ensemble in model order selection is partly motivated by the

problem of image segmentation ensemble combination. In a framework of image segmentation combination, a segmentation ensemble needs to be generated for combination process. Thus, instead of only using it for a combination purpose, we propose to use it for determining the number of regions in a final combined result, as well. Hence, no additional work for generating a segmentation ensemble is required in this approach.

We extensively evaluate our approach on a large dataset. Recent approaches for determining the number of regions based on the median concept (Wattuya et al., 2008) and minimum description length principle (Rao et al., 2009) are described in Section 4 where they are compared with our method. Three simulations with different state-of-the-art segmentation algorithms are presented to illustrate the efficacy of the proposed approach. Each segmentation algorithm simulates different characteristics of experts to build the case base. The experiments show promising results. Our CBR approach can significantly reduce the computational time required by the existing model order selection methods without the loss of segmentation combination accuracy.

The paper is organized as follows. We begin with a review of using CBR on image segmentation problem. Then, our proposed CBR approach is described in Section 3. Experimental results are described in Section 4. Finally, future work is discussed to conclude the paper.

2 RELATED WORK

The concept of CBR as a problem-solving technique has been used in a wide range of application areas including computer vision applications. CBR has been successfully applied not only in low-level tasks in computer vision, such as image segmentation (Perner, 1999; Frucci et al., 2008), but also in high-level tasks, such as image interpretation (Perner et al., 2003) and medical image understanding (Grimnes and Aamodt, 1996). The works (Perner, 1999; Frucci et al., 2008) used CBR for solving a parameter selection problem which differs from our task. They used CBR to find an appropriate parameter setting for each individual image to ensure that good segmentation results are obtained even if the segmentation algorithm is applied to images with different characteristics. However, the problem of parameter selection seems to be less complicated than our problem of determining an appropriate number of clusters. For a particular segmentation algorithm, solutions of parameter selection problem of images with similar characteristics tend to be the

same, while solutions of model order selection problem may be totally different. In the next section, we will describe how we use segmentation ensemble as an expert domain to cope with this difficulty.

3 OUR PROPOSED CBR

CBR is a problem-solving technique that uses previous cases or past experiences to solve new problems (Kolodner, 1993; Leake, 1996). CBR is suitable for dealing with our problem for two reasons. Firstly, the problem of image segmentation is ill-defined. It is hardly possible to explicitly express it in terms of rules or models. However, these rules/models are not required by CBR. Secondly, CBR can learn for itself to solve future problems in an automatic way.

Segmentation ensemble is a key component for building our CBR. We use it as a knowledge domain for generating a case base. Each segmentation solution in an ensemble (computed from different segmenters) can be viewed as a judgment from individual experts on the input image. Having benefited from this information, the CBR will be able to settle naturally the appropriate number of regions underlying a segmentation ensemble. In this section, we describe the details of building our case base. We begin by describing the way we extract knowledge from a segmentation ensemble and the way we represent it as a case in a case base. Then we describe how we retrieve similar cases from the case base, and finally, how we adapt them to the new problem.

3.1 Knowledge Extraction

The first task of building a case base is to determine the problem features that make up a case. These features should define our problem well. Thus, we first investigate the connection between image features, information provided by a segmentation ensemble, and a true number of regions (k^*) in an image.

*Supposition I: Image Features vs. k^**

The first intuitive assumption is about the relation between low-level image features and a number of regions in an image. However, it is not easy to model or define a correlation between them. It is typical that two or more images happen to have similar feature values but different number of regions and it is often that two or more images with different feature values have similar number of regions.

*Supposition II: Median of k_i of Ensemble vs. k^**

The second conjecture was made based on the assumption that the number of regions of segmentations

in an ensemble (will be denoted as k_i - a number of regions of the i th segmentations in an ensemble) could be used to imply k^* . Thus, simple statistical measures such as the mean and the median of k_i have been investigated in (Wattuya et al., 2012). The work presented in (Wattuya et al., 2012) is our first attempt to develop our idea of using CBR for model order selection. In that work we used both low-level features and the median of k_i to represent a case in a case base (we applied the fuzzy c -means algorithm to handle the ambiguity of low-level features). The experiments showed good results, however, we later found out that the median of k_i works well when the values of k_i is in a normal range (e.g. $k \in [2, 60]$). For example, if an image segmentation algorithm used for generating segmentation ensembles tends to produce over-segmentation results (e.g. $\bar{k} > 50$, the median of k_i would be less appropriate for implying a true number of regions. This is true for the mean of k_i . In our experiments, this situation is simulated and the experimental results show the ineffectiveness of using the median and the mean alone.

*Supposition III: Segmentation Quality vs. k^**

The assumption we used in the proposed approach is based on the quality of segmentations in an ensemble. Thus, two pieces of information are extracted from each segmentation solution in an ensemble. For each segmentation S_i , the first piece of information is the number of regions k_i and the second piece is the quality q_i . The first piece of information can be viewed as the decision of an expert on the k^* of a given image, while the second piece reflects the expertise of that expert.

Let us first consider good quality segmentations in an ensemble. What do they tell us? Good quality segmentations, whose q_i is relatively high, tell us that their corresponding k_i are reliable and very likely to be k^* . Based on this assumption, we can further imply that segmentations with similar values of q_i and similar values of k_i tend to have similar values of k^* . On the other hand, segmentations with similar values of q_i but different values of k_i should have different values of k^* . Bad quality segmentations are not absolutely useless. They can be used to help estimate k^* . We consider three situations:

- Low k^* : q_i of high k_i would be low. The further the k_i is away from k^* , the lower the value of q_i is. This implies that the value of k^* is not high.
- High k^* : q_i of low k_i would be low. This implies that the value of k^* of this image is not low.
- Moderate k^* : q_i of too low or too high k_i would be low. This implies that the value of k^* of this image is not too low or too high.

It is possible that some bad segmentations may not follow the above assumption, since the bad segmentations are arbitrarily bad in their own way. However, in our experiments we observed that the majority of them act in the above manner since we used expert segmentation algorithms, not arbitrary ones (e.g. k -means). Nonetheless, we manage to handle this uncertainty in two ways. The first way is to exclude a few of the worst segmentations in each ensemble from consideration. The second way is to generalize a case base. We will discuss how to evaluate the quality of segmentations in an ensemble later in this section. In the rest of the paper, we will call q_i and k_i extracted from segmentations in an ensemble *ensemble features*.

We claim that ensemble features are higher related to k^* than low-level image features. We test this hypothesis by clustering 300 images from BSDS dataset into C clusters using ensemble features and low-level image features, separately. Image features are contrast, correlation, homogeneity, and entropy (Haralick et al., 1973). For each cluster, we compute the standard deviation of k of images within the same cluster. Small standard deviation of k indicates that all images in the same cluster have k close to each other. On the other hand, large standard deviation of k indicates that the number of k within a cluster are highly varied. Therefore, we expect to have a low standard deviation of k for each cluster. The numbers of k used in this test are taken from human segmentations of BSDS dataset. Since one image has multiple human segmentations, the average of them are used. For ensemble features, we generated three sets of features which are based on three different image segmentation algorithms (which will be described later in Section 4). For each set of features, we conducted a series of experiments with several clustering algorithms with $C \in [30, 40, 50, 60, 70]$ clusters. The three clustering algorithms with the largest Silhouette index and Calinski-Harabasz index values (i.e. complete linkage, average linkage, and k -means) show that the average of standard deviations of k of clusterings using ensemble features is smaller than the average of standard deviations of k of clusterings using low-level image features. This indicates that ensemble features are more related to k^* than low-level features.

3.2 Building a Case Base

3.2.1 Case Description

Given a segmentation ensemble $S = \{S_1, \dots, S_N\}$ of N segmentations of the same image, each segmentation can be produced by using different algorithms

or with the same algorithm but different parameter values. Let $K = \{k_1, \dots, k_N\}$ be a set of the number of regions in the segmentations in S . As we discussed earlier, a case is described as a feature vector $\mathbf{x} = \{\langle k_1, q_1 \rangle, \dots, \langle k_r, q_r \rangle\}$. k_i and q_i are the number of regions and the quality value of the i th segmentation in an ensemble. $r = [1, N]$ is the number of segmentations to be used.

Since in practice the ground truth data is not available for evaluating the quality of segmentation solutions, an unsupervised validation method has to be used. In this work, we proposed to apply two different concepts for evaluating the quality of segmentations. The first one is based on the concept of mutual information, while the second one is based on the concept of minimum description length (MDL) principle.

SNMI Quality Measure. The SNMI measure is first introduced by (Strehl and Ghosh, 2002) for designing consensus functions. Given an ensemble E of r clustering solutions denoted by $E = C_1, \dots, C_r$, Strehl and Ghosh suggested that a good consensus clustering should maximize the following criterion: $SNMI(C, E) = \sum_{i=1}^r NMI(C, C_i)$, where NMI is the normalized mutual information which is used as similarity measures in many contexts including image segmentation (Fowlkes et al., 2003; Franek et al., 2012; Wattuya et al., 2008). The NMI value is maximized to be one if two clusterings define the same partition of the data. In contrast, if two clusterings define completely independent partitions, the NMI value is 0. We propose to apply SNMI as a quality index (q) and use it for a case description. Higher SNMI value indicates higher quality of segmentation that can capture more information among other segmentations in an ensemble.

MDL Quality Measure. The MDL principle can be used to encode image information (e.g. texture, boundary, color) and evaluate segmentation quality based on a coding length function. It defines the best fitted segmentation as the one that produces the shortest code length of the image. Hence, a low value of MDL indicates a good segmentation result. In this work we adapt the MDL criterion used for image segmentation as proposed in (Rao et al., 2009). For additional details the readers are referred to (Rao et al., 2009). Since the range of values of k_i and q_i is relatively large, all values are normalized into a range of 0 to 10.

Note that the main advantages of SNMI-based measure over MDL-based measure are that it is fast computation and requires no original image data in computation process. This is very useful when the original data are not available.

3.2.2 Case Solution

A case solution is defined as an approximate number of regions, \hat{k} , in a final segmentation combination result. We estimate the number of \hat{k} through the framework of image segmentation combination. The random walker-based segmentation combination algorithm (Wattuya et al., 2008) is used to produce a series of combination results with different number of $k \in [k_{\min}, k_{\max}]$. Once candidate results are obtained by a combination algorithm, the next step is to evaluate the combination solutions to determine an optimal one. \hat{k} is then defined as the number of regions of the optimal combination result. The accuracy of \hat{k} depends on a segmentation validation index. In this work we proposed two validity criteria for choosing the optimal combination result. The first criterion is based on the NMI concept and the second criterion is based on the MDL principle.

NMI-based Validity Criterion. This criterion follows the concept of mutual information mentioned above. The best combined segmentation result should share the largest amount of information with all corresponding human segmentations. The candidate combination result with the highest NMI value is chosen as the optimal combination result. This criterion is a supervised method and requires ground-truth segmentation. However, the human segmentations is only used for computing the case solutions in a training process. Thus, the on-line process of our approach is still fully automatic and requires no human intervention. We will call this approach *CBR-NMI*.

MDL-based Validity Criterion. The MDL we used here is the same as the one we use for evaluating the quality of segmentations described above. The MDL criterion defines the optimal combination solution as the one that minimizes its total coding length. This criterion is an unsupervised method and requires no human segmentation data. However, the accuracy of \hat{k} would not be as accurate as the supervised method. We will call this approach *CBR-MDL*.

A case in a case base is thus defined as a problem-solution pair $\mathbf{c} = (\mathbf{x}, \hat{k})$, where \hat{k} is a solution corresponding to a problem feature vector \mathbf{x} defined above.

3.3 Case Retrieval

Case retrieval is a process of finding the cases that are closest to the new problem. The cases with the highest similarity values will be chosen and their associated solutions will be adapted to the new problem. Because our case description is extracted from a segmentation ensemble, its accuracy depends directly on the performance of image segmentation al-

gorithm that we use to generate a segmentation ensemble. Practically, a segmentation ensemble typically contains incorrect segmentations. The information about k^* extracted from them is not absolutely accurate. Hence, it would be more beneficial to compensate their errors by generalizing the cases in a case base and use a case prototype as a case representative, instead of using each individual case on its own.

For this purpose, we apply a k -means clustering algorithm to cluster the cases into c clusters. The cases with similar case descriptions will be clustered together. The c cluster centers are used as case representatives of the whole cases in the case base and their solutions are defined as the average of the solutions of the cases in the same cluster. It is not unusual that the solutions of the cases in the same cluster are not exactly the same, however, a small difference among them is observed. Actually, the averaging of the case solutions is a good concept for error compensation of individual case solutions.

It is interesting to note that k -means clustering used in our approach not only benefits for a case base generalization but also for decreasing the number of cases required for matching in the retrieval step. Instead of matching all the original cases in the entire case base, the task of case matching boils down to matching the new case with the c new case representatives only. Since our case is represented as a real-value feature vector, the simple Euclidean distance is used in a case matching process.

We note here that in our previous work (Wattuya et al., 2012) we used the fuzzy c -means algorithm for a case base generalization since we needed it to handle uncertainty in both case descriptions (as discussed earlier) and case solutions. In that work a case solution was also derived from a segmentation ensemble. Hence, it is more uncertain and more inaccurate than the case solution defined in this work.

3.4 Cases Adaptation

According to our case description, it is impossible that a new case will exactly match any case in a case base. Thus, the case that is closest to the new case will be retrieved, and the solution of a retrieved case will be applied directly to the new case. However, there is a situation where the new case is close to more than one case in a case base. For example, the difference among the distances between the new case and those cases is very small (i.e. less than 0.005). In this situation, a group of cases that are closest to the new case will be retrieved. The solution for the new case is then derived by averaging the solutions of the retrieved cases.

4 EXPERIMENTS

4.1 Segmentation Ensemble Generation

We tested our approach on 300 natural images of size 481×321 from the BSDS dataset (Martin et al., 2001), each having multiple manual segmentations. We apply NMI to quantitatively evaluate the segmentation quality against the ground truth. One segmentation result is compared to all manual segmentations and the average NMI (ANMI) is reported. Larger ANMI values indicate better combination results that share more information with the ground truths.

In the experiments, three state-of-the-art segmentation algorithms were used as baseline segmentation algorithms for generating a set of initial segmentations to be combined. The three segmentation algorithms are the graph-based algorithm (FH) (Felzenszwalb and Huttenlocher, 2004), the mean shift-based algorithm (MS) (Comaniciu and Meer, 2002), and the region growing-based algorithm (JSEG) (Deng and Manjunath, 2001). These algorithms were chosen because of their good performance, computational efficiency and different segmentation behavior.

A segmentation ensemble is obtained by varying the parameter values of the same segmentation algorithm in an appropriate range as summarized in Table 1. The total number of parameter combinations for each algorithm is equal to 24 combinations. For each combination of parameters, the segmentation algorithms are run over the complete set of 300 images from the BSDS dataset to form three sets of segmentation ensembles (with 24 segmentations per ensemble). We call the set of 300 segmentation ensembles produced by FH, MS, and JSEG algorithms as '*FH ensembles*', '*MS ensembles*', and '*JSEG ensemble*', respectively.

The three segmentation algorithms have different segmentation behavior, resulting in the difference of segmentation ensemble characteristics as illustrated in Figure 1. The x -axis is the quality of segmentations q (high value of q indicates good segmentation quality) and the y -axis is the number of region k . FH algorithm tends to produce segmentation results with a normal range of $k \in [1, 40]$. Thus, it has nearly uniform distribution over the full range (as shown in the first plot). MS and JSEG algorithms tends to produce segmentation results with a much wider range of k . The dense distribution in the upper left area of the plots correspond to the highly over-segmentation behavior of the segmentation algorithms. Notice that segmentations with overly high k have very low q (close to 0). Our experiments demonstrate that our CBR approach can work well for all data distribution.

Table 1: Segmentation algorithms and their parameters.

Algos.	Parameter values
FH	$\sigma = \{.4, .5, \dots, .9\}, K = \{150, 300, 500, 700\}, M = 1500$
MS	$h_s = \{8, 16\}, h_r = \{7, 11, 15\}, M = \{100, 500, 1000, 1500\}$
JSEG	$l = \{1, 2, 3\}, q = \{300, 600\}, m = \{0.2, 0.4, 0.6, 0.8\}$

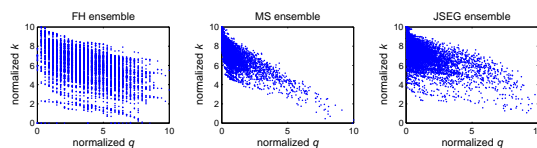


Figure 1: The distribution of three sets of ensembles after normalized into a range [0, 10].

4.2 Model Order Selection Methods

Four methods for determining the number of regions in a final combination results are used for comparison: the generalized median-based validation method (GM) (Wattuya et al., 2008), the MDL-based validation method (MDL) (adapted from (Rao et al., 2009)), and two statistical methods based on the mean and the median of a number of regions in an ensemble.

For the first two methods, the random walker-based segmentation combination algorithm (Wattuya et al., 2008) is applied to generate a series of candidate combination results with different $k \in [k_{\min}, k_{\max}]$. Then, an optimal segmentation combination result is chosen as the one that optimized the validity criterion. The generalized median-based method selects an optimal solution as the one with the minimal average NMI distance (i.e. $1 - \text{NMI}$) among all individual segmentations in an ensemble. This approach is regarded as an approximation of the GM segmentation of the input ensemble by investigating a subspace of U , which consists of the combination segmentations for the considered range of k only. For the MDL-based method, a similar MDL function used for building a case base is used. The optimal segmentation combination result is the one with the shortest coding length. For the last two statistical measures, the mean and the median of a number of regions in an ensemble are first computed. Then, the combination algorithm is applied to generate a combination result with k equals the mean/median of a number of regions in an ensemble.

4.3 Experimental Results

Comparison of Model Order Selection Approaches

For our CBR approach, the 300-image dataset is randomly partitioned into a 200-image training set and a 100-image test set. Three non-overlapping test sets are generated. For each test set we run the k -means

Table 2: Average performance overall three test sets.

Ensemble	Model order selection method							
	GM	MDL	Average	Median	MDL quality measure		SNMI quality measure	
					CBR-NMI	CBR-MDL	CBR-NMI	CBR-MDL
FH	0.6179	0.6167	0.6097	0.6077	0.6162	0.6151	0.6149	0.6127
MS	0.6267	0.6252	0.6072	0.6147	0.6296	0.6291	0.6281	0.6293
JSEG	0.5911	0.5946	0.5889	0.5873	0.5967	0.5914	0.5953	0.5937

Table 3: Model order selection methods, their computational requirements, and their actual computation time (in second).

Approach	Number of combination	Number of MDL computation	Number of NMI computation	Number of Euclidean distance computation	Total time
GM	$ k $ (20.9757)	-	$ k N$ (5.8490)	-	(26.8247)
MDL	$ k $ (20.9757)	$ k $ (48.4283)	-	-	(69.4040)
CBR (MDL measure)	1 (1.1832)	N (21.2152)	-	c (0.0001)	(22.3985)
CBR (SNMI measure)	1 (1.1832)	-	$N(N-1)/2$ (1.4083)	c (0.0001)	(2.5916)

algorithm 100 times and retain only the best results. Note that a small variance of ANMI over 100 runs is observed. The CBR performance reported in Table 2 is the average performance over all three test sets. For all experiments we set the number of ensemble members (r) to 20 and the number of clusters (c) to 50. Note that we have studied the sensitivity of our method on the r and c by varying their values in a proper range. No significant change in the performance is observed.

As expected, the mean and median approach obtained the worst performance in all scenarios. As we discussed earlier, the number of regions in an ensemble depends heavily on the behavior of segmentation algorithm used and has a weak correlation to the k^* . It is not surprising that the performance of CBR-NMI is better than that of CBR-MDL because the case solutions of CBR-NMI are computed based on human segmentations, while the case solutions of CBR-MDL are computed based on the MDL-based objective function. The solutions of CBR-NMI should be closer to the k^* than the solution of CBR-MDL. Even though the CBR based on SNMI achieved slightly lower performance than the CBR based on MDL, it has two dominant advantages, namely, it performs much faster and requires no original image data in the computation.

For MS and JSEG ensembles, the performances of both CBR-NMI and CBR-MDL are slightly better than the performances of GM and MDL, while the performance of CBR-NMI and CBR-MDL are slightly worse than the performances of GM and MDL for an FH ensemble. However, the standard deviation of the performance accuracy of our CBR approach is better than GM and MDL approaches in all cases. This indicates that our approach is more stable than the rest. Even though our CBR approach does

not show a significant improvement over the other methods, we argue that it does show the significant reduction of computation time, especially the CBR approach based on SNMI quality measure. This contribution would make the segmentation combination framework more feasible in practical applications.

Comparison of Computation Time

We summarize the details of the four methods in comparison with the proposed CBR approach in Table 3 in terms of the four computation requirements: i) the number of combination results to be computed; ii) the number of segmentations that need to be evaluated using the MDL quality criterion; iii) the number of times needed for computing a distance between two segmentations based on NMI measure; and iv) the number of case matching using Euclidean distance used in the CBR approach. We did not discuss about the time spend on a generation of segmentation ensembles required by a segmentation algorithm because this step is a preprocessing task needed by all approaches.

In the table $|k|$ refers to a number of all possible combination results with different number of regions in $[k_{\min}, k_{\max}]$ and N refers to a number of segmentations in an ensemble. In practice $|k|$ is typically larger than N . It is obvious that the MDL method requires the highest computation time, while the CBR approach requires the least computation time. In all experiments conducted in this work, $|k|$ is equal to 49, where $[k_{\min}, k_{\max}]$ is $[2, 50]$, and N is equal to 24.

The actual computation time (shown in parenthesis) used to process one input image of each approach implemented in MATLAB on an Intel Core i7, CPU 2.30GHz. The computation time reported in the table are the average of computation times of 20 test images. Experimental results show that the computation time of the CBR based on MDL quality measure is substantially less than that of MDL approach

and slightly less than the GM approach. More interestingly, the computation time of the CBR based on SNMI quality measure is substantially less than both MDL and GM approaches and without the loss of segmentation combination accuracy. We are very positive about the potential of our proposed approach because the only computational burden of our approach is just one step for evaluating quality of segmentations in an ensemble. If a more effective segmentation validation method is available, we can not only greatly reduce the overall computation time, but also improve the accuracy of the combination results. Unlike GM and MDL approaches that have a fix overhead of computing a series of combination results.

5 CONCLUSIONS AND FUTURE WORK

We proposed a new approach for automatically determining the number of regions in a final segmentation combination result. We presented a novel use of cluster ensemble concept to handle this difficult problem. We first studied the correlation between a segmentation ensemble and k^* of an image. The information about k^* is then extracted from an ensemble and used as a knowledge domain for building a case base. We represented the extracted knowledge in terms of the number of regions of segmentations in an ensemble and the qualities of them. The concepts of NMI and MDL are used to evaluate the quality of ensemble members. By utilizing this information, our case-based reasoning is able to settle the true number of regions in the final segmentation combination result as good as the more sophisticated methods. Even though our CBR approach does not show any significant improvement over the existing methods, it does show the significant reduction of computational time without the loss of segmentation combination accuracy. This contribution would make the segmentation ensemble concept more feasible in real-world applications. However, the results of our CBR approach presented in this paper are one of our first attempts. There is room for improvement, actually, in most steps of our approach. More effective and sophisticated methods should be very useful to improve the performance of the proposed approach. It is also interesting to apply our approach in different domains or in a general data cluster ensemble application.

REFERENCES

- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Patt. Anal. Mach. Intel.*, 24:603–619.
- Deng, Y. and Manjunath, B. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Patt. Anal. Mach. Intel.*, 23(8):800–810.
- Dubes, R. C. (1993). Cluster analysis and related issues. *Handbook of Pattern Recognition & Computer Vision*, pages 3–32.
- Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.
- Fischer, B. and Buhmann, J. (2003). Bagging for path-based clustering. *IEEE Trans. Patt. Anal. Mach. Intel.*, 25(11):1411–1415.
- Fowlkes, C., Martin, D., and Malik, J. (2003). Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. In *CVPR*, volume 2, pages 54–61.
- Franek, L., Jiang, X., and Wattuya, P. (2012). Local instability problem of image segmentation algorithms: Systematic study and an ensemble-based solution. *IJPRAI*, 26(5).
- Fred, A. L. N. and Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *IEEE Trans. Patt. Anal. Mach. Intel.*, 27(6):835–850.
- Frucci, M., Perner, P., and Sanniti, G. (2008). Case-based reasoning for image segmentation. *IJPRAI*, 22:1–14.
- Grimnes, M. and Aamodt, A. (1996). A two layer case-based reasoning architecture for medical image understanding. In *European Workshop on Advances in Case-Based Reasoning*, pages 164–178.
- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17:107–145.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Syst, Man, and Cyber*, 3(11):610–630.
- Hardy, A. (1996). On the number of clusters. *Computational Statistics and Data Analysis*, 23:83–96.
- Jiang, Y. and Zhou, Z.-H. (2004). Som ensemble-based image segmentation. 20(3):171–178.
- Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA.
- Leake, D. (1996). *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. AAAI Press, CA.
- Martin, D. R., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pages 416–425.
- Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50:159–179.
- Nguyen, N. and Caruana, R. (2007). Consensus clusterings. In *IEEE ICDM*, pages 607 – 612.
- Perner, P. (1999). An architecture for a cbr image segmentation system. *Journal on Engineering Applications in Artificial Intelligence*, 12(6):749–759.

- Perner, P., Guether, T., and Perner, H. (2003). Airborne fungi identification by case-based reasoning. In *Workshop on CBR in the Health Sciences*.
- Rabinovich, R., Belongie, S., Lange, T., and Buhmann, J. M. (2006). Model order selection and cue combination for image segmentation. In *CVPR*, pages 1130–1137.
- Rao, S., Mobahi, H., Yang, A., Sastry, S., and Ma, Y. (2009). Natural image segmentation with adaptive texture and boundary encoding. In *ACCV*, pages 135–146.
- Rohlfing, T., Russakoff, D. B., and Jr., C. R. M. (2004). Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. Med. Imaging*, 23(8).
- Strehl, A. and Ghosh, J. (2002). Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Machine Learning Research*, 3:583–617.
- Topchy, A. P., Jain, A. K., and Punch, W. F. (2005). Clustering ensembles: Models of consensus and weak partitions. *IEEE Trans. Patt. Anal. Mach. Intel.*, 27(12):1866–1881.
- Vega-pons, S. and Ruiz-shulcloper, J. (2011). A survey of clustering ensemble algorithms. *IJPRAI*, 25(3):337–372.
- Wattuya, P., Jiang, X., Prassni, J., and Rothaus, K. (2008). A random walker based approach to combining multiple segmentations. In *ICPR*, pages 1–4.
- Wattuya, P., Soonthornphisaj, N., and Jiang, X. (2012). Using soft case-based reasoning in model order selection for image segmentation ensemble. In *JSAI*.

WILEY
PRESS
TECHNOLOGY PUBLICATIONS