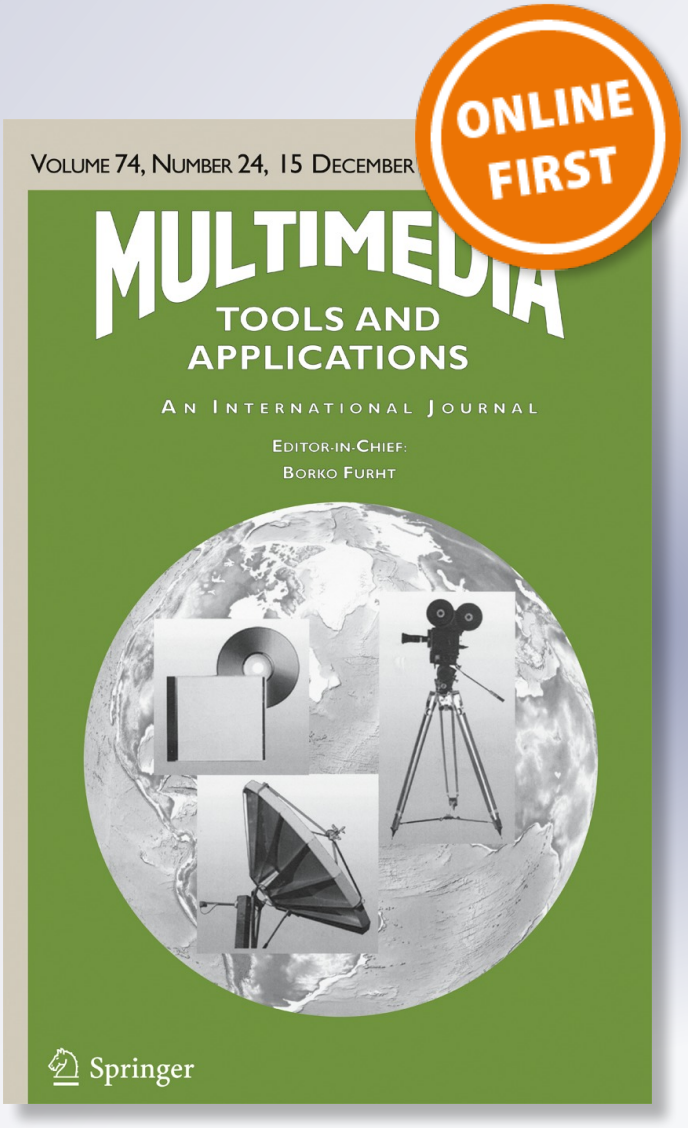
# *Segmentation data visualizing and clustering*

**Ayman Khlif & Max Mignotte**

**Multimedia Tools and Applications**
An International Journal

ISSN 1380-7501

Multimed Tools Appl
DOI 10.1007/s11042-015-3148-6




Springer

Multimed Tools Appl
DOI 10.1007/s11042-015-3148-6

# Segmentation data visualizing and clustering

**Ayman Khlif**[1] · **Max Mignotte**[1]

Received: 22 May 2015 / Revised: 14 October 2015 / Accepted: 7 December 2015


**Abstract** Browsing, searching and retrieving images from large databases based on low level color or texture visual features have been widely studied in recent years but are also often limited in terms of usefulness. In this paper, we propose a new framework that allows users to effectively browse and search in large image database based on their segmentation-based descriptive content and, more precisely, based on the geometrical layout and shapes of the different objects detected and segmented in the scene. This descriptive information, provided at a higher level of abstraction, can be a significant and complementary information which helps the user to browse through the collection in an intuitive and efficient manner. In addition, we study and discuss various ways and tools for efficiently clustering or for retrieving a specific subset or class of images in terms of segmentation-based descriptive content which can also be used to efficiently summarize the content of the image database. Experiments conducted on the Berkeley Segmentation Datasets show that this new framework can be effective in supporting image browsing and retrieval tasks.

**Keywords** Berkeley dataset · Clustering algorithm · Entropy · Database browsing and retrieving images · Hierarchical clustering · $K$-means · Multidimensional visualization · Query-by-drawing · Segmentation data clustering · Descriptive content based image classification · Variation of information · Visualization of image databases.

## 1 Introduction

Clustering is the task of grouping together, in a feature space, data samples in the same group (or cluster) that are similar in the sense of a given distance measure. It is the main task

✉ Max Mignotte
mignotte@iro.umontreal.ca

[1] Département d'Informatique et de Recherche Opérationnelle (DIRO), Université de Montréal, Faculté des Arts et des Sciences, Montréal H3C 3J7 QC, Canada

Published online: 16 December 2015

of data exploration and analysis which is then useful for looking for patterns or structures (and/or correlations) in the data that are of interest.

Image segmentation is a special case of clustering problem where the grouping of image data (or pixels) into clusters must take into account not only their similarity in the feature space but also the requirement of their spatial coherence, since an essential feature of image data is the spatial ordering of its pixels. In the image segmentation case, structures of interest are spatially coherent regions such as consistent parts of objects or of the background sharing similar attributes. The detection and localization of these regions, with a spatial clustering, allows to change and simplify the representation of the image into something that is both compact, and also easier to analyze since this economic description, in terms of detecting homogeneous regions, efficiently describes the geometric content of the input image.

In these two latter cases, the clustering performance and its subsequent usefulness closely depend on the used distance measure. In this sense, depending on this similarity metric, the application herein proposed, consisting of achieving an unsupervised clustering of segmentation results (estimated, for example, from an image database) could either be meaningless or have a certain interest. Indeed, in such application, it is crucial to consider an appropriate distance, across the lattice of possible spatial clusterings, for efficiently comparing two segmentation results in a objective, reliable, comprehensible and perceptual criterion sense which should be also capable of taking into account the inherent variability of each possible perceptually consistent interpretation/segmentation of an input image (possibly segmented at different detail levels by different human segmenter). Nevertheless, it is not trivial to find a true (in the mathematical sense) and meaningful distance between two segmentation maps. Indeed, the two segmentations might have a different number of segments and the correspondence between segments are not known and this is especially true in the case of segmentations obtained from two different images. If such a distance exists, an unsupervised clustering of segmentation results could be efficiently used for obtaining an overview or for browsing and/or retrieving images from large image collections or database or for retrieving a specific subset or class of images in terms of segmentation-based descriptive content (such as the layout and similar arrangement of the different objects segmented in the image) and not in terms of low-level features such as color and/or texture as it is commonly used in image database browsing applications. Moreover, a distance-based clustering can also be exploited, among other things, in order to build an image database with a substantial amount of diversity in the dataset. This can be easily achieved by automatically removing images with to much similar content in terms of region-based descriptive content or to quantify this amount of diversity with a dispersion measure such as the average of all distances obtained for each segmentation pair.

In addition, an automatic clustering procedure often requires the estimation of the center of each cluster. It is the case of the $K$-means clustering procedure which is the most commonly used clustering algorithm so far proposed in the literature [11] or more generally, also the case of all class of (distance-based) clustering algorithms known as iterative refinement algorithms. To this end, the estimation of cluster centers, in a non-Euclidean and non-standard distance sense, when the underlying data are segmentation results, is far from being trivial. Fortunately, this problem has been recently solved in the image processing community in order to provide an interesting alternative to the existing complex and computationally costly segmentation models. Indeed, an effective, simple and commonly accepted segmentation strategy consists in averaging (i.e., efficiently combining or fusing) multiple quickly estimated segmentation results (of the same scene) obtained from some

simple segmentation models (or by the same segmentation algorithm with different values of its internal parameters) to achieve a final improved segmentation result. This strategy has initially been introduced in [9, 10] with the restriction that all input segmentations (to be averaged or fused) should contain the same number of regions and then a little later without this restriction, with an arbitrary number of regions [6, 15–17, 31] in different criteria senses with algebraic, analytic or stochastic procedures. In the case of unsupervised clustering of segmentation results, these cluster centers or prototypes could be efficiently used to quickly visualize the region-based descriptive content of each class of segmentation map ensemble associated to a given image database. Also, these cluster centers could estimate a *clustering meaningfulness* distance allowing to quantify the diversity in the dataset or the performance of each clustering algorithm, among others things. It is true that the $K$-means procedure and more generally, all iterative refinement clustering algorithms have also a major shortcoming in the fact that the number of clusters must be specified in advance.

Finally, it is worth mentioning that, there is no work, reported in the literature (to our knowledge), which proposes and exploits, for several visualizing or browsing applications, an automatic clustering result of segmentation maps estimated from an image database. In the context of browsing large image collections, which is a non trivial task, mainly due to the semantic gap existing between the user subjective notion of similarity and the one according to which a browsing system organizes the images, we can cite several existing techniques. A possible strategy consists of the use of a PCA or MDS mapping-based visualization technique for grouping similar (in terms of color and/or texture-based features) images together on a plane [26], eventually projected onto a sphere [27], on a hierarchical (browsing) tree which can be customized according to user preferences [4], or other graphs (a good survey of existing strategies can be found in [7]). Higher level strategies are proposed for example in [12] where a browsing model integrates high level semantic concepts which allows to help users to narrow a search domain rather than to browse the whole collection [12] or in [29] in which the underlying idea is to mine and interpret the information from the user's interaction in order to understand the user's needs. Based on the Dempster-Shafer theory of evidence and the combination of color and text features, the system's interpretation is used for suggesting new relevant images to the user.

The remainder of this paper is divided into the following sections: First, the proposed distance measure, defined on the space of clusterings, for efficiently comparing two segmentation results, is presented in Section 2.1. The averaging procedure of segmentation results used for estimating the cluster centers based on this aforementioned distance is recalled in Section 2.2. Section 3 shows a variety of applications. Finally Section 4 concludes this paper.

## 2 Distance-based clustering components

### 2.1 Used distance

The variation of information (VoI) metric [13, 14] is a recent information theory based measure for comparing the similarity of two segmentation results (or clusterings). This metric quantifies the information shared between two segmentations by, more precisely, measuring the amount of information that is lost or gained in changing from one segmentation to another. Equivalently (and conceptually), it also represents roughly the amount of

randomness in one segmentation which cannot be explained by the other [13]. The VoI is a true metric on the space of clusterings which is positive, symmetric and obeys the triangle inequality [14]. This VoI metric is currently exploited as a clustering or segmentation quality metric that measures the agreement of the segmentation result with a given ground truth. To this end, it was recently used in image segmentation [18–21, 25, 32] as a quantitative and perceptually interesting measure to compare automatic segmentation of an image to a ground truth segmentation (e.g., a manually hand-segmented image given by an expert) and/or to objectively evaluate the efficiency of several unsupervised segmentation methods.

Let $S^{\mathrm{A}} = \left\{C_1^{\mathrm{A}}, C_2^{\mathrm{A}}, \ldots, C_{R^{\mathrm{A}}}^{\mathrm{A}}\right\}$ and $S^{\mathrm{B}} = \left\{C_1^{\mathrm{B}}, C_2^{\mathrm{B}}, \ldots, C_{R^{\mathrm{B}}}^{\mathrm{B}}\right\}$ be respectively the first and second segmentation (or the segmentation test result to be evaluated and the ground truth segmentation) between which the VoI distance has to be estimated and $R^{\mathrm{A}}$ being the number of regions[1] in $S^{\mathrm{A}}$ and $R^{\mathrm{B}}$ the number of regions in $S^{\mathrm{B}}$. The VoI between $S^{\mathrm{A}}$ and $S^{\mathrm{B}}$ is defined as:

$$\mathrm{VoI}(S^{\mathrm{A}}, S^{\mathrm{B}}) = H(S^{\mathrm{A}}) + H(S^{\mathrm{B}}) - 2 \cdot I(S^{\mathrm{A}}, S^{\mathrm{B}}) \tag{1}$$

where $H(S^{\mathrm{A}})$ and $H(S^{\mathrm{B}})$ denote respectively the classical entropy associated with the segmentation $S^{\mathrm{A}}$ and $S^{\mathrm{B}}$ and $I(S^{\mathrm{A}}, S^{\mathrm{B}})$ the mutual information between these two segmentations. Let $n$ be the number of pixels within the image, $n_i^{\mathrm{A}}$ the number of pixels in the $i$-th cluster of the segmentation $S^{\mathrm{A}}$, $n_j^{\mathrm{B}}$ the number of pixels in the $j$-th cluster of the segmentation $S^{\mathrm{B}}$ and finally $n_{ij}$ the number of pixels which are together in the $i$-th cluster (or region) of the segmentation $S^{\mathrm{A}}$ and in the $j$-th cluster of the segmentation $S^{\mathrm{B}}$. The entropy is always non-negative (it takes a value of 0 only when there is no uncertainty, namely when there is only one cluster) and is defined as:

$$\begin{aligned} H(S^{\mathrm{A}}) &= -\sum_{i=1}^{R^{\mathrm{A}}} P(i)\, \log P(i) = -\sum_{i=1}^{R^{\mathrm{A}}} \frac{n_i^{\mathrm{A}}}{n}\, \log \frac{n_i^{\mathrm{A}}}{n} \\ H(S^{\mathrm{B}}) &= -\sum_{i=1}^{R^{\mathrm{B}}} P(j)\, \log P(j) = -\sum_{i=1}^{R^{\mathrm{B}}} \frac{n_j^{\mathrm{B}}}{n}\, \log \frac{n_j^{\mathrm{B}}}{n} \end{aligned} \tag{2}$$

with $P(i) = n_i^A/n$ being the probability that a pixel belongs to cluster $S^{\mathrm{A}}$ (respectively $P(j) = n_j/n$ being the probability that a pixel belongs to cluster $S^{\mathrm{B}}$) in the case where $i$ and $j$ represent two discrete random variables taking respectively $R^{\mathrm{A}}$ and $R^{\mathrm{B}}$ values and uniquely associated to the partition $S^{\mathrm{A}}$ and $S^{\mathrm{B}}$. Let now $P(i, j) = n_{ij}/n$ represents the probability that a pixel belongs to $C_i^{\mathrm{A}}$ and to $C_j^{\mathrm{B}}$, the mutual information $I(.)$ between the partitions $S^{\mathrm{A}}$ and $S^{\mathrm{B}}$ is equal to the mutual information between the random variables $i$ and $j$ and is expressed in the following way:

$$I(S^{\mathrm{A}}, S^{\mathrm{B}}) = \sum_{i}^{R^{\mathrm{A}}} \sum_{j}^{R^{\mathrm{B}}} P(i, j)\, \log \frac{P(i, j)}{P(i)\, P(j)} \tag{3}$$

The VoI is a true metric across the lattice of possible clusterings (taking a value of 0 when two clusterings are identical and positive otherwise) and is bounded by $\log n$. However, if $S^{\mathrm{A}}$ and $S^{\mathrm{B}}$ have at most $R^{\max}$ clusters (i.e., regions), it is bounded by $2 \log R^{\max}$ [14]. Let us also note that if we have several possible ground truth segmentations for a same

[1] A region is a set of connected pixels belonging to the same class and a class, a set of pixels possessing similar textural characteristics.

scene (which could be possibly segmented at different levels of details by different human segmenters), this measure is able to take into account the inherent variability of possible interpretations between each human observer and more precisely the inherent variability of each possible perceptually consistent interpretation/segmentation of an input image by a simple averaging technique [32]. This variability is also due to the fact that the image segmentation problem is inherently ill-posed (and consequently, this problem has multiple solutions notably for the different possible values of the number of classes not known *a priori*) [17].

### 2.2 Cluster center estimation

Let $\{S_k\}_{k\leq L}$ be a finite ensemble of $L$ segmentations $\{S_k\}_{k\leq L} = \{S_1, S_2, \ldots, S_L\}$ existing in a given cluster. The estimation of the center of these $L$ segmentations (also called the cluster prototype or the consensus segmentation) can be efficiently achieved in the VoI distance sense (see Section 2.1) by the solution of the following optimization (or so-called *median partition* [30]) problem:

$$\begin{aligned}\hat{S}_{\overline{\mathrm{VoI}}} &= \arg\min_{S\in\mathcal{S}_n} \overline{\mathrm{VoI}}\,(S, \{S_k\}_{k\leq L}) \\ &= \arg\min_{S\in\mathcal{S}_n} \frac{1}{L}\sum_{k=1}^{L} \mathrm{VoI}\,(S, S_k) \end{aligned} \tag{4}$$

with $\mathcal{S}_n$ is the set of all possible segmentations using $n$ pixels. Herein, each estimation of the center of a given cluster (i.e., the best compromise segmentation solution resulting in a consensus in terms of contour accuracy or detail level displayed by each segmentations in $\{S_k\}_{k\leq L}$), thus appears as the segmentation solution which minimizes the average pairwise VoI distance between all elements of the cluster. Equivalently, this partition solution can be expressed as the result of a minimization problem on a consensus function (using the $\overline{\mathrm{VoI}}$ distance) which can be solved with a steepest local energy descent procedure [17]. In this iterative minimization procedure, a new label $x$ is assigned to pixel $s$ (initially with label $l_s$), if this pixel is connected to the $x$-th region and if the local decrease in the energy function $\overline{\mathrm{VoI}}(.)_{s:l_s\to x}$ is positive with:

$$\begin{aligned}\Delta\overline{\mathrm{VoI}}\left(\hat{S}^{[p]}_{\overline{\mathrm{VoI}}}, \{S_k\}_{k\leq L}\right)_{s:m\to x} = \\ L\cdot\Bigg\{-\frac{n_m}{n}\log\left(\frac{n_m}{n}\right) - \frac{n_x}{n}\log\left(\frac{n_x}{n}\right) \\ +\frac{(n_m-1)}{n}\log\left(\frac{n_m-1}{n}\right) + \frac{(n_x+1)}{n}\log\left(\frac{n_x+1}{n}\right)\Bigg\} \\ -2\cdot\sum\nolimits_{l=1}^{L}\Bigg\{\frac{n_{m,\mathcal{L}_s^l}}{n}\log\left(\frac{n_{m\mathcal{L}_s^l}}{n}\cdot\frac{n}{n_m}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right) \\ +\frac{n_{x\mathcal{L}_s^l}}{n}\log\left(\frac{n_{x\mathcal{L}_s^l}}{n}\cdot\frac{n}{n_x}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right) \\ -\frac{(n_{m\mathcal{L}_s^l}-1)}{n}\log\left(\frac{(n_{m\mathcal{L}_s^l}-1)}{n}\cdot\frac{n}{(n_m-1)}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right) \\ -\frac{(n_{x\mathcal{L}_s^l}+1)}{n}\log\left(\frac{(n_{x\mathcal{L}_s^l}+1)}{n}\cdot\frac{n}{(n_x+1)}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right)\Bigg\}\end{aligned} \tag{5}$$

where $\mathcal{L}_s^l$ denotes the label at site $s$ of the $l$-th segmentations ($l \leq L$) of the segmentation ensemble $\{S_k\}_{k\leq L}$ and we recall that $n_{m\mathcal{L}_s^l}$ designates the number of pixels which are together in the $m$-th cluster (or region) of the segmentation $S$ and in the $\mathcal{L}_s^l$-th cluster of the segmentation $S_l \in \{S_k\}_{k\leq L}$ (see Algorithm 1).

**Algorithm 1**
**VoI-Based Mean Segmentation Estimation**

| | |
|---|---|
| $\overline{\text{VoI}}$ | Mean VoI (See Equation (4)) |
| $\{S_k\}_{k\leq L}$ | Set of $L$ segmentations to be averaged |
| $T_{\max}$ | Maximal number of iterations |

**1. Initialization**

$$\hat{S}_{\overline{\text{VoI}}}^{[0]} = \arg\min_{S\in\{S_k\}_{k\leq L}} \overline{\text{VoI}}\,(S, \{S_k\}_{k\leq L})$$

or

$\hat{S}_{\overline{\text{VoI}}}^{[0]} \leftarrow$ image divided into $K(=6) \neq$ rectang. regions

**2. Steepest Local Energy Descent**

**while** $p < T_{max}$ **do**

**for** *each pixel with label* $l_s$ *at site* $s$ **do**

- Let $\mathcal{E}$ the set of labels $\neq l_s$ contained in the local (squared) fixed-size ($N_w = 7$) neighborhood of $s$
- Draw a new label $x$ according to the uniform distribution in the set $\mathcal{E}$
- if $x = \emptyset$ **then continue;**
- if pixel $s$ with label $x$ is not 4-connected with the $x$-th region in $\hat{S}_{\overline{\text{VoI}}}^{[p]}$ **then continue;**
- Compute $\Delta\overline{\text{VoI}}\left(\hat{S}_{\overline{\text{VoI}}}^{[p]}, \{S_k\}_{k\leq L}\right)_{s:l_s\rightarrow x}$ (See Equation (5))
- **If** $\Delta\overline{\text{VoI}}\left(\hat{S}_{\overline{\text{VoI}}}^{[p]}, \{S_k\}_{k\leq L}\right)_{s:l_s\rightarrow x} > 0$ Then replace label $l_s$ by label $x$ at site $s$

$p \leftarrow p + 1$

As initialization of this steepest gradient descent, we can start from the segmentation result (among the $L$ segmentation results to be averaged), ensuring the minimal consensus energy in the $\overline{\text{VoI}}$ sense [17]. Another strategy consists in initializing the gradient procedure from a synthetic image spatially divided by $K$ horizontal or vertical rectangles with $K$ different labels and to take, at convergence, the segmentation solution ensuring the minimal consensus energy. In order to improve the convergence, we also use a multiresolution approach by considering the optimization problem at a lower resolution level (with the downsampling of $\{S_k\}_{k\leq L}$ by a scale factor $c$). After convergence of the gradient, the result obtained at this lower resolution level is interpolated and then used as initialization for the

gradient procedure at the full resolution level. This strategy drastically reduces the complexity and computational effort and provides an accelerated convergence toward improved estimate (as noticed, for example, in other energy-based models [21–23]).

## 3 Examples of applications

### 3.1 Visualization of image databases

Due to the huge number of the images in a collection or a database and the limited size of a computer monitor, it may be interesting to find a strategy to provide a quick overview of these images. Generally, it will ensure that these images are displayed as thumbnails and correctly arranged in such a way that the user can quickly and intuitively understand what types of images are contained in the database and their distribution for further analysis. Generally this is typically achieved by mapping-based technique such as principal component analysis or multi-dimensional scaling (MDS) along with a metric between low-level color or texture features and computed for each pair of images of the database. More precisely, MDS is before all, a nonlinear dimensionality reduction technique that attempts to find an embedding from the initial feature vectors in the high dimensional space such that distances (or conceptually, the original relationships of these images in term of a given distance) are preserved in a low dimensional space. The foundational ideas behind MDS were first proposed by Young and Householder [33] and then further developed by Torgerson [28] in which, its original algorithm (called classic MDS) exploits a spectral method which consists of finding embedding coordinates by computing the top eigenvectors of a *double-centered* transformation of the distance matrix (called a Gramian matrix) sorted by decreasing eigenvalue.

Instead of considering low-level color or texture features to arrange these image thumbnails, as it is commonly used in image database browsing or navigation systems, it may be interesting to arrange them according to their descriptive content extracted by a (region-based) segmentation or more precisely, based on the spatial arrangement of the different objects detected or segmented in the image regardless of their own color or texture. This mapping-based visualization technique, made at a higher level of abstraction, is herein possible since the VoI distance (see Section 2.1) is a true metric and also a quantitative and perceptual measure to compare two segmentations, which also inherently takes into account the variability of each possible perceptually consistent interpretation/segmentation of an input image which could be possibly segmented at different detail levels.

The VoI-based true metric allows us to estimate a distance matrix (describing the dissimilarities between each existing pair of segmentations) from which the classic MDS then computes a Gramian matrix having the same properties as the one obtained with a classical Euclidean distance based distance matrix (i.e., positive and semi-definite (PSD)) which then ensures (positive eigenvalues and) a good convergence and accuracy of the MDS algorithm (and more generally of all MDS methods based on eigen-decomposition). As explained in [5], the use of a distance which is not a true metric is somewhat equivalent to considering a noise corrupted version of the Gramian or distance matrix and consequently an inaccurate and unreliable (visualization) mapping.

We first present an MDS image overview of the Berkeley Dataset (BSDS300) [24], based on the VoI distance (see Fig. 1). The BSDS300 consists of 300 color images of size $481 \times 321$. For each color image of, a set of ground truth segmentations, provided by human observers (between 4 and 7), is also available. In our application, we can either exploit the

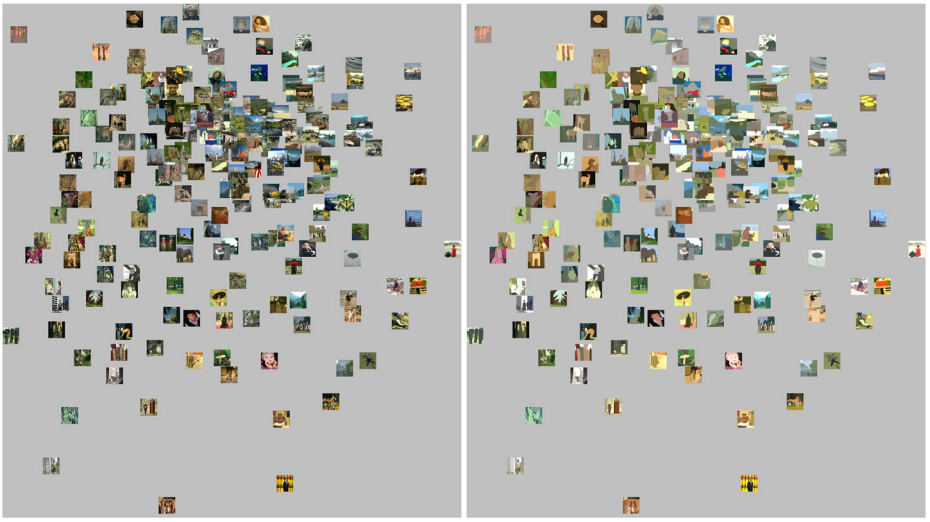


**Fig. 1** A MDS visualization map of the BSDS300 based on the VoI distance between segmentations. Left: view map of the (300) color images mapped according to their similarity in term of region-based segmentation result. Right: view map of the segmentation results according to their similarity in the VoI distance sense

result of a segmentation algorithm or use the ground truth segmentations when these one are, of course, available.[2]

In our case, the ground truth segmentation map with the median value of the number of regions estimated by the set of human observers (amongst 4 and 7) for each image of the BSDS300 has been here exploited.[3] The images are also squared by stretching them in order to get some invariance in the spatial arrangement of the different parts and object shape segmented in the image. The visualization 2D map of the image thumbnails of the images and their associated ground truth segmentation based on the VoI distance computed for each pair of segmentations is shown in Fig. 1. Images (to the left) with a similar layout and arrangement of the different objects detected or segmented in the image regardless of their own color or texture are placed close to each other while images with dissimilarity arrangement are far from each other. In this example, the target output dimension of the MDS-based mapping technique equals to 2 dimensions which is the size of the output mapping. A mapping on a cube would have been also possible with a target output equals to 3 dimensions. It is worth mentioning that we can also efficiently evaluate the reliability of the MDS-based mapping technique (as a function of the target output dimension) by using the correlation-based metric which is simply the correlation of the VoI distance between pairwise segmentations and their corresponding (pairwise) 2D Euclidean distances in the target space [2]. In the absence of perceptual error and thus, with no loss of information in this dimensionality reduction problem, the ideal correlation metric is one. For a 2D MDS visu-

[2]Our approach is tolerant to different types of image degradation (e.g., noise, blur, distortions) insofar as the segmentation method is able to give a segmentation map which is robust enough for these types of degradation.

[3]It is worth mentioning that we could also exploit all the set of ground truth segmentations for each image by computing an average VoI distance computed across all the existing ground truth segmentations.

alization map of the BSDS300 based on the VoI distance between segmentations, we obtain a correlation metric $\rho = 0.856$ which means that there are only 14.4 percent of pairs of images whose 2D Euclidean distance between them does not preserve the monotonicity of the initial VoI distance (i.e., or 14.4 loss of information of this 2D MDS mapping according to the correlation metric). Figures 2 and 3 show respectively some magnified details of Fig. 1 and the 6 nearest neighbors of a given image in the database illustrating well the efficiency of our image mapping method based on either the similar layout and spatial arrangement or (to some extend) the similarities in the geometrical shapes of the objects detected and segmented in the scene (and despite the fact that the BSDS300 exhibit a great diversity of images).

Let us note that we can easily quantify the amount of diversity, in term of our main criterion (layout and similar arrangement of the different objects detected or segmented in the image regardless of their own color or texture) by computing the average over all the distances obtained for each segmentation pair:

$$\mathcal{D} = \frac{\sum_{k,l \;\; k \neq l} \mathrm{VoI}(S_k, S_l)}{N(N-1)} \tag{6}$$

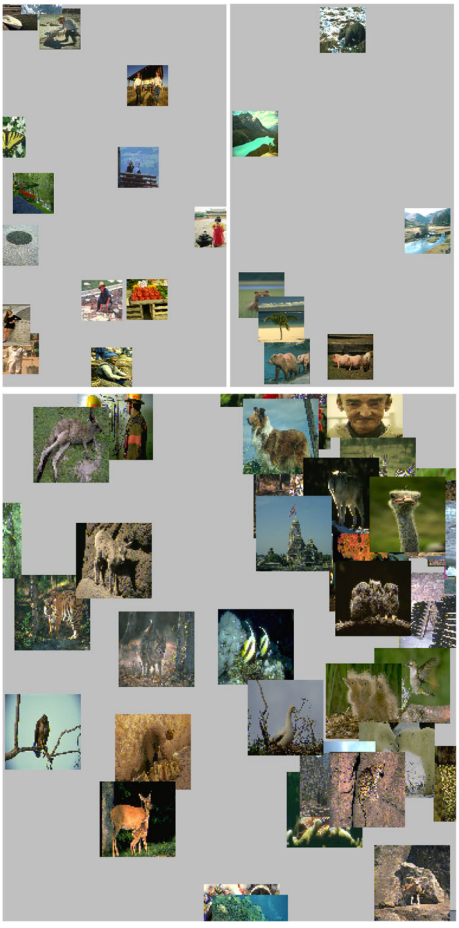


**Fig. 2** Magnified details of Fig. 1 showing the grouping of images of the BSDS300 based on the similar layout and arrangement of the objects detected and segmented in the scene. In lexicographic order, a group of pictures mostly showing one or two people at the center of the image (top left) or showing that the three images of bears of the BSDS300 are close (top right) and finally a group of pictures showing that the six birds of the BSDS300, perched on a branch are relatively close or showing, for the bottom or left part of the image, an animal on a mountainside or located in the middle of the image

where $N$ is the number of images or segmentations and $N(N-1)$ the number of pair of segmentations. This diversity measure $\mathcal{D}$ will be even closer to 1 that the diversity will be better in the image database, according to our visualization criterion. For the BSDS300, we obtain a diversity measure metric $\mathcal{D} = 3.05$. In order to increase the diversity in the image database, in our criterion sense, it consists in searching the two most similar segmentations and removing one of them. To attain this goal, a hierarchical clustering (see Section 3.4) can be exploited.

**Algorithm 2**
**VoI-Based MDS+$K$-means**

| | |
|---|---|
| VoI | VoI distance (See Equation (1)) |
| $\{S_k\}_{k\leq N}$ | Set of $N$ segmentations to be clustered |
| $D_{\text{out}}$ | Output dimension |
| $K$ | Number of classes |
| $T_{\max}$ | Maximal number of iterations |

**Initialization**

**i.** Compute the distance matrix $M_{ij}$ describing the dissimilarities between each pair of segmentations with the VoI distance

**ii.** Based on $M_{ij}$, use the MDS-based dimensionality reduction to estimate a $D_{\text{out}}$-dim. mapping of $\{S_k\}_{k\leq N}$. Let $\{s_k\}_{k\leq N}$ be this mapping

**iii.** Choose $K$ initial cluster centers $c_1^{[1]}, \ldots, c_K^{[1]}$ among the $L$ $D_{\text{out}}$-dim. vectors $\{s_k\}_{k\leq N}$

**1. while** $c_i^{[p+1]} \neq c_i^{[p]}\ \forall i$ **or** *iter. number* $< T_{max}$ **do**

**1.** At the $p^{\text{th}}$ step, assign segmentation $S_k$ $(k \leq N)$ to cluster $i$ if: $\| s_k - c_i^{[p]} \| < \| s_k - c_j^{[p]} \| \quad \forall j \neq i$

**2.** Determine new cluster centers by:

$$c_i^{[p+1]} = \frac{1}{N_i} \sum_{s_m \in Cl_i^{[p]}} s_m$$

where $N_i$ is the number of samples in $Cl_i^{[p]}$

**2.** Determine each new cluster centers, in term of segmentation map (prototype), by Algorithm 1

Figure 4 shows respectively the two images of the BSDS300 which are respectively the closest and the farthest, in the segmentation-based mean VoI distance sense, from the center of the 300 pictures of the BSDS300 with their associated ground-truth segmentation. It is interesting to note that the estimation of the center of the BSDS300, in term of segmentation map (see Algorithm 1), namely; $\arg\min_{S\in\mathcal{S}_n} \overline{\text{VoI}}\,(S, \{S_k\}_{k\leq L})$), is the blank image (exhibiting one segment/region for the entire image) which is not too far from $\arg\min_{S\in\{S_k\}_{k\leq L}} \overline{\text{VoI}}\,(S, \{S_k\}_{k\leq L})$, the segmentation associated to the center of the

BSDS300 (see Fig. 4a at right). This estimation result is also comprehensible, since, in the VoI distance sense, any existing segmentations are also a refinement of this one region segmentation [32].

**Algorithm 3**
**VoI-Based $K$-means**

| | |
|---|---|
| VoI | VoI distance (See Equation (1)) |
| $\{S_k\}_{k\leq N}$ | Set of $N$ segmentations to be clustered |
| $K$ | Number of classes |
| $T_{\max}$ | Maximal number of iterations |

**Initialization** Choose $K$ initial cluster centers $c_1^{[1]}, \ldots, c_K^{[1]}$ among the $N$ segmentations

**while** $c_i^{[p+1]} \neq c_i^{[p]}\ \forall i$ **or** *iteration number* $< T_{max}$ **do**

**1.** At the $p^{\text{th}}$ step, assign $S_k$ $(k \leq N)$, to cluster $i$ if: $\text{VoI}\,(S_k, c_i^{[p]}) < \text{VoI}\,(S_k, c_j^{[p]}) \quad \forall j \neq i$

**2.** Determine new cluster centers by Algorithm 1

### 3.2 Segmentation-based clustered visualization

With the VoI distance defined in Section 2.1, and the estimation procedure of the center of each cluster or ensemble of segmentations presented in Section 2.2, an unsupervised clustering of segmentation results can be efficiently designed for reducing the number of images that are required to be displayed by grouping images with a similar layout and spatial arrangement of the different objects detected or segmented in the image regardless of their own color or texture or for retrieving a specific subset or class of images in terms of (segmentation-based) descriptive content. To this end, several strategies (that we will also evaluate the reliability later) are possible.

The first strategy (Algorithm 2) consists first of using the MDS based dimensionality reduction technique presented in Section 3.1 and then exploiting the reduced data of this mapping in a classical $K$-means algorithm. At the convergence of the $K$-means algorithm, each new cluster center, in term of segmentation map, is estimated by Algorithm 1. The second strategy (Algorithm 3) consists in directly using the non-reduced data, i.e., the segmentation maps along with, at each iteration of the $K$-means algorithm, the estimation of the center (or prototype) of each cluster (or consensus segmentation) until convergence is achieved.

In order to evaluate and compare the reliability of these two above-described clustering strategies, we can estimate the class separability or the Fisher's distance [3] on each clustering result. This distance is simply the within-class inertia divided by the between-class inertia. In our case, this distance is meaningful if this one is computed in the VoI metric sense such as:

$$\mathcal{F} = \frac{\sum_{k=0}^{K} \sum_{S_m \in Cl_k} \text{VoI}(S_m, c_k)}{\sum_{k=0}^{K} \sum_{l=0\ l\neq k}^{K} \text{VoI}(c_k, c_l)} \tag{7}$$

**Fig. 3** 6 nearest neighbors of a given (the leftmost) image belonging to the BSDS300. From top to bottom, group of images mainly showing 1-] a small and elongated animal or object on the grass/sand or in the sky/water. 2-] an animal or a group of animals. 3-] pyramidal or (highly) elongated structures (pyramid or mountain). 4-] wild mammals. 5-] group of men. 6-] small pyramidal structures in land/sea-scape

where $Cl_k$ is the $k$-th cluster, $\{S_k\}$ the set of segmentations to be clustered and $\{c_k\}$ the set of prototype centers. This number $\mathcal{D}$ can be also viewed as a *clustering meaningfulness* metric since it clearly measures just that and we should expect that this value be close to zero for good clustering results. Figure 5 illustrates the Fisher's distance obtained by the two different above-mentioned clustering strategies as a function of $K$, the number of clusters. We can easily notice that the reliability of the two strategies are comparable, in term of

**(a)** **(b)**

**Fig. 4** From left to right, the closest and the farthest image, in the mean VoI distance sense, from the center of the BSDS300 with their associated segmentation. The segmentation map which is the exact center of the BSDS300, as estimated by the Algorithm 1, is the predictable blank image or one region segmentation (which is not included in the BSDS300 but not too far from the segmentation of the leftmost image)

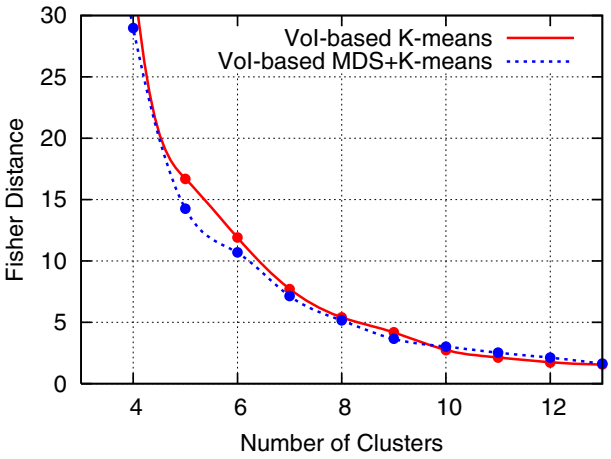



**Fig. 5** Fisher's distance (or cluster separability measure) for the two clustering strategies as a function of the number of clusters

cluster separability. More precisely, slightly better for the VoI-Based $K$-means (Algorithm 3) for $K > 8$ and slightly better for the MDS based mapping (Algorithm 2) for $K < 8$. Computationally speaking, the two algorithms require approximately one minute in order to estimate a partition into a specified number of clusters or classes.

In our case, the $K$ cluster prototypes (centers), (or more simply the $K$ segmentation maps which are the closest of these $K$ prototypes) can be exploited to efficiently summarize the content of the image database and to check, according to this set of cluster prototype segmentations if the database is correctly diversified. Figure 6 shows some cluster prototypes/centers, in term of segmentation maps obtained with the VoI-based $K$-means for different numbers of clusters. Let us note that the optimal number of clusters can be empirically defined, in our application, by checking if the cluster prototypes are (visually or in the VoI distance sense) pairwise different or as soon as a cluster is just composed by one component only. This condition is fulfilled from 10 classes (for the two different clustering

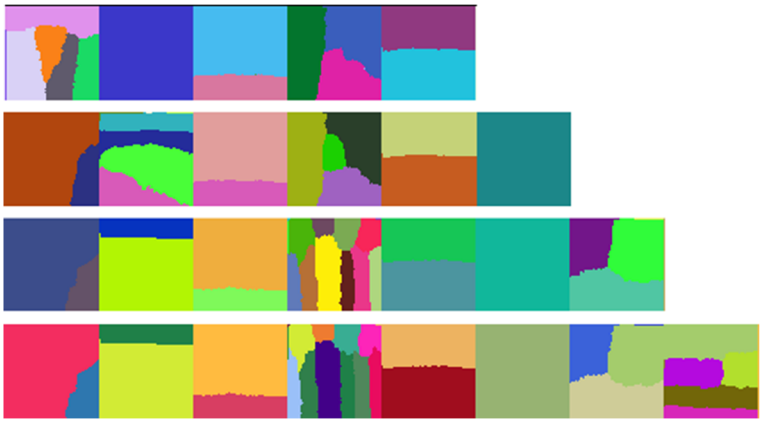


**Fig. 6** Some prototype (segmentation) centers obtained with the VoI-based $K$-means for 5, 6, 7 and 8 clusters

strategies and for the BSDS300) from 10 classes (and this upper bound can be an interesting cue to quantify the diversity of an image-base). Figure 7 shows us all the selected images from the BSDS300 assigned to the 4th cluster of the VoI-based $K$-means procedure (see Fig. 6 for $K = 8$) whose cluster prototype, in term of segmentation map, is recalled on the leftmost image. On this segmentation map, we can see several (four or five) elongated structures (with a sort of head above them) which have been automatically clustered and retrieved from the BSDS300. Indeed, among them, we can see images exhibiting between two or four elongated structures such as persons or statues or ears of corn (or reeds). All the other clustering results can be consulted at the web page of the author's website.[4]

**Algorithm 4**
**VoI-Based HAC**

| | |
|---|---|
| VoI | VoI distance (See Equation (1)) |
| $\{S_k\}_{k \leq N}$ | Set of $N$ segmentations to be clustered |

**Initialization.** Compute the distance matrix $M_{ij}$ describing the dissimilarities between each pair of segmentations with the VoI distance

**while** *all images are in one cluster* **do**

**1.** Find the two most similar (center of) clusters or images (by searching through $M_{ij}$

**2.** Join the two clusters or images to produce a new cluster

**3.** Determine new cluster centers by Algorithm 1

**4.** Update $M_{ij}$ by computing the VoI distance between the center of this new luster and all other clusters or images

### 3.3 Query-by-drawing search

The proposed framework allows us to easily design and perform a query-by-drawing or query-by-sketch search procedure which would allow a user to formulate a query by simply (and coarsely) drawing a desired configuration or layout of the different geometric shapes of the objects he wants to search and to automatically retrieve in the database. Figure 8 shows some examples of a schematic drawing showing one or several geometric shapes and the three nearest neighbor images, in the segmentation-based VoI distance sense, retrieved in the BSDS300. It is interesting to note that the image which is at the center of the BSDS300

[4]Source code (in C++ language) of our algorithm with the set of clustering and mapping results for each clustering strategy are publicly available at the following http address http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/scvoi.html

**Fig. 7** Images from the BSD300 belonging to the cluster corresponding to the cluster prototype, in term of segmentation map, shown on the leftmost image. Images exhibiting elongated structures such as persons or statues or ears of corn (or reeds) have been clustered and retrieved from the BSDS300

(see Fig. 4a) appears in all the test examples. This result is comprehensible, since we recall that, in the VoI distance sense, any existing segmentations are a refinement of this (almost similar) one region segmentation [32]. In addition, this latter segmentation which is very close to the one-region segmentation is also very close of a segmentation map exhibiting one or two simple geometrical segmented shapes as the tested drawing examples. This is also true for the snake and eagle images of the BSDS300.

### 3.4 Hierarchical clustering

Another grouping strategy consists in using the VoI based distance between two segmentation maps (defined in Section 2) along with a Hierarchical Agglomerative Clustering (HAC) based visualization approach. The HAC-based visualization method produces an informative nested hierarchy of similar groups of object or clusters and iteratively builds the hierarchy from the individual elements by progressively merging clusters. It outputs a dendrogram showing all $N$ levels of agglomerations where $N$ is the number of images in the image database, in terms of their region-based descriptive content (and without requiring any parameters such as the number of clusters as the $K$-means procedure). The first agglomeration corresponds to the most similar pair of images in the database and also define the $(N-1)$ clusters existing in the image-base. The last agglomeration allows us to define the two main clusters existing in the image database. Between the first and last iteration, one can also easily search in the dendrogram a data partitioning or segmentation with a specified number of clusters. Algorithm 4 outlines the VoI distance-based HAC algorithm. It is worth mentioning that the HAC algorithm does not make implicit assumptions on cluster shapes, contrary to the $K$-means based clustering Algorithms 2 and 3 which *a priori* assume (sometimes wrongly) that the considered clusters are spherical with equal volumes (or the

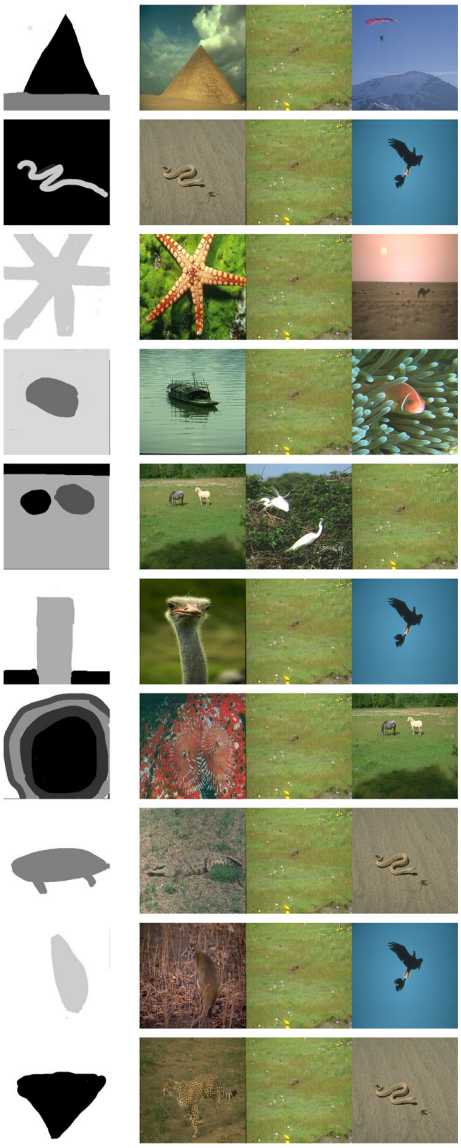


**Fig. 8** Examples of query-by-drawing search (three nearest neighbors) in the BSDS300 in the VoI distance sense

presence of Gaussian distributions with identical covariance matrix) [3]. Computationally speaking, the HAC algorithm requires approximately two hours on the BSDS300 in order to compute the nested hierarchy of clusters, integrating together the different partitioning of the BSDS300 into different numbers of clusters. Figure 9 shows us a dendrogram on the first 50 images of the BSDS300. The full dendrogram estimated for the entire database is available on the author's website. Figure 10 presents us the images related to the first agglomerations, i.e., the set of the most similar images from the BSDS300, in terms of their region-based descriptive content.

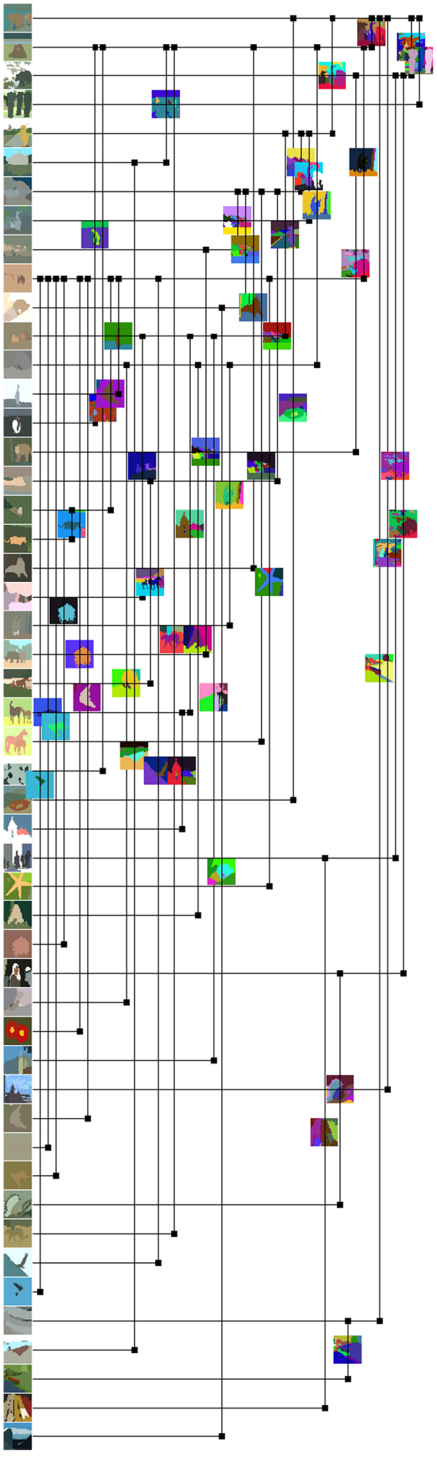


**Fig. 9** A dendrogram or hierarchical agglomerative clustering based on the VoI distance between segmentations on the first 50 images of the BSDS300, showing the 50 agglomerations and the average segmentation between each similar groups of images, in terms of their region-based descriptive content

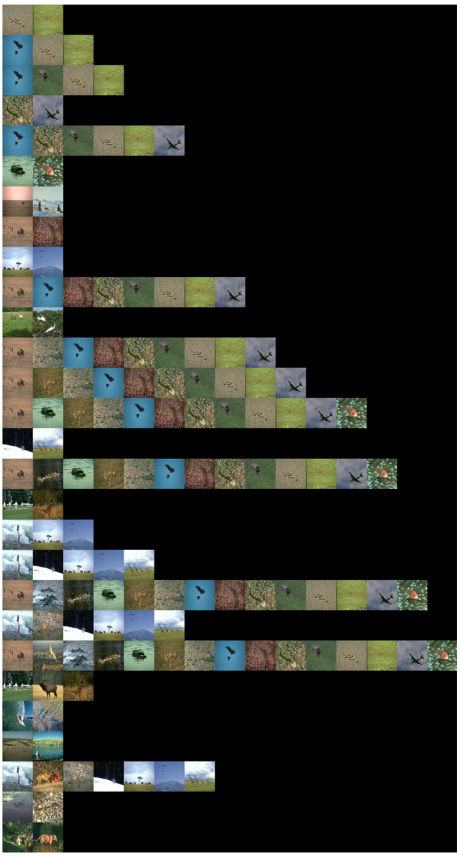


**Fig. 10** Images from the BSDS300 related to the first agglomerations of the HAC, or most similar images, in terms of their region-based descriptive content

### 3.5 Discussion

In order to better understand the behavior and properties of our search algorithm and the differences of our indexing strategy with previous approaches, we have asked Google image[5] to search similar images to a particular image of the BSDS300 in which we can see three stone totems (i.e., three vertical elongated stone sculptures) with a specific texture and color (i.e., the image number 101085) which was one of the image contained in the cluster shown in Fig. 7 (more precisely, this image was assigned to the 4$^{th}$ cluster of our VoI-based $K$-means procedure, see Section 3.2). A Google search online returned several images of statues (see Fig. 11) with the same color and texture (than the query image) but among the top 12 retrieved images, there is no retrieved images with three vertical structures (four of them exhibit two elongated vertical structures) contrary to our approach (see Fig. 7) which have naturally grouped images with a similar layout and spatial arrangement (all the three vertical structures existing in the BSDS300) for the different objects detected in the image regardless of their own color or texture.

[5]At this stage, it is important to recall that Google search image will not seek in a specific database (with 300 images as the BSDS300), but on the whole web image database and therefore it will have more choice to refine its search process which will be more accurate and efficient.

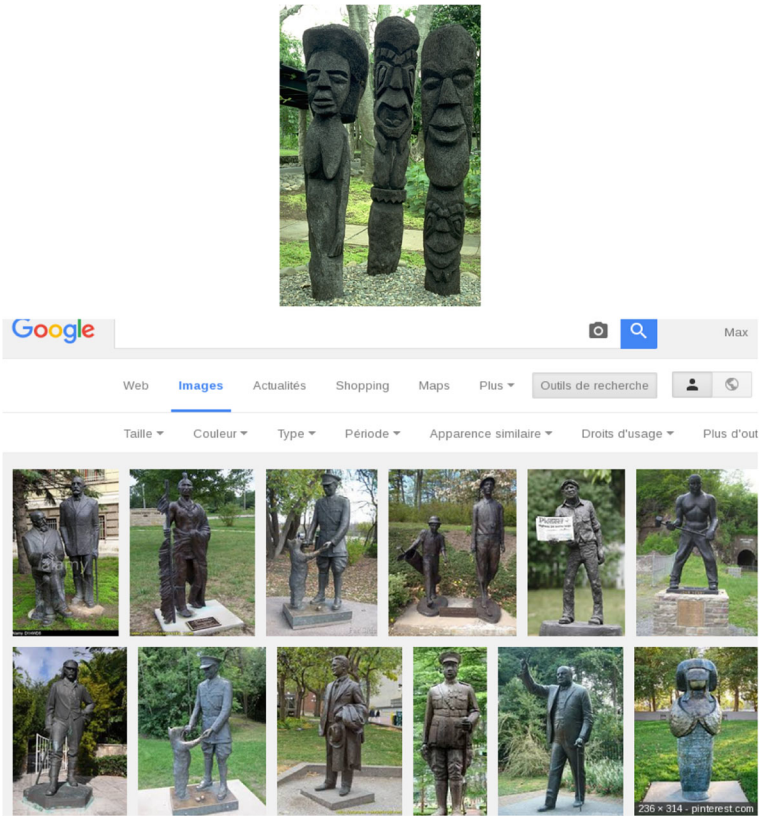



**Fig. 11** Top: Image 101085 from the BSDS300. Bottom: Top 12 most similar images returned by Google image search. A Google search online returned several images of statues with the same color and texture between them (and enough similar to the query image) but none of them show three elongated vertical structures with possibly a different color or texture as our method can give (see Fig. 7)

## 4 Conclusion

In this paper, we have proposed a new MDS based visualization map[5], only based on the geometrical layout and shapes of the different objects detected and segmented in the scene which provides promising image overviews for large image database. Besides, we have also presented various ways and tools for efficiently clustering or for retrieving a specific subset or class of images in terms of their segmentation-based descriptive content[6]. This descriptive segmentation-based information, provided at a higher level of abstraction, can be a significant and complementary information which can be combined with the commonly

[6]This map has also been estimated on 3 other image databases, namely;
1) the Weizmann database (1 & 2 objects) (200 images),
2) the Microsoft Research Cambridge Object Recognition Image Database (MSRC) (591 images),
3) The Stanford Background Dataset (DAGS) (715 images).
The references of these image databases and the obtained visualization maps are publicly available (with the source code of our algorithm) at the following http address:
www.iro.umontreal.ca/~mignotte/ResearchMaterial/scvoi.html

used color or texture cues or SIFT features or text information [1, 8] in order to further help the user to browse through large collections in a more efficient and intuitive manner.

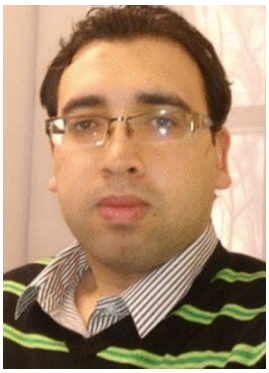
It is worth recalling that the VoI-based distance and average segmentation estimation process can be exploited by all procedures requiring (iteratively or not) the *mean* observations of a sample of (possibly partitioned) data, such as, the mean shift or (more generally) mode seeking based procedures which does not require prior knowledge of the number of clusters. It is also the case of statistical procedures such as PCA (principal component analysis) which could be also generalized in order to study the variability existing in a medical image segmentation database of specific segmented anatomical structures.

In addition, a similar approach could be applied to image data containing a temporal dimension such as video image sequence and the 3D-generalized proposed method could be useful and exploited, in the same way, for video structure analysis (in terms of their segmentation-based descriptive content) or for video indexing problems and retrieval. It could also include query interfaces and video clustering with, for this 3D temporally coherent data, the estimation of a visual (spatio-temporal) prototype (center) model for each cluster which could be subsequently exploited for video classification.

## References


1. Alvarez C, Id-Oumohmed A, Mignotte M, Nie J-Y (2004) Toward cross-language and cross-media image retrieval. In: 5th Workshop on Cross Langage Evaluation Forum, CLEF 2004 Lecture notes in Computer science, Multilingual information access for text, speech and images, vol 3491, Bath, United Kingdom, pp 676–688
2. Borg I, Groenen P (2005) Modern Multidimensional Scaling: Theory and Applications. Springer
3. Banks S (1990) Signal processing image processing and pattern recognition. Prentice Hall
4. Bartolini I, Ciaccia P, Patella M (2006) Adaptively browsing image databases with PIBE. Multimed Tools Appl 31(3):269–286
5. Cayton L, Dasgupta S (2006) Robust euclidean embedding. In: Proceedings of the twentythird International Conference on Machine learning. ACM Press, pp 169–176
6. Ghosh S, Pfeiffer J, Mulligan J (2009) A general framework for reconciling multiple weak segmentations of an image. In: Proceedings of the Workshop on Applications of Computer Vision, (WACV'09), Utah, USA, pp 1–8
7. Heesch D (2008) A survey of browsing models for content based image retrieval. Multimed Tools Appl 40(2):261–284
8. Id-Oumohmed A, Mignotte M, Nie J-Y (2005) Semantic-based cross-media image retrieval. In: 3rd International Conference on Advances in Pattern Recognition, ICAPR'05, Lecture Notes in Computer Science, volume LNCS 3686, Pattern Recognition and Data Mining, Proceedings Part 2, Bath, United Kingdom (UK), pp 414–423
9. Jiang Y, Zhou Z-H (2004) SOM ensemble-based image segmentation. Neural Process Lett 20(3):171–178
10. Keuchel J, Kuttel D (2006) Efficient combination of probabilistic sampling approximations for robust image segmentation. In: 28th Annual Symposium of the German Association for Pattern Recognition. DAGM-Symposium, Lecture Notes in Computer Science, Berlin, Germany, pp 41–50
11. Lloyd SP (1982) Least squares quantization in PCM. IEEE Trans Inform Theory 28(2):129–136
12. Leelanupab T, Feng Y, Stathopoulos V, Jose JM (2009) A simulated user study of image browsing using high-level classification. In: Semantic Multimedia, 4th International Conference on Semantic and Digital Media Technologies, SAMT 2009, Graz, Austria, December 2-4, 2009, Proceedings, pp 3–15
13. Meila M (2005) Comparing clusterings - an axiomatic view. In: Proceedings of the 2005 22nd International Conference on Machine Learning (ICML'05), Bonn, Germany, pp 577–584
14. Meila M (2007) Comparing clusterings–an information based distance. J Multivar Anal 98(5):873–895
15. Mignotte M (2008) Segmentation by fusion of histogram-based K-means clusters in different color spaces. IEEE Trans Image Process 17(5):780–787
16. Mignotte M (2010) A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation. IEEE Trans Image Process 19(6):1610–1624

17. Mignotte M (2014) A label field fusion model with a variation of information estimator for image segmentation. Fusion Information
18. Ma Y, Derksen H, Hong W, Wright J (2007) Segmentation of multivariate mixed data via lossy coding and compression. IEEE Trans Pattern Anal Mach Intell 29(9):1546–1562
19. Mignotte M (2011) A de-texturing and spatially constrained K-means approach for image segmentation. Pattern Recogn lett 32(2):359–367
20. Mignotte M (2011) MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation. IEEE Trans Neural Netw 22(3):447–460
21. Mignotte M (2014) A non-stationary MRF model for image segmentation from a soft boundary map. Pattern Anal Appl 17(1):129–139
22. Mignotte M (2004) Nonparametric multiscale energy-based model and its application in some imagery problems. IEEE Trans Pattern Anal Mach Intell 26(2):184–197
23. Mignotte M (2010) A multiresolution markovian fusion model for the color visualization of hyperspectral images. IEEE Trans Geosci Remote Sens 48(12):4236–4247
24. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th International Conference on Computer Vision (ICCV'01), vol 2, pp 416–423
25. Rachid H, Mignotte M (2009) A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images. In: Proceedings of the 16th IEEE International Conference on Image Processing (ICIP'09), pp 1365–1368
26. Schaefer G (2012) Interactive browsing of image repositories - (invited paper). In: Computer Vision and Graphics - International Conference, ICCVG 2012, Warsaw, Poland, September 24-26, 2012. Proceedings, pp 236–244
27. Schaefer G (2010) A next generation browsing environment for large image repositories. Multimed Tools Appl 47(1):105–120
28. Torgerson W (1952) Multidimensional scaling: I. theory and method. Psychometrika 17:401–419
29. Urban J, Jose J, Van Rijsbergen C (2006) An adaptive technique for content-based image retrieval. Multimed Tools Appl 31(1):1–28
30. Vega-Pons S, Ruiz-Shulcloper J (2011) A survey of clustering ensemble algorithms. Int J Pattern Recogn Artif Intell, IJPRAI 25(3):337–372
31. Wattuya P, Rothaus K, Prani J-S, Jiang X (2008) A random walker based approach to combining multiple segmentations. In: Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08), Florida, USA, pp 1–4
32. Yang AY, Wright J, Sastry S, Ma Y (2008) Unsupervised segmentation of natural images via lossy data compression. Comput Vis Image Underst 110(2):212–225
33. Young G, Householder A (1938) Discussion of a set of points in terms of their mutual distances. Psychometrika 3



**Ayman Khlif** received the engineering diploma in computer science from the Tunisian engineering ENIS-SFAX in June 2012 He is currently a PhD student in computer science at the University of Montreal under the supervision of Pr. Max Mignotte. in the Department of Computer Science and Operational Research (DIRO). His research interests include web browsing, multimedia, image segmentation, fusion and change detection.

**Max Mignotte** received the DEA (Postgraduate degree) in Digital Signal, Image and Speech processing from the INPG University, France (Grenoble), in 1993 and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidentale (UBO) and the digital signal laboratory (GTS) of the French Naval academy, France, in 1998. He was an INRIA post-doctoral fellow at University of Montreal (DIRO), Canada (Quebec), from 1998 to 1999. He is currently with DIRO at the Computer Vision & Geometric Modeling Lab as a Professor at the University of Montreal. He is also a member of LIO (Laboratoire de recherche en imagerie et orthopedie, Centre de recherche du CHUM, Hopital Notre-Dame) and researcher at CHUM. His current research interests include statistical methods, Bayesian inference and hierarchical models for high-dimensional inverse problems.