

# Unsupervised Superpixel-based Segmentation of Histopathological Images with Consensus Clustering

Fouad, Shereen; Randell, David; Galton, Antony; Mehanna, Hesham; Landini, Gabriel

DOI:

[10.1007/978-3-319-60964-5\\_67](https://doi.org/10.1007/978-3-319-60964-5_67)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Fouad, S, Randell, D, Galton, A, Mehanna, H & Landini, G 2017, Unsupervised Superpixel-based Segmentation of Histopathological Images with Consensus Clustering. in Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11-13, 2017, Proceedings. Communications in computer and information science, vol. 723, Springer, pp. 767-779, Medical Image Understanding and Analysis (MIUA) 2017, United Kingdom, 11/07/17. [https://doi.org/10.1007/978-3-319-60964-5\\_67](https://doi.org/10.1007/978-3-319-60964-5_67)

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Unsupervised Superpixel-Based Segmentation of Histopathological Images with Consensus Clustering

Shereen Fouad<sup>1</sup>(✉), David Randell<sup>1</sup>, Antony Galton<sup>2</sup>, Hisham Mehanna<sup>3</sup>,  
and Gabriel Landini<sup>1</sup>

<sup>1</sup> School of Dentistry, Institute of Clinical Sciences,  
University of Birmingham, Birmingham, UK  
{s.a.fouad,d.a.randell,g.landini}@bham.ac.uk

<sup>2</sup> Department of Computer Science, University of Exeter, Exeter, UK  
apgalton@ex.ac.uk

<sup>3</sup> Institute of Head and Neck Studies and Education,  
University of Birmingham, Birmingham, UK  
h.mehanna@bham.ac.uk

**Abstract.** We present a framework for adapting consensus clustering methods with superpixels to segment oropharyngeal cancer images into tissue types (epithelium, stroma and background). The simple linear iterative clustering algorithm is initially used to split-up the image into binary superpixels which are then used as clustering elements. Colour features of the superpixels are extracted and fed into several base clustering approaches with various parameter initializations. Two consensus clustering formulations are then used, the Evidence Accumulation Clustering (EAC) and the voting-based function. They both combine the base clustering outcomes to obtain a single more robust consensus result. Unlike most unsupervised tissue image segmentation approaches that depend on individual clustering methods, the proposed approach allows for a robust detection of tissue compartments. For the voting-based consensus function, we introduce a technique based on image processing to generate a consistent labelling scheme among the base clustering outcomes. Experiments conducted on forty five hand-annotated images of oropharyngeal cancer tissue microarray cores show that the ensemble algorithm generates more accurate and stable results than individual clustering algorithms. The clustering performance of the voting-based consensus function using our re-labelling technique also outperforms the existing EAC.

**Keywords:** Superpixel segmentation · Consensus clustering · Histology · Histopathology · Image analysis

## 1 Introduction

The automatic segmentation of digitised histological images into regions representing different anatomical or diagnostic types is of fundamental importance

© The Author(s) 2017

M. Valdés Hernández and V. González-Castro (Eds.): MIUA 2017, CCIS 723, pp. 767–779, 2017.

DOI: 10.1007/978-3-319-60964-5.67

for developing digital pathology diagnostic tools. Superpixel segmentation is an advanced method to group image pixels with similar colour properties into atomic regions to simplify the data in the pixel grid [1]. Recently, superpixel methods have been combined with pattern recognition techniques for image segmentation (e.g. [2]) where certain features (e.g. colour, morphology) are extracted from the superpixels and then fed to pattern recognition procedures that assign each superpixel to an expected histological class. Supervised methods are built from labelled training sets to predict the classes of novel unlabelled data and they require access to ground truth reference images for the training. In contrast, unsupervised approaches (clustering analysis) do not require pre-labelled training sets for their learning but instead rely on certain similarity measures to group data into separate homogeneous clusters. In histopathological imaging analysis, clustering is particularly useful as an exploratory tool as it can provide information about hidden anatomical or functional structures in images.

Clustering algorithms use different heuristics and can be sensitive to input parameters, i.e. repeatedly applying different clustering methods on the same dataset often yields different clustering results. Furthermore, a given clustering algorithm may give rise to different results for the same data when the initialisation parameters change. Consensus Clustering (CC) [3] methods have addressed this issue by combining solutions obtained from different clustering algorithms into a single consensus solution. In unsupervised learning, this enables more accurate and robust estimation of clusterings when compared to single clustering algorithms. CC is often performed in two main steps, (a) the cluster ensemble generation, and (b) the consensus function, which finds a consensual opinion of the ensemble. CC techniques have proved efficient in a variety of practical domains; their application to histological image segmentation is, however, relatively new.

In this work, we investigate the use of CC in the context of superpixel-based segmentation of haematoxylin and eosin (H&E) stained histopathological images. We suggest a multi-stage segmentation process. First, we use the recently proposed Simple Linear Iterative Clustering (SLIC) superpixel framework [1, 4] to segment the image into compact regions. Colour features from each dye are extracted from the superpixels and used as input to multiple base clustering algorithms with various parameter initializations. The generated results (denoted here as partitions) pass through an ensemble selection scheme which generates a more effective ensemble based on partitions diversity. Two consensus functions are considered here, the Evidence Accumulation Clustering (EAC) [5] and the voting-based consensus function (e.g. [6, 7]).

Unlike supervised methods, labels resulting from unsupervised techniques are symbolic (i.e. labels do not represent a meaningful class), and consequently an individual partition in the ensemble includes clusters that do not necessarily have labels that correspond to other clusters in different partitions of the ensemble. In the voting-based consensus function the label mismatch is defined as the problem of finding the optimal re-labelling of a given partition with respect to a reference partition. This problem is commonly formulated as a weighted bipartite

matching formulation [6, 7], and it is solved by inspecting whether data patterns in two partitions share labels more than with other clusters. In this paper, we present an alternative simple, yet robust, implementation for generating a consistent labelling scheme among the different partitions of the ensemble. Our approach considers the space occupied by each individual cluster in an image and exploits the fact that pairs of individual clusters from different partitions would match when their pixels largely overlap in a segmented image.

## 2 Related Work

SLIC [1] is an advanced supervoxel method that generates compact, mostly uniform supervoxels by agglomerating pixels based on colour similarity and proximity in the image plane. Achanta et al. [4] conducted an empirical comparison of SLIC with other state-of-the-art supervoxel algorithms, which revealed the superiority of SLIC in terms of performance and speed. They also showed that SLIC is easy to use and implement, low in computational cost and requires fewer parameters than other algorithms. All these features are potentially useful for automatic segmentation of large, complex and variable histopathological images. SLIC supervoxels have been used before to facilitate and improve unsupervised segmentation of histopathological images. SLIC was applied in [2] as a pre-processing step to decrease the complexity of large histopathological images. Colour descriptors of the generated regions were then used in an unsupervised learning formulation of the probabilistic models of expected classes using the Expectation-Maximisation (EM) [8].

Consensus Clustering (CC) methods have emerged for improving robustness, stability and accuracy of unsupervised learning solutions. Contributions in this field include the EAC [5] and voting-based algorithms. A comprehensive survey of existing clustering ensemble algorithms is presented in [3]. The voting-based literature utilizes different heuristics in attempting to solve the labelling correspondence problem. This problem is commonly formulated as a bipartite matching problem [6], where the optimal re-labelling is obtained by maximizing the agreement between the labelling of an ensemble partition with respect to a reference partition. The agreement is estimated by constructing a  $K \times K$  contingency table between the two partitions, where  $K$  is the number of clusters in each partition<sup>1</sup>. Each entry of the contingency table holds the number of cluster label co-occurrences counted for the same set of objects in the two partitions.

There have been previous publications on CC in unsupervised histopathological segmentation, but to the best of our knowledge, its application to supervoxel-based segmentation remains unexplored. Simsek et al. [9] defined a set of high-level texture descriptors of colonic tissues representing prior knowledge, and used those in a multilevel segmentation where they used a cluster ensemble to combine multiple partitioning results. Khan et al. [10] proposed ensemble clustering for pixel-level classification of tumour vs. non-tumour regions in breast

---

<sup>1</sup> The two partitions should contain the same number of clusters,  $K$ .

cancer, where random projections of low-dimensional representations of the features and a consensus function combined various partitions to generate a final result.

### 3 Unsupervised Superpixel-Based Segmentation with Consensus Clustering

#### 3.1 Dataset and Preprocessing

Our data consisted of H&E stained tissue images (paraffin sections) of human oropharyngeal cancer processed into tissue micro arrays (TMAs), prepared at the Institute of Cancer and Genomic Sciences, University of Birmingham, UK. H&E is the commonest staining method used in routine diagnostic microscopy; haematoxylin primarily stains nucleic acids and nuclei in blue/violet while the eosin counter-stain primarily stains proteins in the intra- and extra-cellular compartments in pink. TMAs are usually used for the analysis of tumour markers of multiple cases (cores) in single batches where there is a need to identify various components in the samples. Samples were digitised using an Olympus BX50 microscope with a x20 magnification objective (N.A. 0.5, resolution  $0.67 \mu\text{m}$ ) and a QImaging Retiga 2000R camera and a tunable liquid crystal RGB filter (Surrey, BC, Canada).

Tissue core images were  $\approx 3300 \times 3300$  pixels (inter-pixel distance of  $0.367 \mu\text{m}$ ). Fifty five images were used for the analysis (ten for training and forty five for testing), which provided the range of variations in tissue distributions typically found in this type of histological material (2.3 to 98.8% of epithelium tissue component and 25.5 to 83.2% of background out of the whole image).

As a preprocessing step, colour deconvolution [11] was applied to the H&E image  $I$  to separate the RGB information into haematoxylin-only and eosin-only images. With this procedure, up to three dyes (in our case, H&E) can be separated into ‘stain’ channels. This can be applied when the colours of the dyes on their own are known and combine as light-absorbing dyes. In the case of two-dye stains, a third component is a residual channel of the deconvolution process. The results of the colour deconvolution can be combined into a “stain” RGB image here denoted  $I^*$  to better represent the dye absorption of the different tissue types. In  $I^*$  the R, G and B channels now hold the light transmittance of the haematoxylin, eosin and residual images, instead of containing the RGB components. The feature extraction discussed in Sect. 3.3 is applied to this image  $I^*$ .

#### 3.2 Superpixel-Based Segmentation

The SLIC segmentation spits-up the original image  $I$  into a set of superpixels held in a binary image  $S$ . The superpixels tend to be compact and relatively uniform. They are formed by pixel grouping based on colour similarity and spatial proximity. In detail, a  $k$ -means algorithm [12] is used to cluster a five-dimensional vector consisting of the 3 components of a pixel colour in CIELAB space and the

pixel spatial coordinates. A special similarity measure is then exploited, replacing the standard Euclidean distance, which weighs the distance in the colour and spatial domain. This measure weighs the relative importance between color similarity and spatial proximity in the five-dimensional space. Furthermore, it allows the size and compactness of the resulting supervoxels to be adjusted, providing some control over the number of supervoxels generated.

In our experiments, we used the recently proposed jSLIC [13], a Java implementation of SLIC that is faster than the original (in [14]). Unlike the original, jSLIC avoids computing the same distances between data by exploiting pre-computed look-up tables. Borovec et al. showed that the jSLIC is able to segment large images with intricate details into uniform parts, which is particularly useful for complexity-reduction problems (as is the case here). The authors also defined a function  $f$  that compromises between supervoxel compactness and the alignment of object boundaries in the image. This is expressed as:  $f = m \cdot z^2$ , where  $m$  is the initial supervoxel size and  $z$  is a regularisation parameter which affects the supervoxel compactness. The value of  $z$  lies within the range [0,1], where 1 yields nearly square segments and 0 produces very ‘elastic’ supervoxels. To ensure an effective segmentation, we performed a cross validation procedure for the configuration of these two parameters, as discussed in the Experiments and Evaluation section.

### 3.3 Feature Extraction

Colour features are known for their relevance in visual perception and are exploited here for the discrimination of supervoxels into different histological regions. Our H&E images contain at least three types of regions that uptake dyes differently: (1) stratified squamous epithelial tissue (a ‘solid’ tissue with densely packed cells which appear more darkly stained than the rest), (2) connective stroma, which is less cellular and contains abundant extracellular matrix, blood vessels, inflammatory cells, and sometimes glandular tissue, and (3) background areas, often appearing white or neutral grey. First the colour descriptors for each supervoxel in image  $S$  are computed but instead of referring to the original  $I$ , these are extracted from the data in image  $I^*$ , so they become ‘stain features’ that quantify the distribution of the stain uptake in the supervoxels. We used eleven measures for each stain (mode, median, average, average deviation, standard deviation, minimum, maximum, variance, skew, kurtosis and entropy) for each of the three colour deconvolution components (haematoxylin, eosin and the residual channel), forming a vector of thirty-three colour descriptors per supervoxel.

### 3.4 Consensus Clustering (CC) Frameworks

The CC framework exploited here involves three main steps (1) creation of an ensemble of multiple cluster solutions, (2) selection of an effective sub-set of cluster solutions based on their diversity measure, and (3) generation of a final partition via the so-called consensus function. A clustering algorithm takes the

set  $X = \{x_1, x_2, \dots, x_n\}$  of  $n$  superpixels as an input, and groups it into  $K$  clusters (epithelium, stroma and background regions) forming a data partition  $P$ . Note that  $x_i$  is characterized here by the 33-dimensional colour features described in the previous section.

**Ensemble Generation and Selection.** First, a number of  $q$  clustering results are generated for the same  $X$ , forming the cluster ensemble  $E$ , where  $E = \{P_1, P_2, \dots, P_q\}$ . To this end, we used five different clustering algorithms and ran each of those multiple times while varying their parameters. There are two factors that influence the performance of this approach: one is the accuracy of the individual clusters ( $P_i$ ) and the other is the diversity within the ensemble  $E$ . Accuracy is maintained by tuning a set of effective clustering methods to obtain the best set of results. Regarding the diversity of  $E$ , it was shown in [15] that a moderate level of dissimilarity among the ensemble members ( $E$ ) improves the consensus results. For this, we studied the diversity within  $E$ , using the Rand Index (RI) similarity measure [16], and created a more effective sub-set of cluster solutions to represent the new ensemble, denoted here as  $E'$ . This new ensemble was obtained by pruning out significantly inconsistent partitions as well as identical or closely-similar partitions.

Given clustering solutions  $P_i$  in the original ensemble  $E$ , in order to decide whether  $P_i$  is included in  $E'$ , we measure how well  $P_i$  agrees with each of the clustering solutions ( $P_j$ ) contained in  $E$ , where  $i = 1, \dots, q$ , as follows:

$$\text{similarity}(P_i, E) = \frac{1}{q-1} \sum_{j=1}^q \text{RI}(P_i, P_j), \quad (1)$$

where  $(P_i, P_j \in E)$  and  $(i \neq j)$ . The RI counts the pairs of points (in our case superpixel pairs) on which two clusterings agree or disagree and it is computed as:

$$\text{RI}(P_i, P_j) = \frac{TP + TN}{TP + FP + TN + FN}, \quad (2)$$

where  $TP$  and  $TN$  are the number of pairs correctly grouped in the same, and different clusters, respectively.  $FP$  is the number of dissimilar pairs assigned to the same cluster and  $FN$  is the number of similar pairs grouped in different clusters. The RI lies between 0 and 1, where 1 implies the two partitions agreeing perfectly and 0 that they completely disagree. We defined two thresholds  $T_1$  and  $T_2$  that correspond to the minimum and maximum accepted levels of diversity among the partitions. If  $P_i$  exhibits an acceptable level of diversity with respect to the rest of the population in  $E$  (i.e.  $\text{similarity}(P_i, E) \geq T_1$  and  $\text{similarity}(P_i, E) \leq T_2$ ) then it is considered as an eligible voter and is added to the new ensemble  $E'$ . If the opposite applies then the partition is excluded from the new ensemble. The total number of selected partitions in  $E'$  is denoted here as  $q'$ , where  $q' \leq q$ .  $E'$  is formed as follows,

$$E' = \{P_i \mid \text{similarity}(P_i, E) \in [T_1, T_2]\}. \quad (3)$$

The next step consists of finding the consensual partition, denoted here as  $P^*$ , based on the information contained in  $E'$ . For this we use two consensus functions described below.

**Evidence Accumulation Consensus (EAC) Function [5].** This method, denoted here as EAC-CC, considers the co-occurrences of pairs of patterns in the same cluster as votes for their association. In particular, the algorithm maps the  $q'$  partitions in  $E'$  into an  $n \times n$  co-association matrix  $M$ . Each entry in  $M$  is defined as  $M_{ij} = u_{ij}/q'$ , where  $u_{ij}$  is the number of times the pattern pair  $(i, j)$  are grouped together in the same cluster among the  $q'$  partitions. The more frequent a pair of objects appear in the same clusters, the more similar they are. Note that  $M$  is needed here because of the label correspondence problem occurring among partitions of  $E'$ .  $M$  can now be viewed as a new similarity measure among the data patterns and it comprises real numbers ranging from 1 (perfect consensus among partitions) down to 0 (no association). The consensus cluster  $P^*$  is obtained by applying an appropriate similarity-based clustering algorithm on  $M$  (e.g. the hierarchical agglomerative clustering algorithm [17]). The final clustering output here ( $P^*$ ) is represented in another image, namely  $S'$ . Although the interpretation of the results of the EAC are intuitive, it has a quadratic complexity in the number of patterns,  $O(n^2)$ .

**Voting-Based Consensus Function.** This method, denoted here as Vote-CC, utilizes a majority voting technique to find the  $P^*$  that optimally summarizes  $E'$ , first, however, it solves the problem of labelling correspondence among different partitions in  $E'$ . Here we propose a simple re-labelling algorithm using imaging processing tools to match the symbolic cluster labels between the different partitions in  $E'$ . The method finds the optimal re-labelling of a given partition  $P$  with respect to a reference fixed partition  $P'$ .  $P'$  is selected from  $E'$  as the one with highest RI with respect to the ensemble (see Eq. 1).

As we are dealing with images, we first assign the labels resulting from the  $P'$  and  $P$  to the corresponding regions (or supervoxels in this case) located in the binary segmented image  $S$ . The labelled regions are displayed in  $K$  unique colours in two images denoted here as  $IMG'$  and  $IMG$  for  $P'$  and  $P$ , respectively. For example, supervoxels with cluster assignments of '1', '2' and '3' in  $P$  will be represented in  $IMG$  as blue, red and green, respectively. We assume that the number of clusters ranges from 1 to  $K$  and the partitions in  $E'$  group the data (supervoxels) into three clusters (epithelium, stroma and background regions). However, due to the label mismatching problem a pair of correlated clusters from different partitions may be assigned different labels. Our target is therefore to permute the labels, so the cluster labels in  $P$  are in the most likely agreement with the labels in  $P'$ .

To this end, individual clusters displayed in images  $IMG'$  and  $IMG$ , denoted here as  $k_{p'}$  and  $k_p$ , are visualized in two binary images  $IMG'_{k_{p'}}$  and  $IMG_{k_p}$ , respectively. Note that  $k_{p'} \in P'$  and  $k_p \in P$ . The algorithm then estimates the degree of overlapping between  $IMG'_{k_{p'}}$  and  $IMG_{k_p}$ , to assess the similarity between the individual clusters ( $k_{p'}$  and  $k_p$ ). The similarity is obtained using



the Jaccard Index (JI) [16], defined as the ratio between the pixel-counts of the intersection and union of  $IMG'_{k_{p'}}$  and  $IMG_{k_p}$  as follows:

$$JI_{(IMG'_{k_{p'}}, IMG_{k_p})} = \frac{|IMG'_{k_{p'}} \cap IMG_{k_p}|}{|IMG'_{k_{p'}} \cup IMG_{k_p}|}, \quad (4)$$

JI values range from 0 (denoting no matching between  $IMG'_{k_{p'}}$  and  $IMG_{k_p}$ , and hence between  $k_{p'}$  and  $k_p$ ) to 1 (denoting perfect matching). For every label  $k_{p'} \in P'$  we compute  $JI_{(IMG'_{k_{p'}}, IMG_{k_p})}$  obtained against all  $k_p \in P$ . Then, we find the maximum JI value which gives the most similar cluster in  $P$  to  $k'_{p'}$ . If  $k'_{p'}$  and its highest similar  $k_p$  have different labels then the match is achieved by swapping the labels in the original image  $IMG$  and therefore the labels in  $P$ . The procedure then stores the swapped labels as well as their corresponding JI in two variables. These are needed in order to track whether a label pair of  $(k_p, k_{p'})$  has already been swapped in a previous iteration. If true, then swapping  $k_{p'}$  and  $k_p$  is only performed if they have higher JI value than before (i.e. the swapped pair of  $(k_p, k_{p'})$ ). The process is repeated until all labels in  $IMG$  have been inspected against the ones in  $IMG'$ , and therefore clusters in  $P$  are matched with  $P'$ . Note that  $P'$  remains unchanged throughout the re-labelling process. The procedure is summarized in Algorithm 1 and it has a complexity of  $O(K^3)$ . The now aligned labels for all the partitions are combined into a final consensus partition  $P^*$  via a majority voting technique. In exceptional cases, where the number of votes are equal we select the vote of the partitions that produce the highest total similarity (RI) with respect to the ensemble  $E'$  (Eq. (1)). As before,  $P^*$  will be represented in image  $S'$ .

The idea of cluster re-labelling based on a similarity assessment has been proposed before in relation to voting-based consensus methods. However, those approaches are implemented based on inspection of the labels of data points (i.e. samples as abstract objects with no shape or size) while our re-labelling captures the similarity in a different way, based on the overlap of the superpixels, which in turn represent image regions with their own shapes and sizes.

## 4 Experiments and Evaluation

The effectiveness of the proposed methodology—CC applied to superpixel-based segmentation—was evaluated in the context of clustering accuracy obtained against five standard clustering approaches: (1) **k-means** [12], a centroid based algorithm, (2) **Unsupervised Learning Vector Quantization (LVQ)** [18], LVQ algorithm for unsupervised learning, (3) **EM** [8], a distribution based method (4) **Make Density Based (MDB)** [19], a density based algorithm, and (5) **Agglomerative Hierarchical Clustering (AH)** [17], a pairwise distance based approach. These algorithms were chosen to include a range of different clustering strategies to ensure diversity in the ensemble.

All imaging procedures and machine learning algorithms were implemented on the ImageJ platform [20] using the WEKA data mining JAVA libraries [21]

**Algorithm 1: Label matching algorithm for the Vote-CC method**

**Input:**  $P, P', S, n, K$

**Output:** Labels matched for  $P$  with respect to  $P'$

```

for (s=1 to n) do
    Assign label of  $P'_s$  to superpixel  $S_s$  and save in image  $IMG'$ 
    Assign label of  $P_s$  to superpixel  $S_s$  and save in image  $IMG$ 
end for
for ( $k'_p=1$  to  $K$ ) do
    for ( $k_p=1$  to  $K$ ) do
        Threshold  $k'_p$  in  $IMG'$ , convert to mask and save in  $IMG'_{k'_p}$ 
        Threshold  $k_p$  in  $IMG$ , convert to mask and save in  $IMG_{k_p}$ 
        Compute  $JI(IMG'_{k'_p}, IMG_{k_p})$  using eq. (4)

    end for
     $MaxJI = \max\{JI(IMG'_{k'_p}, IMG_{k_p}), \text{where } k_p = \{1 \dots K\}\}$ 

    if ( $k_{p'} \neq k_p$ )
        SwappedLabels = ( $k_{p'}, k_p$ )
         $JISwappedLabels_{(k_{p'}, k_p)} = MaxJI$ 
        if ( $k_p, k_{p'} \notin SwappedLabels$ )
            Swap  $k_{p'}$  and  $k_p$  and save result in  $IMG$ 
        else if ( $k_p, k_{p'} \in SwappedLabels$  and  $JISwappedLabels_{(k_{p'}, k_p)} > JISwappedLabels_{(k_p, k_{p'})}$ )
            Swap  $k_{p'}$  and  $k_p$  and save result in  $IMG$ 
        end if
    end if
end for
Assign the new labels in  $IMG$  to partition  $P$ 

```

running on an Intel R core(TM) i7-4790 CPU running at 3.60 GHZ, with 32 GB of RAM and 64-bit Linux operating system. All the algorithms were quantitatively evaluated by comparing their results with forty five gold-standard H&E stained images (denoted here as  $R$ ) from oropharyngeal cancer TMAs. A set of  $R$  images were obtained by manually labelling them into epithelium, stroma and background areas by one of us (GL) with a background in Oral Pathology.

We used three well-known clustering measures [16] to evaluate the algorithm results: (1) **The Rand Index (RI)** was used to compare the final consensus clustering solution given in image  $S'$  with their corresponding reference partition given in the gold-standard image  $R$  and it is estimated as,  $RI(S', R)$  (see Eq. (2)), where  $TP, TN, FP,$  or  $FN$  were calculated by considering the overlapping superpixels of  $S'$  and  $R$  (as explained before). (2) **F1-score** that is defined as:

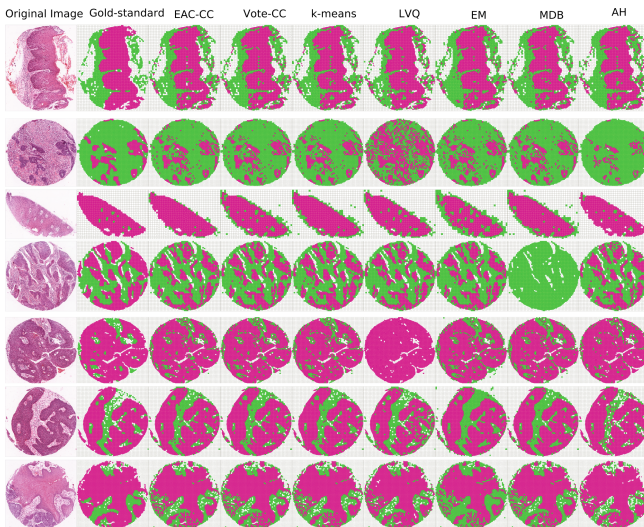
$$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \quad (3) \text{ Jaccard Index (JI)} \text{ that is defined as: } JI = \frac{|S' \cap R|}{|S' \cup R|},$$

In all experiments, (hyper)parameters of jSLIC and CC methods were tuned via a cross-validation procedure on a training set of ten additional images. For the supervoxel segmentation, the regularisation parameter  $z$  and the initial supervoxel size  $m$  were tuned over the values of (0.2, 0.3, 0.4) and (40, 50, 60), respectively. We found that the optimal values were at 0.3 and 60 for  $z$  and  $m$ , respectively. The number of clusters was fixed to three in all experiments, corresponding to three most distinct types of content: epithelium, stroma and background regions. The ensemble of cluster solutions was generated by running the five aforementioned clustering algorithms multiple times with various para-

meter settings. The number of seeds in  $k$ -means and EM algorithms were chosen randomly from the range [10, 300]. Learning rates in the LVQ algorithm were set at the values of 0.05, 0.07, 0.09, 0.1 and 0.3. The AH algorithm was used with Complete and Mean link types. The ensemble generation process yielded a total of thirty one clustering solutions, stored in the initial pool of cluster solutions,  $E$ . The diversity selection strategy was applied to form another better performing ensemble  $E'$ . For this, we assigned the values of 0.5 and 0.9 to the diversity acceptance thresholds  $T_1$  and  $T_2$ , respectively.

Table 1 presents a quantitative comparison of the EAC-CC and Vote-CC methods with five individual clustering approaches (mentioned above). For each of the individual clustering algorithms, we selected the result of the best performing run (out of the multiple runs) then we evaluated its mean RI, F1-score, JI and the standard deviations across the forty five images. Figure 1 provides a visual comparison of our output against the clustering methods. For display purposes we randomly selected one clustering output (out of the multiple runs) to represent the performance of the individual clustering approaches.

The results show that EAC-CC and Vote-CC following jSLIC segmentation produce the most accurate results out of the individual clusterings tested (81% and 82%, respectively). The accuracy of the Vote-CC comes very close to the one in EAC-CC. However, Vote-CC significantly outperformed EAC-CC in execution time. This is due to EAC-CC having a large complexity of the order  $O(n^2)$  (in our case,  $n$  reached



**Fig. 1.** Examples of tissue regions detection in seven H&E images. From left, the original image, gold-standard, EAC-CC, Vote-CC and the individual clustering methods after superpixel segmentation. Black, white, magenta and green colours correspond to the segmentation lines, background, epithelium and stroma regions, respectively. (Color figure online)

**Table 1.** Performance evaluation of the EAC-CC and Vote-CC frameworks compared against five individual clustering approaches in terms of mean RI, F1-score and JI along with standard deviations ( $\pm$ ) across the forty five images. The best results (Vote-CC method) are marked in **bold font**.

Measure	EAC-CC	Vote-CC	$k$ -means	LVQ	EM	MDB	AH
RI( $\pm$ )	0.81 $\pm(0.05)$	<b>0.82</b> $\pm(\mathbf{0.05})$	0.78 $\pm(0.15)$	0.77 $\pm(0.11)$	0.76 $\pm(0.16)$	0.79 $\pm(0.07)$	0.72 $\pm(0.12)$
F1-score( $\pm$ )	0.74 $\pm(0.08)$	<b>0.75</b> $\pm(\mathbf{0.09})$	0.71 $\pm(0.15)$	0.72 $\pm(0.12)$	0.69 $\pm(0.14)$	0.73 $\pm(0.09)$	0.68 $\pm(0.15)$
JI( $\pm$ )	0.71 $\pm(0.10)$	<b>0.72</b> $\pm(\mathbf{0.10})$	0.69 $\pm(0.18)$	0.61 $\pm(0.12)$	0.66 $\pm(0.20)$	0.68 $\pm(0.13)$	0.60 $\pm(0.17)$
Time (ms)	748.95	<b>31.02</b>					

up to 5000 in some images) while the complexity of Vote-CC is  $O(K^3)$  (with  $K = 3$ ). The results also reveal that CC methods result in greater consistency in performance over individual clustering methods as illustrated by lower standard deviations of the RI and F1-scores. This consistency can be seen visually by comparing the results in Fig. 1. In particular, despite the apparent satisfactory clustering results obtained by the single algorithms across most images, they all failed to perform well in some cases (e.g. notice the unstable performance of the LVQ, MDB and AH in the seven examples depicted in Fig. 1).

## 5 Conclusion

We presented a method of tissue segmentation of histopathological images using supervoxels and Consensus Clustering (CC), a combination that, to our knowledge, has not been applied before in quantitative microscopy. Our approach decreases the spatial complexity of images while retaining important information about their contents which are essential for enabling automated pre-screening and guided searches on histopathological imagery. The proposed method performs an unsupervised detection of image regions that correspond to three classes of interest: epithelium, connective and background areas. A supervoxel segmentation was initially performed that was followed by a CC technique which combines the ‘opinions’ of several clustering algorithms into a single, more accurate and robust result. Our work exploited two CC functions, the EAC and the voting-based. For the latter, we introduced a label matching technique which imposed consistency to the different base clustering outcomes. The method is easy to understand and implement and specially tailored for unsupervised imaging segmentations. Qualitative and quantitative results tested on a set of forty five hand-segmented H&E stained tissue images verified that the CC methods outperform the individual clustering approaches in terms of the accuracy of the results and consistency. Furthermore, the voting-base CC using our re-labelling technique outperforms the EAC in terms of execution time.

**Acknowledgments.** This work was supported by the EPSRC through funding under grant EP/M023869/1 “Novel context-based segmentation algorithms for intelligent microscopy”.

## References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels. Technical report, EPFL no. 149300 (2010)
2. Borovec, J.: Fully automatic segmentation of stained histological cuts. In: 17th International Student Conference on Electrical Engineering, Prague, pp. 1–7 (2013)
3. Vega-Pons, S., Ruiz-Shulcloper, J.: A survey of clustering ensemble algorithms. *Int. J. Pattern Recogn.* **25**(03), 337–372 (2011)
4. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012)
5. Fred, A.L., Jain, A.K.: Combining multiple clusterings using evidence accumulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(6), 835–850 (2005)
6. Topchy, A.P., Law, M.H.C., Jain, A.K., Fred, A.L.: Analysis of consensus partition in cluster ensemble. In: 4th IEEE International Conference on Data Mining, pp. 225–232 (2004)
7. Dudoit, S., Fridlyand, J.: Bagging to improve the accuracy of a clustering procedure. *Bioinformatics* **19**(9), 1090–1099 (2003)
8. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B* **39**(1), 1–38 (1977)
9. Simsek, A.C., Tosun, A.B., Aykanat, C., Sokmensuer, C., Gunduz-Demir, C.: Multilevel segmentation of histopathological images using cooccurrence of tissue objects. *IEEE Trans. Biomed. Eng.* **59**(6), 1681–1690 (2012)
10. Khan, A.M., El-Daly, H., Rajpoot, N.: RanPEC: random projections with ensemble clustering for segmentation of tumor areas in breast histology images. In: Medical Image Understanding and Analysis (MIUA), Swansea, pp. 17–23 (2012)
11. Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Hist.* **23**(4), 291–299 (2001)
12. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: a k-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **28**(1), 100–108 (1979)
13. Borovec, J., Kybic, J.: jSLIC: superpixels in ImageJ. In: Computer Vision Winter Workshop, Praha (2014)
14. The jSLIC-superpixels plugin. [http://imagej.net/CMP-BIA\\_tools](http://imagej.net/CMP-BIA_tools)
15. Hadjitodorov, S.T., Kuncheva, L.I., Todorova, L.P.: Moderate diversity for better cluster ensembles. *Inf. Fusion.* **7**(3), 264–275 (2006)
16. Hubert, L., Arabie, P.: Comparing partitions. *J. Classif.* **2**(1), 193–218 (1985)
17. Defays, D.: An efficient algorithm for a complete link method. *Comput. J.* **20**(4), 364–366 (1977)
18. Kohonen, T.: Learning vector quantization. In: The Handbook of Brain Theory and Neural Networks, 2nd edn, pp. 631–634. MIT Press (2003)
19. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: 2nd International Conference on Knowledge Discovery and Data Mining, pp. 226–231, August 1996
20. Rasband, W.S.: ImageJ. US National Institutes of Health, Bethesda, Maryland, USA (1997). <http://imagej.nih.gov/ij/>
21. Frank, E., Hall, M., Witten, I.H.: The WEKA Workbench, 4th edn. Morgan Kaufmann, Burlington (2016)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

