**Abstract**—Clustering ensemble is a promising technique to face data clustering problems. Similarly, the combination of different segmentations to obtain a consensus one could be a powerful tool for addressing image segmentation problems. Such segmentation ensemble algorithms should be able to deal with the possible large image size and should preserve the spatial relation among pixels in the image. In this paper, we formalize the segmentation ensemble problem and introduce a new method to solve it, which is based on the kernel clustering ensemble philosophy. We prove that the Rand index is a kernel function and we use it as similarity measure between segmentations in the proposed algorithm. This algorithm is experimentally evaluated on the Berkeley image database and compared to several state-of-the-art clustering ensemble algorithms. The achieved results ratify the accuracy of our proposal.

**Index Terms**—Consensus segmentation, clustering ensemble, Rand index, kernel function.

I. INTRODUCTION

Clustering ensemble has become a popular technique to deal with data clustering problems [1]. It is known that there is no clustering algorithm that effectively works for every clustering problem. When different clustering algorithms are applied to the same dataset, different clustering results can be obtained. Instead of trying to find the “best” one, the idea of combining these individual results in order to obtain a consensus has gained an increasing interest in the last years. In practice, such a procedure could produce high quality final clusterings.

A similar situation can be found in the image segmentation context. There is no segmentation algorithm to correctly work with all type of images. Different segmentation algorithms or the same segmentation algorithm with different parameter initializations could produce very different segmentations of the same image. Therefore, if there are several segmentations of the image, the idea of combining them following the philosophy of the clustering ensemble algorithms seems to be appropriate.

A first approximation could be the use of the clustering ensemble algorithms directly for segmentation combination problems. In other words, assuming that an image is a set of pixels and a segmentation is a partition of this set of pixels, any clustering ensemble algorithm could be applied to obtain a consensus segmentation. Indeed, some clustering ensemble algorithms (e.g. [2], [3]) have been evaluated on image segmentation problems; see [4] for a recent study of applying general clustering ensemble algorithms on image segmentation combination problems. However, this way of facing the segmentation ensemble problem has two fundamental drawbacks:

- Images are normally composed of a large number of pixels. For example, a rather small image of $500 \times 500$ pixels would lead to a dataset of $250000$ objects.
- We are loosing the spatial relation among pixels if we assume that an image is only a set of individual pixels. Besides, if this spatial relation is not taken into account, inappropriate segmentations could be considered as potential consensus result.

In this paper we first formalize the segmentation ensemble problem and then propose a segmentation combination algorithm based on the philosophy of clustering ensembles, which is able to overcome the above limitations. The rest of this paper is organized as follows. Section II briefly introduces the general clustering ensemble problem and presents some clustering ensemble algorithms. Section III formalizes the segmentation ensemble problem. Section IV presents the proposed method and the proof that the Rand index is a kernel function. This proof allows the use of the Rand index as a similarity measure between segmentations in the proposed algorithm. In Section V several experiments are shown and discussed, and finally Section VI concludes this paper.

II. CLUSTERING ENSEMBLE

Clustering ensemble methods combine partitions of the same dataset into a final consensus clustering. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of objects, $P = \{P_1, P_2, \ldots, P_m\}$ is a clustering ensemble, where each $P_j = \{C_{j1}^1, C_{j2}^1, \ldots, C_{jd_j}^j\}$ is a partition of the set of objects $X$ with $d_j$ clusters, for all $j = 1, \ldots, m$. We also denote $\mathbb{P}_X$ the set of all possible partitions of $X$ and the consensus partition is represented by $P^*$. The consensus partition $P^*$ is usually defined through the median partition problem:

$$P^* = \arg \max_{P \in \mathbb{P}_X} \sum_{i=1}^{m} \Gamma(P, P_i)$$

where $\Gamma$ is a similarity measure\(^1\) between partitions.

Over $\mathbb{P}_X$ a partial order relation $\preceq$ ("nested in") can be defined. For all $P, P' \in \mathbb{P}_X$, $P \preceq P'$ if and only if, for all $i$

\(^1\)This problem can be equivalently defined by minimizing the dissimilarity with all partitions in the clustering ensemble.
cluster $C' \in P'$ there are clusters $C_{i_1}, C_{i_2}, \ldots, C_{i_m} \in P$ such that $C' = \bigcup_{i=1}^{m} C_{i_i}$. This way a lattice structure is associated to $\mathbb{P}_X$ (see example in Fig. 1). In the last years, several clustering ensemble methods, where the consensus partition is defined through the median partition problem (1), have been proposed. Most of these methods consist of heuristics based on the use of different mathematical and computational tools (see [1] for a survey). The segmentation ensemble method proposed in this paper is based on the Weighted Partition Consensus via Kernels (WPCK) method introduced in [5], which is briefly explained in the next section.

A. Weighted Partition Consensus via Kernels

In the Weighted Partition Consensus via Kernel (WPCK) method, the consensus partition is theoretically defined\(^2\) as:

$$P^* = \arg\max_{P \in \mathbb{P}_X} \sum_{i=1}^{m} k(P, P_i)$$

where $k$ is a positive definite kernel\(^1\) function [6].

If $k$ is a kernel function, there is a function $\phi : \mathbb{P}_X \rightarrow \mathcal{H}$ that allows mapping the problem (2) into the Reproducing Kernel Hilbert Space $\mathcal{H}$ associated to $k$. The exact solution $\psi$ in $\mathcal{H}$ can be found as the sum of the input partitions mapped in $\mathcal{H}$ by $\phi$, divided by its norm, i.e.:

$$\psi = \sum_{i=1}^{m} \phi(P_i) \left/ \sum_{i=1}^{m} \phi(P_i) \right.$$

Then, the consensus partition $P^* \in \mathbb{P}_X$ is the solution of the pre-image problem

$$P^* = \arg\min_{P \in \mathbb{P}_X} \|\phi(P) - \psi\|^2$$

where $\|\phi(P) - \psi\|^2$ can be rewritten only in terms of the similarity measure $k$ in the following way:

$$\|\phi(P) - \psi\|^2 = 2 - 2 \sum_{i=1}^{m} \sum_{j=1}^{n} k(P_i, P_j)$$

where the denominator is a constant value that does not depend on any particular partition $P$, therefore can be computed only once. By this last equation we know how close each partition in $\mathbb{P}_X$ is to the consensus partition. Thus, this equation can be used as a global measure to evaluate any candidate partition $P \in \mathbb{P}_X$ as a possible consensus. This is of great importance because by directly working with equation (2) we can compute the sum of similarity values of a given partition, but we cannot estimate its deviation to the consensus solution. Therefore, equation (3) allows defining an iterative procedure that starting from an initial partition, tries to find a new partition closer to the consensus partition. In [5] the simulated annealing meta-heuristic is used to avoid the convergence to a low quality local optimum.

III. SEGMENTATION ENSEMBLE PROBLEM FORMULATION

As was previously mentioned, a direct application of clustering ensemble algorithms on segmentation combination problems present some drawbacks, e.g. generation of very large datasets and loss of spatial relation among pixels.

In order to overcome the first limitation, the idea of the super-pixel representation [3], [4] of an image is used. The use of super-pixels is motivated by the fact that neighboring pixels that were grouped in the same region in all segmentations do not have to be analyzed in a separated way. It is expected that such pixels are grouped in the same region in the consensus segmentation. Therefore, a representative element (super-pixel) could be selected from each of these regions.

Thereby, a super-pixel is a connected component in the image, formed by pixels that were grouped in the same cluster in all the segmentations to be combined. Moreover, two super-pixels $\sigma_a, \sigma_b$ are neighbors if there are at least two neighboring pixels $\pi_a, \pi_b$ such that $\pi_a \in \sigma_a$ and $\pi_b \in \sigma_b$.

Given an image and a set of segmentations of the image, the super-pixels computation can be efficiently performed in $O(w \cdot b \cdot m)$, where $w$ and $b$ are the image dimensions and $m$ the number of segmentations to be combined. In the worst case the number of super-pixels could be the number of pixels in the image. In practice, however, the number of objects is substantially reduced after the super-pixel computation.

Our segmentation ensemble problem formulation is based on the definition of the concepts Image and Segmentation by using super-pixels. These definitions also consider the spatial relation among pixels in the image. As it can be seen in Figure 2, we can associate a graph to an image, which represents the neighborhood relation among the super-pixels in the image. This graph is essential in the following definitions.

Definition 1: An image is a pair $I = (I_{\Sigma}, I_{G})$, where $I_{\Sigma} = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ is a set of super-pixels, $I_{G} = (V, E)$ is the connectivity graph of super-pixels, where there is a node $\nu_i \in V$ associated to each super-pixel $\sigma_i$ and there is an edge $e_{ij} \in E$ if and only if the super-pixels $\sigma_i$ and $\sigma_j$ are neighbors.

Proposition 1: The graph $I_{G}$ is a planar graph.

This proposition is given without a proof, since it can be easily verified from the graph definition, i.e., this graph connects neighboring objects in the two-dimensional plane.

Definition 2: A segmentation $S = \{R_1, R_2, \ldots, R_d\}$ is a set of regions (clusters) $R_i \subseteq I_{\Sigma}$ that holds the following four properties:

- $R_i \neq \emptyset, \forall i = 1, \ldots, d$
- $\bigcup_{i=1}^{d} R_i = I_{\Sigma}$
- $R_i \cap R_j = \emptyset, \forall i, j = 1, \ldots, d$, with $i \neq j$
This equation is used to know how close to the consensus segmentation any segmentation \( S \in \mathbb{S}_1 \) is. Therefore, the simulated annealing meta-heuristic can also be used. The first state (segmentation) in the process is defined as the segmentation in the segmentation ensemble \( \mathbb{S} \) closest to the theoretical consensus, which can be computed as follows:

\[
S_0 = \arg \min_{S \in \mathbb{S}} \| \phi(S) - \psi \|^2
\]

Other parameters of the simulated annealing like the temperature, cost function, etc. can be computed as in [5]. However, in this case, the search space of the problem is \( \mathbb{S}_1 \) instead of \( \mathbb{P}_1 \). Then, the state (segmentation) neighborhood should be modified. In order to guarantee the good performance of the simulated annealing, it is important that the new neighborhood definition fulfills the following conditions:

- Given a state (segmentation), the process of generating a new neighbor state must ensure that the new state holds the proposed segmentation definition (without an additional verification).
- Starting from an initial segmentation, any segmentation in \( \mathbb{S}_1 \) can be reached in \( O(n) \) steps.

Given an image \( I = (I_S, I_G) \), with \( I_G = (V, E) \), and a segmentation \( S \) of this image, the following neighborhood segmentation generation process is proposed. A super-pixel \( \sigma \in I_S \) and an edge \( \{\sigma, \sigma'\} \in E \) are randomly selected. If the super-pixel \( \sigma' \) does not belong to the same region (cluster) of \( \sigma \) in \( S \), the super-pixel \( \sigma \) is moved to the region where \( \sigma' \) belongs. On the other hand, if \( \sigma \) and \( \sigma' \) are in the same region, a new empty region is created and \( \sigma \) is moved to this new region. This procedure satisfies both properties mentioned above.

Finally, as the graph \( I_G = (V, E) \) is planar, \( |E| \leq 3 \cdot |V| - 6 \) and therefore \( |E| = O(|V|) \). In other words, we can store in memory and iterate over the set of edges \( E \). The space requirement and time complexity are linear with respect to the number of super-pixels in the image. Hence, the proposed neighboring segmentation generation process does not affect the algorithm’s complexity. On the other hand, the reduction of the search space to \( \mathbb{S}_1 \) allows finding good solutions with a smaller number of iterations.

We call the proposed algorithm **Segmentation Ensemble via Kernels (SEK)**. Given an image and a set of segmentations of this image, the algorithm steps are the following:

1) Computation of the set of super-pixels \( I_S \) and the connectivity graph \( I_G \). Selection of the kernel similarity measure \( k \) between segmentations to be used in the problem formulation.

2) Computation of the initial state (segmentation) \( S_0 \) by solving equation (6).

3) Application of the simulated annealing. Take \( S_0 \) as current solution and while a maximum number of iterations is not reached, try to find a new segmentation in \( \mathbb{S}_1 \) better than the current solution. The move from one segmentation to another is done by using the proposed segmentation neighboring relation.
4) In the solution given by the simulated annealing, each super-pixel is substituted by the set of pixels that it represents in order to form the final consensus segmentation of the image.

A. Rand index as a kernel similarity measure between segmentations

In the experiments we will use the proposed algorithm with two kernels as similarity measures between segmentations. The first is the subset significance based measure introduced in [5]. Given a set of objects \( X \) and the set of all possible partitions of these objects \( \mathbb{P}_X \), this measure is defined by

\[
SS(P_a, P_b) = \sum_{S \subseteq X} \mu(S, P_a) \mu(S, P_b)
\]

where \( \mu(S, P) = \frac{|S|}{|P|} \) if \( \exists C \in P \) such that \( S \subseteq C \), and \( \mu(S, P) = 0 \) otherwise. The second measure is the Rand index, which is a widely used similarity measure between clusterings and segmentations. We present in this section a proof that the Rand index measure is a positive definite kernel function.

The Rand index \([7]\) is a function \( RI : \mathbb{P}_X \times \mathbb{P}_X \rightarrow \mathbb{R} \), which can be defined between two partitions \( P_a \) and \( P_b \) as:

\[
RI(P_a, P_b) = \frac{N_{00} + N_{11}}{n(n-1)/2}
\]

where \( N_{11} \) is the number of pairs of objects that belong to the same cluster in both partitions \( P_a \) and \( P_b \), while \( N_{00} \) is the number of pairs of objects that belong to different clusters in \( P_a \) and \( P_b \). These measures can be written as:

\[
N_{11} = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}^a \cdot \delta_{ij}^b
\]

where

\[
\delta_{ij} = \begin{cases} 1, & \text{if } (i \neq j) \land (\exists C \in P_a \mid x_i \in C \land x_j \in C); \\ 0, & \text{otherwise.} \end{cases}
\]

\[
N_{00} = \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}^a \cdot \gamma_{ij}^b
\]

where

\[
\gamma_{ij} = \begin{cases} 1, & \text{if } \exists C \in P_a \mid x_i \in C \land x_j \in C; \\ 0, & \text{otherwise.} \end{cases}
\]

Then, the Rand index can be rewritten as:

\[
RI(P_a, P_b) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}^a \cdot \delta_{ij}^b + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}^a \cdot \gamma_{ij}^b}{n(n-1)}
\]

\( \text{(8)} \)

Proposition 2: The Rand index is a positive definite kernel.

Proof: In order to prove that the Rand index (RI) is a kernel function we need to prove that:

- RI is symmetric.
- \( \forall t \in \mathbb{N}, \forall \alpha_1, \alpha_2, \ldots, \alpha_t \in \mathbb{R} \) and \( \forall P_1, P_2, \ldots, P_t \in \mathbb{P}_X \):

\[
\sum_{a=1}^{t} \sum_{b=1}^{t} \alpha_a \alpha_b \cdot RI(P_a, P_b) \geq 0
\]

The symmetry is evident from the definition of the Rand index. The second property can be proven by substituting equation (8) in the above expression.

\[
\sum_{a=1}^{t} \sum_{b=1}^{t} \alpha_a \alpha_b \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}^a \cdot \delta_{ij}^b + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}^a \cdot \gamma_{ij}^b \right) \geq 0
\]

Reorganizing the terms we have

\[
\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \sum_{a=1}^{t} \alpha_a \delta_{ij}^a \right)^2 + \left( \sum_{a=1}^{t} \alpha_a \gamma_{ij}^a \right)^2 \geq 0
\]

and reorganizing again the terms we obtain

\[
\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \sum_{a=1}^{t} \alpha_a \delta_{ij}^a \right)^2 + \left( \sum_{a=1}^{t} \alpha_a \gamma_{ij}^a \right)^2 \geq 0
\]

and the proposition is proven.

V. Experimental Results

We used the color images from the Berkeley dataset \([8]\) for the experiments. It is widely used for image segmentation evaluation and is composed of 300 natural images of size 481 × 321. For each image, we used two state-of-the-art segmentation algorithms to generate two ensembles: TBES ensembles and UCM ensembles. Each ensemble is composed of 10 segmentations obtained by varying the parameter values of the segmentation algorithms. TBES ensembles were generated with the TBES algorithm \([9]\), which is based on the MDL-principle and has as parameter the quantization level (\( \epsilon \)). We varied \( \epsilon = 40, 70, 100, 130, \ldots, 310 \) to obtain the 10 segmentations in the ensemble. Furthermore, UCM ensembles were generated with a segmenter based on ultrametric contour map (UCM) \([10]\). Its only parameter is the threshold \( l \), we choose \( l = 0.03, 0.11, 0.19, 0.27, 0.35, 0.43, 0.50, 0.58, 0.66, 0.74 \).

Each image in this dataset has a human segmentations (ground-truth). In the experiments, we compared the obtained results with the human segmentations of each image by using three well-known measures to evaluate the algorithm results: Normalized Mutual Information (NMI) \([11]\), Variation of Information (VI) \([12]\), and Rand Index (RI) \([7]\). NMI and RI are similarity measures that take values in the range \([0, 1]\), where 1 means a perfect correspondence between the segmentation and the ground-truth. On the other hand, VI is a dissimilarity measure that takes values in \([0, +\infty]\), where 0 means a perfect correspondence. In order to show the experimental results in a homogeneous way we present 1-NMI and 1-RI values, which are dissimilarity versions of the original measures. Thus, for all measures lower values mean better correspondence.

In Table I, the comparison of the obtained results with the best human segmentation is presented in the column denoted bestGT and the average value of the comparisons with all the human segmentations is presented in column allGT. In this table, the proposed algorithm (SEK) is compared with several clustering ensemble algorithms applied to the segmentation combination problem. These algorithms are BOK and BOEM \([13]\); CSPA, HGPA and MCLA \([11]\); EA-AL \([2]\) and QMI
As can be seen in Table I, for almost all ensembles and validity measures, one of the two variants of the proposed algorithms obtained the best results. In addition, a superiority of the SEK + RI configuration can be appreciated. This ratifies that the Rand index is an appropriate similarity measure between segmentations and thanks to the proof given in Section IV-A, it can be used in the proposed algorithm. From this table we can experimentally corroborate that the application of general clustering ensemble algorithms in segmentation ensemble problems is possible. However, the design of an algorithm that takes into account the peculiarities of the segmentation ensemble problem produces better results. Finally, we can see from Table I that the proposed algorithm obtained very close results to the supervised parameter learning approach, even when supervised learning makes use of the ground-truth information. It is important to notice that supervised learning is not applicable in many practical scenarios because ground-truth is not available.

VI. CONCLUSION

The segmentation ensemble problem differs from the general clustering ensemble problem. When clustering ensemble algorithms are applied to segmentation combination problems, they are not capable of taking into account and benefiting from the spatial relation among pixels in the image. Besides, the large number of pixels in an image could affect this kind of algorithms. In this paper, we formalized the segmentation ensemble problem and proposed an algorithm based on kernel clustering ensembles, which is able to overcome the above mentioned limitations. The super-pixel representation of the image and the connectivity graph of super-pixels are fundamental tools in our approach. On the other hand, we proved that the Rand index, which is a widely used similarity measure between partitions and segmentations, is a kernel function. This extends the application scope of this measure. In particular, it is successfully used in the proposed algorithm. Finally, the comparison with several state-of-the-art clustering ensemble algorithms on image segmentation combination problems showed the accuracy of the proposed algorithm.

REFERENCES