An experimental study on the automatic segmentation of in situ hybridization-derived images

Pablo Mesejo, Stefano Cagnoni

An experimental study on the automatic segmentation of in situ hybridization-derived images

Pablo Mesejo and Stefano Cagnoni
Intelligent Bio-Inspired Systems laboratory
Department of Information Engineering
University of Parma, Italy
{pmesejo, cagnoni}@ce.unipr.it

***Abstract*—Histology is a well established branch of biology focused on the study of the microscopic anatomy of cells and tissues. Despite its importance, it is surprising to check that most literature devoted to histological image processing and analysis is focused on registration and 3D reconstruction of the whole brain. Therefore, there is a lack of research about automatic segmentation of anatomical brain structures in histological images. To bridge this gap, this paper introduces a comparative study of different segmentation techniques applied to this kind of images.**

A wide and representative image set has been collected to run experiments on the hippocampus, due to the importance of this anatomical district in learning, memory and spatial navigation. Seven approaches, from deterministic to non-deterministic ones, and from recent trends to classical computer vision techniques, have been compared using different standard metrics (Dice Similariry Coefficient, Jaccard Index, Hausdorff Distance, True Positive Rate and False Positive Rate). Proper statistical tests have been performed to draw accurate conclusions about the results. The best performance on this particular problem was obtained by a combination of Active Shape Models (optimized using Differential Evolution) with a refinement step based on Random Forests. This approach achieved an average Dice Similarity Coefficient of 0.89 with a standard deviation of 0.03.

***Index Terms*—Biomedical Image Segmentation, In Situ Hybridization-derived Images, Histological Images, Hippocampus, Deformable Models, Soft Computing, Comparative Study**

I. Introduction

Classically, image segmentation is defined as the partitioning of an image into non overlapping regions that are homogeneous with respect to some visual feature such as intensity or texture [1]. In particular, image segmentation plays a crucial role in many biomedical imaging applications (detection of lesions, measurement of organ size/volume, quantitative tissue analysis, computer-integrated surgery,...), by automating and facilitating the delineation of anatomical structures and different regions of interest (ROIs). As one can easily check taking a look at the literature [2]–[6], most of the segmentation algorithms developed so far have been applied to MRI and CT images, not paying too much attention to other important imaging modalities like ultrasound or microscopy.

Among the different anatomical structures which make up the mammalian brain, the hippocampal formation (HPF) (see Figure 1) is particularly interesting. From an anatomical viewpoint, the HPF, composed by the Hippocampus and the Subiculum (SUB), is located within the medial temporal lobe. In turn, the Hippocampus is composed by the Dentate Gyrus (DG) and Ammon's Horn (CA), which is further composed by three different regions (CA1, CA2, and CA3).

The hippocampus has long been known for its crucial role in learning and memory processes [7]–[10]. Moreover, it has recently been demonstrated that the volume of the hippocampus is an early biomarker for Alzheimer's disease [11]. Therefore, there is a great interest in understanding the cellular and molecular events that take place in this structure, under both normal and abnormal conditions. From this point of view, a precise gene expression map at the cellular and subcellular level within this region can provide crucial information for understanding such biological mechanisms.

A very promising data source to derive this map has recently been provided by the Allen Brain Atlas (ABA), a huge, publicly available database that contains high-resolution images mapping the expression patterns of most genes contained in the genomes of the analyzed organisms. The first release of the ABA was focused on mouse, and contained the expression patterns of about 20,000 genes obtained by In Situ Hybridization (ISH) of mouse brain sections [12]. In a more recent release, a similar resource, based on spatially mapped microarray data, has also been provided for the human brain [13], [14]. The availability of such a large number of images, containing important morphological and functional information, has made it extremely important to develop completely automatic methods that can accurately and robustly extract information from the hippocampus. In the ABA, for each gene under consideration, several images are provided, corresponding to different sections of the brain, and each image is labelled according to the corresponding image of a reference atlas (comprising 132 coronal and 21 sagittal sections spaced at 100 μm and 200 μm intervals, respectively).

ISH is a technology for gene expression analysis that uses a labeled complementary DNA or RNA strand to localize a specific DNA or RNA sequence in a portion or section of tissue (in situ) [15]. As its name suggests, it is a method that localizes

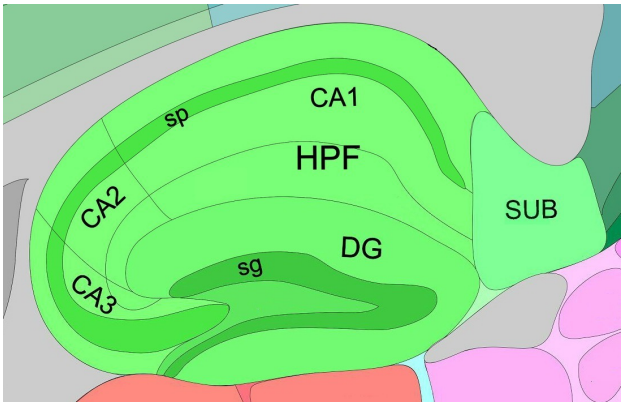

Fig. 1. Regions in Hippocampal Formation (HPF).

and detects specific mRNA sequences in morphologically preserved tissue sections or cell preparations, by hybridizing the complementary strand of a nucleotide probe to the sequence of interest.

Since the task under study in this work is the automatic segmentation of the hippocampus in histological images obtained by ISH, it is important to describe the main characteristics of such images in general, and of the hippocampus region in particular. These features can be summarized as follows:

- natural variability of brain structure shapes in different subjects;
- fuzziness of the hippocampus boundary;
- limited relevance of color for detecting anatomical structures: regions with similar colors may represent different structures and vice versa, depending on the dye used as well as on local image acquisition settings;
- contrast variability between structures: different genes are not expressed equally in the same anatomical region, making it difficult to construct a consistent model for each landmark in all images. Moreover, grained patterns with many irregularities hamper the classification of individual pixels as belonging to the anatomical structures under consideration;
- orientation issues: the different structures may be rotated or displaced on the slice with respect to a "standard" alignment;
- lighting issues: within the same set of images, some are much brighter than others.
- variable resolution even within the same image: high-resolution regions coexist with low-resolution ones;
- presence of artefacts: tears, scraps, bubbles, streaks in tissues, partial cut-off of regions;
- large image size (the typical resolution of complete ABA images is about 15,000 $\times$ 7,000 pixels, and the ROIs which we work with about 2,500 $\times$ 2,000 pixels).

More specifically, and from a low level point of view, these ROIs have the following average histogram (Figure 4), where the darker parts represent, among other things, the hippocampus. These ROIs are automatically extracted from the original images, as explained in [16] (Section 3.1).

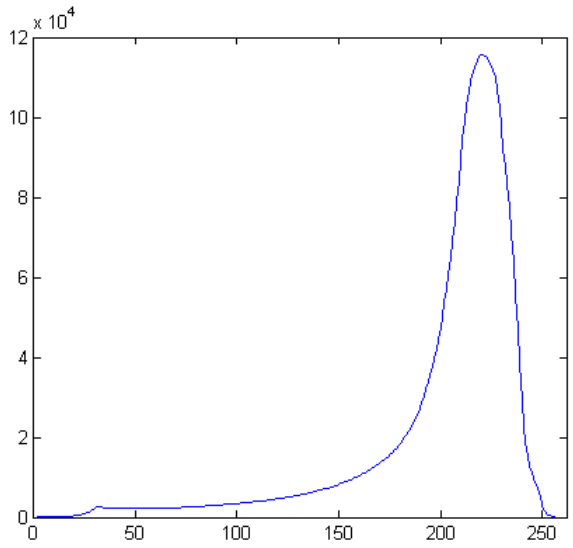

Fig. 3. Average histogram of the images under study.

The main contribution of this paper to the state-of-the-art is threefold: first, the application of the level set method to histological images, which is not common in literature; second, the hybridization of geometric and parametric DMs (two of the methods use a parametric deformable model to initialise a level set); and third, it is one of the few comparative studies devoted to the segmentation of anatomical structures in histological images (where classical pixel- and intensity-based methods, deformable models (DM), and hybridizations with soft computing techniques, are tested and compared)[1].

II. Theoretical Background and Related Work

In this section, we describe the techniques used and their essential working, as well as the main contributions to the field found in literature.

The methods here compared are the following:

- Active Shape Models (and Iterative Otsu Thresholding Method) refined using Random Forests (ASM+RF) [16], [17]. This method, published in 2012, uses a training set of shapes to create a model with its limits for the deformation. The approach is inspired by Active Shape Models (ASM) [18], [19], but takes into account only grey level intensities (not color nor texture), and uses a medial-based shape representation in polar coordinates, with the objective of creating simple models that can be managed in an easy and fast manner (the complexity of the models has been reduced as much as possible, minimizing the number of control points).
  Such a parametric model is moved and deformed by Differential Evolution (DE) [20] according to an intensity-based similarity function between the model and the object itself. After that, Otsu's thresholding method [23]

[1]The methods here compared are all automatic segmentation techniques (manual/semi-automatic methods have not been included).

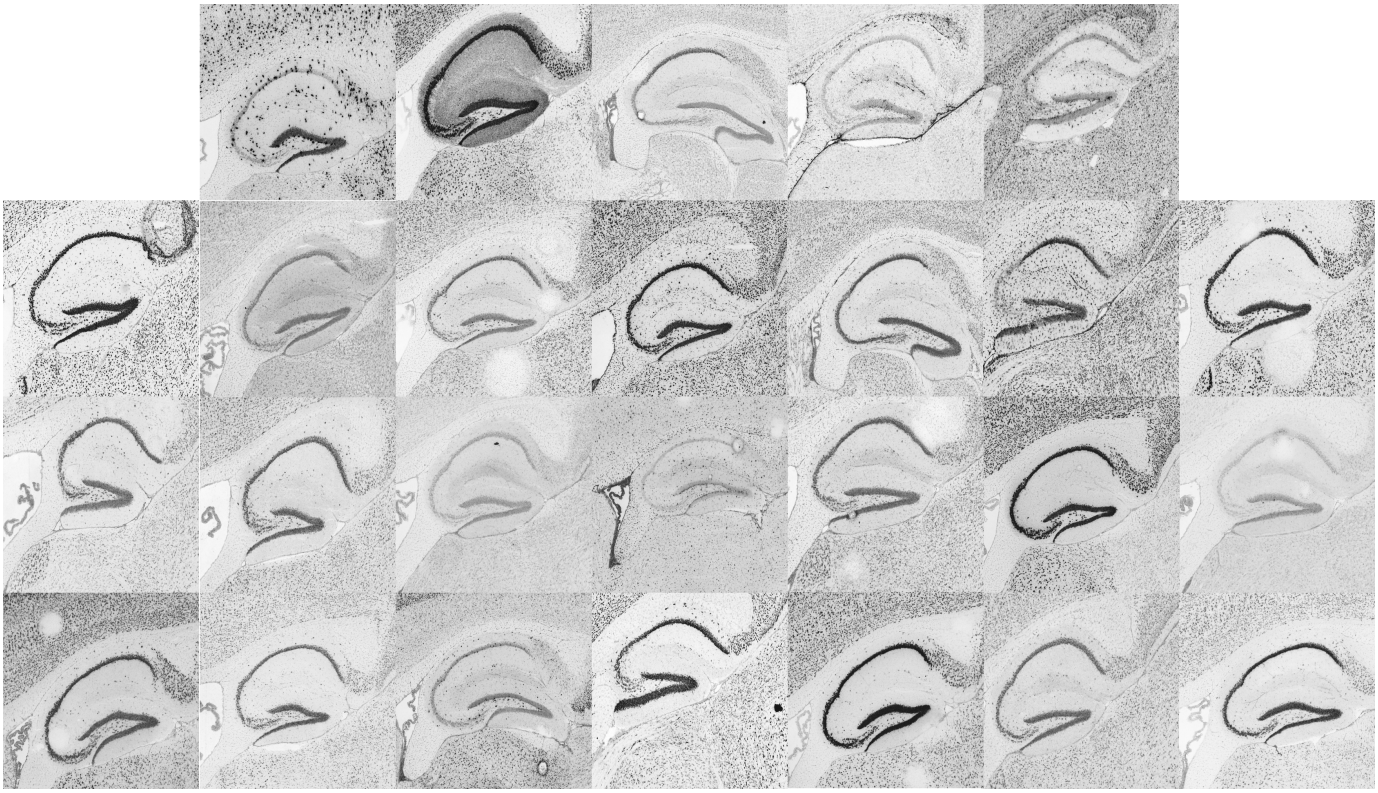

Fig. 2. All images used in the experiments. From up to down, and from left to right: Gad1, Camk2a, 0610010D24Rik, Gfap, Mbp, B230215L15Rik, 3830406C13Rik, A030009H04Rik, Gapdh, 5430437P03Rik, Atp1b2, Tubb3, Azin1, 1300018I05Rik_94, Trem1, Cd9, Wars, Slc17a7, Zim1, Atrx, 1300018I05Rik_86, Cutl2, Camk2g, Camk2b, Wbp7, and Nmt1.

is iteratively applied on every region identified by the located control points and, in every iteration, only the largest segmented component is kept. Therefore, the whole segmentation process can be seen as an intelligent technique for localizing areas where a fast and simple segmentation method is to be applied.

Finally, in order to extend the segmentation towards the parts that have not been segmented yet (mainly because the localization did not cover completely the hippocampus), an Ensemble Classifier (in this case Random Forests [21]), trained using textural features, is applied to detect/classify which points belong to the hippocampus and expand, like a region growing method, the segmented area to the regions that were not properly localized.

- Soft Thresholding (ST) [22]. This method, presented in 2010, is based on relating each pixel in the image to the different regions via a membership function, rather than through hard decisions, and such a membership function is derived from the image histogram. As a consequence, each pixel will belong to different regions with a different level of membership.

  In a first stage, the normalized histogram of the image under study is calculated (using the maximum grey level to bound the histogram to the interval [0, 1]). Then, a sum of weighted known distributions is fit to the histogram, and each probability distribution represents the probability for a pixel with a certain value to belong to the corresponding region.

  This segmentation technique has the following advantages: (1) it is totally automatic, and does not require human intervention, which makes it suitable for automatic processes; (2) the hard decision is postponed to the final stage. So, all the spatial operations are performed before taking into account the different memberships and obtaining the thresholded image; and (3) the spatial operations performed make the thresholding more robust to noise and artifacts.

  Having been succesfully applied to CT, MRI and ultrasound, it seemed interesting to apply it also to histological imaging, to check its performance with this image modality.
- Otsu Thresholding Method (OTSU) [23]. This method, introduced in 1979, is proposed from the viewpoint of discriminant analysis, and it automatically selects an optimal threshold. Otsu's thresholding method involves iterating through all possible threshold values and calculating a measure of spread for the pixel levels on each side of the threshold. The aim is to find the threshold value that minimizes the sum of foreground and background spreads, i.e. it tries to minimize the within-class-variance by maximizing the between-class-variance (which is less computationally intensive to calculate).

The algorithm assumes that the image to be thresholded contains two classes of pixels, or that its histogram is bi-modal (e.g. foreground and background), and then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal. Since ASM+RF uses Otsu in an iterative and local way, it also seemed interesting to test the proper Otsu method directly over the whole ROI, in order to check the improvement obtained by the ASM+RF method when including prior shape knowledge, as well as texture information.

- Geodesic Active Contours (GAC) [24]. This method is a geodesic approach for object segmentation which connects classical 'snakes' based on energy minimization and geometric active contours based on the theory of curve evolution. The technique, introduced in 1997, is based on active contours that evolve in time according to intrinsic geometric measures of the image: the evolving contours naturally split and merge, allowing the simultaneous detection of several objects and both interior and exterior boundaries.
  The Partial Differential Equation of the GAC is the following:

$$u_t = \alpha \cdot div(g\nabla u/|\nabla u|)|\nabla u| + \beta \cdot g|\nabla u| \qquad (1)$$

  where g is a positive and strictly decreasing function, ∇ is the gradient operator computed on the image I, div is the divergence operator (that measures the magnitude of a vector field's source or sink at a given point) , and α and β are the contour (internal force) and expansion (external force) weights, respectively. The higher α, the more regularized/smoothed the contour. If β is positive it expands outwards, if negative it expands inwards (moving faster and crossing gradients, the smaller its value).
  The problem is formalized as the minimization of the energy:

$$E(C) = \int_0^{|C|_\varepsilon} g(|\nabla I(C(s))|)ds, \qquad (2)$$

  where $|C|_\varepsilon$ is the Euclidean length of a contour C, s is the arc length of the contour, and ds is equivalent to $|C'(q)|dq$, being $C(q)$ a parameterized planar curve. This method is equivalent to the minimization of the length of the curve C according to a Riemannian metric, and such a metric depends here on the local gradient of the image I.
  "Geodesic active contours" minimize Equation 2 via a gradient descent scheme and a level-sets representation of the curve. Unfortunately, the method is sensitive to initialization and the global minimum of Equation 2 is not always found.
  In this work, two implementations of GAC have been tested. The first one uses as initial contour the whole image, while the second one, called GAC+ASM, employs the localization method of ASM+RF (i.e. an ASM using DE) to create the initial contour of the geometric DM. Except the initial contour, all the remaining parameters were exactly the same in both approaches.
- Chan&Vese Level Set Model (C&V) [25]. While many segmentation methods heavily rely in some way on edge detection, the "Active Contours Without Edges" method by Chan and Vese (2001) ignores them completely. They have proposed an active contour model that can detect objects whose boundaries are not necessarily defined by gray level gradients. Instead, the segmentation boundary is represented implicitly with a level set function, which allows the segmentation to handle topological changes more easily than explicit snake approaches, and the method is based on trying to separate the image into regions based on intensities.
  The Chan&Vese model is a region-based model for detecting objects in an image, based on a restriction Mumford-Shah model which divides an image into regions and represents each region by a piecewise constant (the minimal partition problem).
  More formally, the Euler-Lagrange equation (as derived in the original paper) representing the time motion of the curve C is shown in Equation 3.

$$\frac{\partial \phi}{\partial t} = \delta_\varepsilon(\phi)[\mu \ \mathrm{div}(\frac{\bigtriangledown \phi}{|\bigtriangledown \phi|}) - \nu - \lambda_1(u_0 - c_1)^2 + \lambda_2(u_0 - c_2)^2] \qquad (3)$$

  The parameter μ is a scaling parameter for the length of the curve represented in terms of curvature. The smaller μ, the more the length of the curve can increase without penalizing the minimization. This allows the model to detect small objects and holes. The larger μ, the less freedom is there for the curve to increase in length, thus, it will only be able to detect larger objects. The parameter ν is also a scaling term for the area of the curve, but it can be set to 0 in the Euler-Lagrange derivation, since μ is sufficient to scale the curve according to the objects that need to be detected. Finally, λ_1 and λ_2 are weighting parameters for the forces inside the curve and outside the curve respectively. Since we want to give both forces equal weight, we have set $\lambda_1 = \lambda_2 = 1$ in all experiments. Therefore, the evolution of the curve is influenced by two terms: the "curvature" that regularizes the curve and makes it smooth during evolution, and the "region term" that affects its motion.
  The main theoretical advantages of this approach are that it automatically detects interior contours and both type of boundaries (with and without gradient), and the initial curve can be placed anywhere in the image.
  Also in this case, like in GAC, two implementations have been tested. The first one uses the whole image as initial contour, and the second one employs the localization results obtained by ASM to create the initial contour.

Regarding the segmentation of structures in histological images, our work is related to the investigations of Senyukova et al. [26]. There, the segmentation is treated as a classification

problem using RF and Markov Random Fields, which refine the results at the pixel level, but the method requires previous knowledge about the reference slice associated to that image. In [27], the segmentation using three-dimensional Gaussian mixtures and Level Sets is carried out slice-by-slice, where the successful segmentation of one section provides a prior for the subsequent one, assuming that the segmentation of few sparsely sampled slices is done manually (so it is not a completely automatic method).

On the other hand, when CV, GAC or ASM are hybridized with DE, this implies a combination between DM and evolutionary computation techniques. There are also some examples in this regard, like [28], where "Genetic snakes" are active contour models with an optimization procedure based on GA, or [29], where a GA evolves a population of medial-based shapes, using prior shape knowledge to produce feasible deformations while also controlling the scale and localization of these deformations. Also in [30] a GA is used to perform level set curve evolution using texture and shape information to automatically segment the prostate in CT and MRI pelvic images. Other example can be found in [31], where DE, in combination with a greedy search algorithm, is used to evolve Topological Active Nets for CT image segmentation.

III. Experimental Results

It is important to notice that all methods used in the experimentation, apart from ASM+RF, and GAC and CV when using ASM to obtain the initial contour, are deterministic techniques, i.e. given a particular input, they will always produce the same output. This is not the case of the other three approaches, since stochastic methods, like DE, are embedded in these algorithms, and can exhibit different behaviors on different runs. Due to this, it is essential to run such algorithms several times to estimate and compare their performances.

In order to evaluate the accuracy of all methods, tests on real images were run on an Intel® Core™ i5-2410M CPU @ 2.30GHz with 4.00 GB of RAM, using MATLAB as programming language. Table II reports the results obtained using 26 real images, employing 25 runs per image (for non-deterministic methods), for a total of 2054 experiments. The results are organized per blocks of five rows each. Every block corresponds to a different metric [32]: Dice Similarity Coefficient, Hausdorff Distance, Jaccard Index, True Positive Rate, and False Positive Rate. Inside every block, the average, median, standard deviation, best value, and worst values are shown. These calculations are computed over 26 or 650 values depending on the nature of the method under consideration (deterministic or non-deterministic, respectively).

A ground truth was created by manually segmenting the hippocampus in these 26 images. In order to avoid erroneous or incomplete manual segmentations, these were supervised by an expert in molecular biology. Every image was manually segmented 5 times and the intersection and union images were calculated. Nevertheless, the difference between the maximum (union) and minimum (intersection) area segmented in these (equally valid) segmentations showed an average difference of 28.44% with a standard deviation of 5.61%. Simultaneous truth and performance level estimation (STAPLE) [33] was used to create a consensus ground truth, since it is considered one of the state-of-the-art methods to create a ground truth from several manual segmentations [34].

A post-processing phase was applied to remove isolated pixels: a median filtering (15x15), only in the case of ST and Otsu, and the removal of connected components with less than 5000 pixels in all methods, except ASM+RF. The manually tuned parameters used on each method are shown in Table I. For GAC, C&V, GAC+ASM and C&V+ASM, a median filtering (10x10) was applied during pre-processing in order to regularize the surface of the image.

The test set is composed of 26 representative genes downloaded from the ABA database (see Figure 2). In particular, the images downloaded, and used for the comparative study, were the ones associated with the following gene symbols: Wbp7, 1300018I05Rik_86, 0610010D24Rik, 1300018I05Rik_94, 5430437P03Rik, Atrx, B230215L15Rik, Nmt1, Wars, Tubb3, Trem1, 3830406C13Rik, A030009H04Rik, Atp1b2, Azin1, Camk2a, Camk2b, Camk2g, Cd9, Cutl2, Gad1, Gapdh, Gfap, Mbp, Slc17a7, and Zim1.

IV. Discussion and Conclusions

In order to check the statistical significance of the results obtained, a Kruskal-Wallis test was performed with a level of confidence of 0.01. Since the normality and homoscedasticity assumptions were not accomplished, as checked through the application of Kolmogorov-Smirnov and Bartlett's tests, non-parametric tests were used.[2] The p-value is near zero, suggesting that at least one sample median is significantly different from the others.

Some conclusions can be extracted from the numerical data in Table II and from some of the representative results included in Figure 5. The first conclusion could be the difficulty of tackling satisfactorily these images, since established and well-known segmentation methods, like C&V and GAC, did not obtain as good results as one could expect. In particular, GAC presents problems with images that have small variations of their gradient values, i.e. when the difference, in terms of intensity, between the hippocampus and the background is low.

According to Table II, the best results were obtained by ASM+RF, being the best in all metrics but in TP rate, where C&V, C&V+ASM and Otsu achieved better results. This is explained by the tendency of these three methods to oversegment the images and, therefore, to segment the hippocampus and many other regions in the image. In general terms, the best methods to segment these images have been ASM+RF, C&V+ASM, GAC+ASM, and ST.

From the values of HD interesting conclusions can also be drawn. This metric represents the largest of all the distances

[2]Note that the Kruskal-Wallis test is a nonparametric version of the classical one-way ANOVA, and an extension of the Wilcoxon rank sum test to more than two groups. It returns the p-value for the null hypothesis that all samples are drawn from the same population.

TABLE I

PARAMETERS USED IN TESTING DIFFERENT SEGMENTATION TECHNIQUES.

ST	OTSU	ASM+RF	C&V	GAC
L = 2 regions Relative max normalization	No parameters to tune	$Cr = 0.9$ $F = 0.7$ Uniform Crossover DE/target-to-best/1 Population Size = 80 Iterations = 250 Median Filter [25×25] RF with 500 trees	number of iterations = 500 μ = 0.1 (length term) ν = 0 (area term) $\lambda_1 = \lambda_2$	number of iterations = 500 β = -1 (expansion weight) α = 3 (contour weight)

TABLE II

SEGMENTATION RESULTS USING 5 DIFFERENT METRICS: DICE SIMILARITY COEFFICIENT (DSC), JACCARD INDEX (JI), HAUSDORFF DISTANCE (HD), TRUE POSITIVES (TP), AND FALSE POSITIVES (FP). THE BEST RESULTS FOR EVERY METRIC ARE SHOWN IN BOLD LETTERS.

		Methods						
		OTSU	ST	C&V	C&V+ASM	GAC	GAC+ASM	ASM+RF
DSC	Average	0.5755	0.6798	0.6164	0.6488	0.6373	0.7469	**0.8882**
	Median	0.5788	0.7505	0.6906	0.7577	0.6985	0.7971	**0.8975**
	Std	0.1820	0.2276	0.2432	0.2052	0.2723	0.1993	**0.0347**
	Best	0.8341	0.9084	0.9120	0.8909	0.9189	0.9198	**0.9350**
	Worst	0.2868	0.0687	0.2183	0.2358	0.0000	0.0214	**0.7422**
JI	Average	0.4263	0.5531	0.4869	0.5121	0.5205	0.6283	**0.8006**
	Median	0.4090	0.6007	0.5274	0.6099	0.5367	0.6626	**0.8141**
	Std	0.1823	0.2330	0.2465	0.2117	0.2782	0.2080	**0.0541**
	Best	0.7154	0.8321	0.8382	0.8033	0.8500	0.8514	**0.8779**
	Worst	0.1674	0.0356	0.1225	0.1336	0.0000	0.0108	**0.5901**
HD	Average	936.2499	660.4281	566.7963	397.8478	543.2649	247.6652	**110.8424**
	Median	988.7435	723.4520	602.9961	378.8250	565.7718	212.2499	**93.9415**
	Std	220.5734	354.8354	340.3507	253.4867	392.5249	174.1109	**47.6210**
	Best	506.7790	57.0088	33.5410	40	67.0820	**30**	44.7214
	Worst	1.3837e+003	1.3911e+003	1.1942e+003	1.0330e+003	1.3067e+003	721.9591	**296.0152**
TP	Average	89.4691	80.0974	95.6464	**97.2764**	85.1330	82.4358	87.3440
	Median	95.4543	90.2218	98.5369	**99.2808**	97.4342	95.2620	87.7858
	Std	15.5791	23.5970	9.0898	**6.8649**	25.9801	25.5602	4.6220
	Best	99.9584	99.9659	**100**	**100**	99.9364	99.9364	94.9942
	Worst	35.8155	5.7835	**64.1182**	**64.1182**	0.0000	1.3607	60.5404
FP	Average	54.7366	34.2408	50.3108	47.9808	41.3426	28.1579	**9.2846**
	Median	58.7041	26.9258	46.5393	38.9488	27.4716	22.1582	**7.8321**
	Std	19.3433	26.0209	25.0742	21.5470	30.7668	19.4360	**5.5229**
	Best	25.6864	**1.5257**	11.6259	16.0377	7.8152	3.6460	3.0877
	Worst	81.2674	91.5339	87.7290	86.5843	100	95.0183	**32.4049**

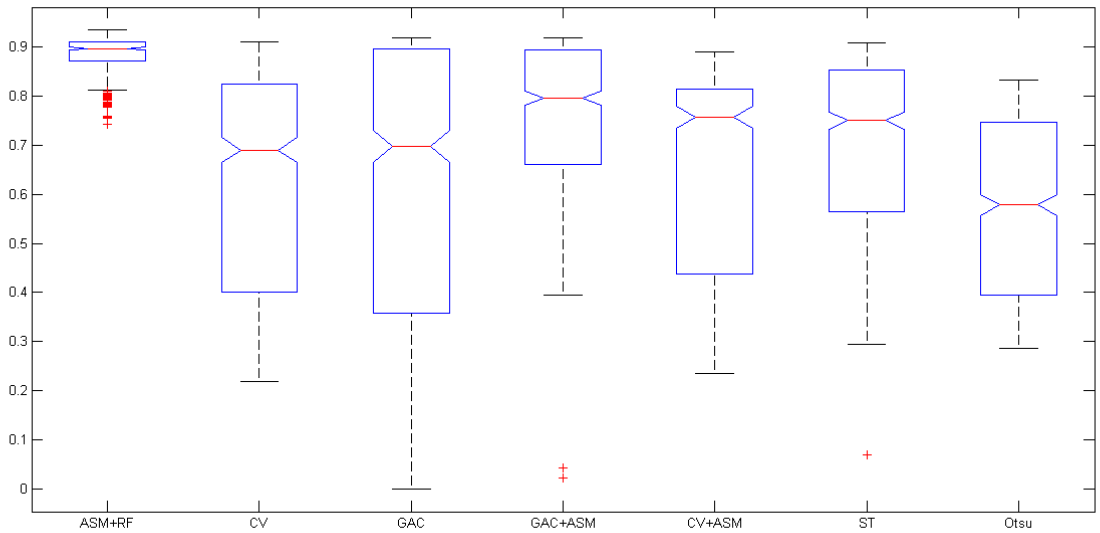

Fig. 4. Box-plot representing the DSC for all methods. The best results are obtained by ASM+RF.

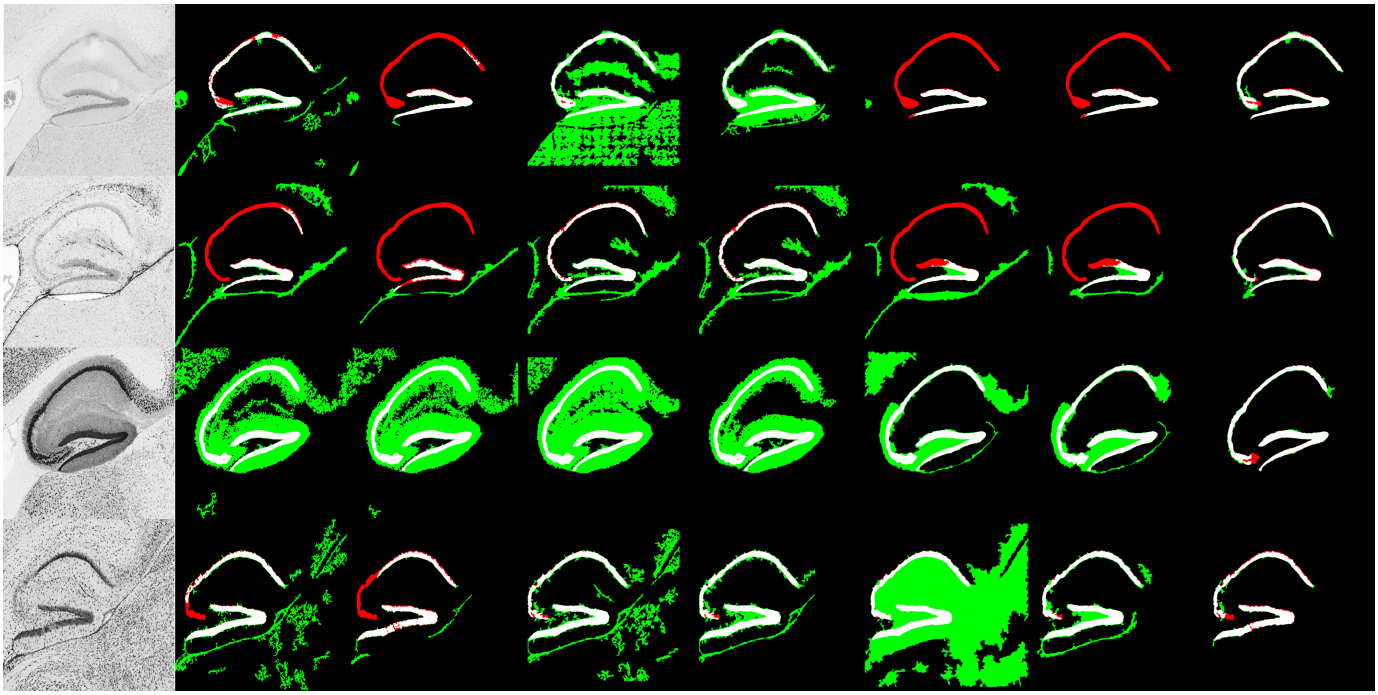

Fig. 5. Results for some images (from up to down: Zim1, Gfap, Camk2a, and Atp1b2). The order of methods is (from left to right): OTSU, ST, C&V, C&V+ASM, GAC, GAC+ASM and ASM+RF. True Positives are represented in white, False Positives in green and False Negatives in red.

from a point in one set (the ground truth) to the closest point in the other set (the result of the automatic segmentation). It is interesting to check how ST has a higher (worst) HD with respect to C&V, even if it obtained better DSC results than the latter. This could be justified by the absence in the former of any kind of shape restrictions or regularization terms, that makes that the pixels are segmented only taken into account their grey level values.

On the other hand, Otsu and ST are the fastest methods since they work only, and directly, on the histogram. The difference in performance between Otsu and ASM+RF demonstrate the great improvement obtained with the introduction of prior knowledge.

Geometric DM have shown a good ability of changing topology and evolving naturally the contour, but on the other hand, they are very sensitive to initialization, pre-processing stages[3], and parameter selection. GAC and C&V suffer with bad initializations (the use of a better initial contour favors a fast convergence to the contour and better results), but this affects specially to GAC (as can be cheked in the improvement obtained when using the localization method based on ASM). If the initial contour is located partially inside and outside of the structure of interest (as sometimes happens due to a bad localization), only the boundaries will be segmented and not the internal part of the hippocampus, because the high gradient values are located on the boundary and not in the internal part of the structure.

Regarding both geometric approaches (C&V and GAC, with and without localization stage), the results would improve in great manner if statistical shape priors for level sets were used (thickness, length, statistical information about the image to segment,...). The only piece of prior knowledge that GAC and C&V are using now is the existence of high gradient values in the boundary of the objects to segment, and the prior knowledge included in the parameter selection (weighting the different components of the energy). Another aspect to take into account is that, sometimes, the inner part of sg is completely segmented without separating the two extremes of this region. This owes to the fact that the regularization factor does not allow the evolving contour to "pass" through small narrow regions and, therefore, it fills completely the empty space of sg. Furthermore, the problem with GAC and C&V in noisy images is that, if the images are smoothed too much, it will be impossible for the method to stop at the gradients, but if they are not smoothed, the noise would provoke the segmentation of many noisy parts in the image.

As previously said, ASM+RF obtained the best results, but it is fair to underline its ad-hoc nature. It needs a training set of shapes to create the template and its possible deformations, and it also needs a training set of textural patterns for the expansion phase. Also it is not able to manage topological changes in a natural way, as geometric DMs can do. So, it is a very accurate method but it sacrifices its general applicability to segment other shapes.

Possible future works include the use of a local C&V

[3]A brief example to show the influence of the type of pre-processing applied. If the size of the median filter were 20x20 instead of 10x10 the results would get worse for CV (DSC Mean: 0.586159 — DSC Median: 0.639210 — DSC Std: 0.256346) and they would improve for GAC (DSC Mean: 0.668008 — DSC Median: 0.841995 — DSC Std: 0.344408)

instead of a global one [35]; introducing segmentation through registration algorithms in the comparison; and trying to create a new segmentation algorithm combining the good features of different methods (like the combination of the ability of geometric DM to manage, naturally, topological changes, with the global search capabilities of Evolutionary Algorithms, or initializing them using existing segmentation methods).

ACKNOWLEDGMENT

Pablo Mesejo is funded by the European Commission (MIBISOC Marie Curie Initial Training Network, FP7 PEOPLE-ITN-2008, GA n. 238819). All the images were downloaded from the Allen Mouse Brain Atlas [http://mouse.brain-map.org]. Seattle (WA): Allen Institute for Brain Science. ©2009.

The authors want to thank Ferdinando Di Cunto, from the University of Torino, for facilitating the ground truth for these images.

REFERENCES

[1] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315–337, 2000.

[2] U. Maulik, "Medical image segmentation using genetic algorithms," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, pp. 166–173, 2009.

[3] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3d medical image segmentation: a review," *Medical Image Analysis*, vol. 13, pp. 543–563, 2009.

[4] C. Petitjean and J.-N. N. Dacher, "A review of segmentation methods in short axis cardiac MR images." *Medical image analysis*, vol. 15, no. 2, pp. 169–184, 2011.

[5] L. He, Z. Peng, B. Everding, X. Wang, C. Y. Han, K. L. Weiss, and W. G. Wee, "A comparative study of deformable contour methods on medical image segmentation," *Image and Vision Computing*, vol. 26, no. 2, pp. 141–163, 2008.

[6] T. Mcinerney and D. Terzopoulos, "Topology adaptive deformable surfaces for medical image volume segmentation," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 840–850, 1999.

[7] K. A. Norman, "How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model," *Hippocampus*, vol. 20, pp. 1217–1227, 2010.

[8] H. Eichenbaum, "A cortical-hippocampal system for declarative memory," *Nature Reviews Neuroscience*, vol. 1, pp. 41–50, 2000.

[9] R. C. O'Reilly, "Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function," *Psychological Review*, vol. 108, pp. 311–345, 2001.

[10] C. L. Thompson, S. D. Pathak, and others, "Genomic Anatomy of the Hippocampus," *Neuron*, vol. 60, pp. 1010–1021, 2008.

[11] J. Barnes, J. W. Bartlett, L. A. van de Pol, C. T. Loy, R. I. Scahill, C. Frost, P. Thompson, and N. C. Fox, "A meta-analysis of hippocampal atrophy rates in Alzheimer's disease," *Neurobiology of Aging*, vol. 30, pp. 1711–1723, 2009.

[12] Allen Institute for Brain Science, "Allen Reference Atlases," http://mouse.brain-map.org, 2004-2006.

[13] A. R. Jones, C. C. Overly, and S. M. Sunkin, "The Allen Brain Atlas: 5 years and beyond," *Nature Reviews Neurosciences*, vol. 10, no. 11, pp. 821–828, 2009.

[14] M. J. Hawrylycz *et al.*, "An anatomically comprehensive atlas of the adult human brain transcriptome," *Nature*, vol. 489, no. 7416, pp. 391–399, 2012.

[15] L. Jin and R. V. Lloyd, "In situ hybridization: Methods and applications," *Journal of Clinical Laboratory Analysis*, vol. 11, pp. 2–9, 1997.

[16] P. Mesejo, R. Ugolotti, F. Di Cunto, M. Giacobini, and S. Cagnoni, "Automatic Hippocampus Localization in Histological Images using Differential Evolution-Based Deformable Models," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 299 – 307, 2013.

[17] P. Mesejo, R. Ugolotti, S. Cagnoni, F. Di Cunto, and M. Giacobini, "Automatic Segmentation of Hippocampus in Histological Images of Mouse Brains using Deformable Models and Random Forest," in *Proc. on 25th International Symposium on Computer-Based Medical Systems (CBMS'12)*, 2012, pp. 1–4.

[18] T. F. Cootes, A. Hill, C. J.Taylor, J.Haslam, and M. M. Pt, "The Use of Active Shape Models For Locating Structures in Medical Images," 1994.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Underst.*, vol. 61, pp. 38–59, 1995.

[20] S. Das and P. Suganthan, "Differential Evolution: A Survey of the State-of-the-Art," *IEEE Transactions on Evolutionary Computation*, vol. 15, pp. 4–31, 2011.

[21] L. Breiman, "Random forests," *Maching Learning*, vol. 45, pp. 5–32, 2001.

[22] S. Aja-Fernandez, G. Vegas-Sanchez-Ferrero, and M. Martin Fernandez, "Soft thresholding for medical image segmentation," in *Proc. on Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'10)*, 2010, pp. 4752 –4755.

[23] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62 –66, 1979.

[24] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, pp. 61–79, 1995.

[25] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 266 –277, 2001.

[26] O. V. Senyukova, A. S. Lukin, and D. P. Vetrov, "Automated atlas-based segmentation of Nissl-stained mouse brain sections using supervised learning," *Programming and Computing Software*, vol. 37, pp. 245–251, 2011.

[27] T. Riklin-Raviv, N. Sochen, N. Kiryati, N. Ben-Zadok, S. Gefen, L. Bertand, and J. Nissanov, "Propagating distributions for segmentation of brain atlas," in *Proc. IEEE International Symposium on Biomedical Imaging, ISBI '07*, 2007, pp. 1304–1307.

[28] L. Ballerini, "Genetic snakes: active contour models by genetic algorithms," in *Genetic and Evolutionary Computation in Image Processing and Computer Vision*, ser. EURASIP Book Series on SP & C, 2007, pp. 177–194.

[29] C. McIntosh and G. Hamarneh, "Medial-based deformable models in non-convex shape-spaces for medical image segmentation using genetic algorithms," *IEEE Trans. on Medical Imaging*, vol. 31, no. 1, pp. 33–50, 2012.

[30] P. Ghosh, M. Mitchell, J. A. Tanyi, and A. Hung, "A genetic algorithm-based level set curve evolution for prostate segmentation on pelvic ct and mri images," in *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*. IGI Global, 2010, pp. 127–149.

[31] J. Novo, J. Santos, and M. G. Penedo, "Topological active models optimization with differential evolution," *Expert Systems with Applications*, vol. 39, no. 15, pp. 12 165–12 176, 2012.

[32] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. F. Cootes, M. Jenkinson, and D. Rueckert, "Comparison and evaluation of segmentation techniques for subcortical structures in brain mri," in *Proc. of MICCAI'08*, 2008, pp. 409–416.

[33] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, pp. 903–921, 2004.

[34] A. M. Biancardi, A. C. Jirapatnakul, and A. P. Reeves, "A comparison of ground truth estimation methods," *Int. J. Computer Assisted Radiology and Surgery*, vol. 5, no. 3, pp. 295–305, 2010.

[35] X.-F. Wang, D.-S. Huang, and H. Xu, "An efficient local chan-vese model for image segmentation," *Pattern Recogn.*, vol. 43, no. 3, pp. 603–618, Mar. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2009.08.002