Exploration and Visualization of Segmentation Uncertainty Using Shape and Appearance Prior Information

Ahmed Saad, Ghassan Hamarneh, Senior Member, IEEE, and Torsten Möller, Member, IEEE

Abstract—We develop an interactive analysis and visualization tool for probabilistic segmentation in medical imaging. The originality of our approach is that the data exploration is guided by shape and appearance knowledge learned from expert-segmented images of a training population. We introduce a set of multidimensional transfer function widgets to analyze the multivariate probabilistic field data. These widgets furnish the user with contextual information about conformance or deviation from the population statistics. We demonstrate the user's ability to identify suspicious regions (e.g. tumors) and to correct the misclassification results. We evaluate our system and demonstrate its usefulness in the context of static anatomical and time-varying functional imaging datasets.

Index Terms—Uncertainty visualization, Medical imaging, Probabilistic segmentation.

1 Introduction

Medical image segmentation is the procedure of delineating an image into its constituent structures. Each structure is assigned a semantic label that characterizes regions with similar anatomical information (e.g. for 3D magnetic resonance imaging (MRI) or computed tomography (CT) images) or similar functional information (e.g. for dynamic positron emission tomography (dPET) or dynamic single photon emission computed tomography (dSPECT) images). Image segmentation is often the precursor to patient diagnosis and treatment evaluation, which is facilitated through a thorough statistical analysis and quantification aided by a variety of visual tools.

In direct volume rendering (DVR), the transfer function plays the role of the image segmenter by assigning optical properties (color and opacity) to different regions in the image. Several methods have been proposed for transfer function design [13]. The assumption behind the design of most transfer functions is that the segmentation is done solely using local features of the dataset currently examined. Those features might be corrupted due to different image degradation factors, such as low signal-to-noise ratio (SNR), partial volume effect (PVE), etc. Thus feature extraction becomes an error-prone process. Further, the extracted features usually depend on local information and do not leverage the fact that, in medical imaging, the structures of interest usually have known but variable global shape and appearance properties.

In the medical imaging analysis domain, it has been demonstrated that medical image segmentation techniques can be greatly improved and made more robust by incorporating prior shape and appearance knowledge. The prior information can typically be learned from a training set of images and their corresponding expert segmentations. The major obstacle of prior-based segmentation is the insufficient availability of expert segmentation samples, which hinders the ability to accurately capture the large variability in the shape and appearance of anatomical structures or of functional activity. Further, automated image analysis methods typically rely on the optimization of a complex energy functional, which encodes multiple competing image or prior-driven cost terms, in an attempt to mimic the cognitive capabilities and years of experience of expert users (e.g. radiologists). Consequently, these methods suffer from sensitivity to parameter tuning, entrapment into erroneous local minima solutions, and a heavy demand for computational resources. Furthermore, the prior models employed are typically built from healthy subjects, which renders the automatic detection of suspicious pathological regions (with unexpected shape, appearance, or function) quite challenging. It is, therefore, undeniable that a fully automatic segmentation algorithm, that is not only fast and accurate but also robust to the vast variability and pathology, is still beyond our reach. We advocate that involving the user in exploring and refining the segmentation is unavoidable. We further argue that the user should not only be focusing on the case under study, but rather be granted with supplementary information that places the studied case in the context of the population, i.e. how it conforms or deviates from the data and analyses resulting from a corpus of training data.

The primary technical contribution of this work is the design of a set of multidimensional transfer function widgets, which utilize the information obtained from a probabilistic segmentation technique, as well as the probabilistic prior information learned from previous expert segmentations. Going beyond the use of simple likelihood information ([27], [30] and [46]), we adopt a Bayesian framework to incorporate more informative priors (than the simple uniform prior) for both shape and appearance as learned from expert-segmented images of a training population. We further explore the relationship between the likelihood and prior terms to allow for the identification of regions of disagreement with the prior. Those regions are potentially the result of misclassification, in which case the user can edit the result and provide a satisfactory, correct result. Alternatively, those regions might be suspicious or pathological (e.g. tumors), which must then be highlighted in order to draw the user's attention to them and react appropriately. Our tool, geared toward research clinicians, will allow the user to explore the uncertainty in the resulting segmentation with respect to the population. In addition, medical image analysis researchers can use this tool to analyze the behavior of their probabilistic segmentation algorithms, in an effort to improve their performance (e.g. by altering the underlying energy functional, improving the optimization mechanism, etc.).

2 Related work

2.1 Segmentation techniques

Medical image segmentation techniques can be classified into automatic and interactive methods. The so called automatic methods still require parameter setting and initialization, but their core operation is automated and the user is not capable of intervening during the segmentation. When the results are unsatisfactory, the user resets the parameters and/or updates the initialization, then re-runs the automatic algorithm. We will limit our discussion of related work to the class

- *Ahmed Saad and Ghassan Hamarneh are with the Medical Image Analysis Laboratory (MIAL), School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6.*
- *Ahmed Saad and Torsten Möller are with the Graphics, Usability, and Visualization (GrUVi) Laboratory, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6.*

E-mail: {aasaad,hamarneh,torsten}@cs.sfu.ca

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

of interactive methods. Three main forms of user input have been adopted for interactive segmentation [37], which entail (1) specifying a nearby complete boundary that evolves to the desired boundary (e.g. the classical Snakes algorithm [24] and the plethora of variants [33]); (2) specifying a small set of voxels belonging to the desired boundary, followed by the application of minimal path techniques to extract the complete boundary (e.g. Intelligent Scissors [36, 2] and 3D Live-wire [39]); or (3) specifying a small set of voxels belonging to the core of different regions to overcome the limitation of the ill-defined boundaries (e.g. using Graph Cuts [4] or Random Walker [17]), followed by some optimal labelling of all unlabelled voxels.

The framework proposed in this paper is independent of the particular segmentation technique as long as it provides probabilistic segmentation results, similar to those that employ Gaussian mixture models [51], the Random Walker method [17], the probabilistic support vector machine approach [38], or a variant of graph cuts [28].

The main building block of segmentation algorithms is the ability to assess data values at every voxel (e.g. intensity, gradient, texture, or other features) in an attempt to label the voxel with a particular class label. Typically, adopting a voxel-wise decision (independent of other voxels) is sensitive to noise and may result in irregular object boundaries or holes and islands in the segmentation result. Therefore, regularization approaches make up the second building block of segmentation algorithms, which are employed to favour smooth segmentation. As medical image modalities are often corrupted by different types of degradation, e.g. low SNR, PVE, and patient motion artifacts, incorporating prior knowledge about the segmented objects becomes necessary to disambiguate the raw data during segmentation. Hence, a third building block is incorporated into segmentation methods, which is essentially designed to favour segmentations belonging to a particular object class.

One of the important algorithms incorporating prior knowledge is the seminal work of active shape model by Cootes et al. [7]. Here, coherently placed landmark points on all the training contours are used to reduce the number of degrees of freedom of the principal modes of variation during principal component analysis (PCA). A point (landmark) distribution model is thus created that captures the variability observed in the training set. Although successfully applied to various types of shapes (hands, faces, organs), the reliance on a parameterized representation and the manual positioning of the landmarks, particularly tedious in 3D images, seriously limited the applicability of this method. To circumvent the need for landmarking, Leventon et al. [29] performed PCA on the signed distance functions of the set of training shapes. The resulting statistical shape model was then integrated into a geodesic active contour framework. The result was an energy-minimizing segmentation method, requiring initialization and parameter setting to balance the contributions of different cost terms. Cremers et al. [10] introduced a nonlinear shape prior based on kernel PCA under the assumption that the training data forms a Gaussian distribution following a nonlinear mapping to a high-dimensional feature space. In general, segmentation techniques that incorporated a shape prior required the projection of the current estimate of the segmentation surface into the plausible shape space learned from a training set. The iterative nature increases the computational demand of these algorithms and therefore renders them unsuitable for interactive exploration.

In applications where there is no well-defined relationship between low level information associated with a voxel (e.g. intensity or gradient) and the label that should be assigned to the voxel, a complementary approach for incorporating prior knowledge is adopted. The weak relationship between voxel data and the label assignment is fairly obvious when we remember that we often seek to label *anatomical structures* rather than *tissue types*. For example, different structures composed of the same tissue (e.g. different bones) cannot be distinguished from one another by merely looking at their intensity values in an image. What distinguishes these structures instead is their location and their spatial relationship to other structures. In such cases, spatial information (e.g. neighborhood relationships) must to be taken into consideration and included in the segmentation process. A common embodiment of such relationships is through what is called an *atlas*. Similar to how the world atlas captures geographical locations of and border relationships between countries, an anatomical atlas captures the locations and shapes of anatomical structures, and the spatial relationships between them. An anatomical atlas can, for example, be generated by manually segmenting one image. It can also be obtained by integrating information from multiple segmented images, for example from different individuals, yielding a probabilistic description rather than a crisp one. Given an atlas, an image can be segmented by mapping its coordinate space to that of the atlas in an anatomically correct way, a process commonly referred to as (atlas-to-image) registration. Labeling an image by mapping it to an atlas is consequently known as atlas-based segmentation, or registration-based segmentation [44]. Collins et al. [6] introduced an automated procedure called ANIMAL (Automatic Nonlinear Image Matching and Anatomical Labeling) to objectively segment gross anatomical structures from 3D MR images of normal brains. The procedure is based on a nonlinear registration of the novel image with a previously labeled target brain, followed by applying a numerical inverse transformation of the labels to the native MRI space. Baillard et al. [1] employed an automatic registration of the current image to a reference pre-segmented image to initialize the evolution of a level-set surface as an alternative solution to manual initialization. Atlas images rely largely on building intensity and shape priors from datasets of healthy subjects. In general, the suspicious regions, which are typically of greater interest than normal-behaving regions, are difficult to identify since they come from the more challenging pathological cases. These pathological cases usually do not have well-defined models. The core goal of this paper is to identify and highlight those regions which disagree with the prior in terms of either their shape or their appearance.

2.2 Transfer functions in volume visualization

In the visualization community, volumetric segmentation is chiefly obtained through the application of transfer functions. Different methods have been proposed for transfer function design in direct volume rendering [13]. For 3D medical volumes, data-driven features such as grey-level intensity, gradient values [26], curvature [25], spatial information [43, 31], and the occlusion spectrum [9] are represented as multidimensional histograms. Also, transfer functions play an important role in visualizing time-varying datasets [21, 15]. To perform the segmentation using transfer functions, the user is required to identify different regions in the feature space, which might be challenging when the data under study is noisy or the anatomical structures within are rather complex. Tzeng et al. [50] used ISODATA hierarchical clustering to classify the volumetric data and later [49] incorporated an artificial neural network and a support vector machine as a supervised learning mechanism for high quality classification. As a supervised learning algorithm, only a small number of voxels was used for training and all the remaining voxels were used as a test set. Thus, a longer interaction time is needed for complex structures. Blaas et al. [3] introduced interactive dynamic coordinated views to facilitate the exploration of multifield medical datasets using pattern classification and clustering approaches. A major assumption behind most of the transfer function widgets for medical imaging is that the low-level image information (e.g. intensity, gradient, curvature, etc.) is sufficient to identify different objects of interest. Low level information is usually not enough to precisely identify complex objects under the different image degradation factors. Further, editing the classification result in all previous algorithms is solely based on user knowledge of the data with no guidance to enhance the classification result more quickly. Recently, Praßni et al. [40] introduced a shape-based transfer function that assigns skeleton regions to one of three primitive classes: tube, blob, or surface. Our method is different from [40] in that we try to model the whole global shape observed in the training set instead of primitive smaller shapes.

2.3 Uncertainty visualization

Uncertainty visualization is considered one of the top visualization research challenges [22]. Few research papers tackled the problem of

visualizing the classification uncertainty given a probabilistic segmentation result. Kniss et al. [27] proposed a Bayesian risk minimization framework, in which the final classification can be altered by changing the risk of each class instead of applying the maximum-a-posteriori (MAP) estimate directly on the probabilistic result. Lundström et al. [30] applied different animation schemes to convey the uncertainty at each voxel obtained from the likelihood information. Recently, Saad et al. [46] allowed the user to visualize different tissue relationships by analyzing the probabilistic field. Although we consider the methods of Kniss et al. [27], Lundström et al. [30] and Saad et al. [46] to be the closest works to our contribution, we differ in several aspects: From a Bayesian perspective, instead of only trying to explore the likelihood information and considering a uniform prior, we incorporate more informative prior for both shape and appearance learned from expert-segmented images of a training population. Further, we explore the relationship between the likelihood and prior terms to allow the identification of regions of disagreement with the prior. Those regions might be the result of misclassification and in that case we allow the user to correct it. Alternatively, they might be suspicious regions (e.g. tumors) that need to be brought to the user's attention.

Few papers tackled the segmentation editing problem [37]. This is partly due to the fact that scientific publications on segmentation mostly emphasize the automatic part while necessary manual corrections are considered flaws in the automated process. Kang et al. [23] introduced a set of 3D segmentation editing tools based on morphological operators for hole filling and surface editing. The introduced tools required a good segmentation to start with. Grady et al. [18] formulated the editing task as an energy minimization problem. The user-supplied seeds, the pre-segmentation, and the image content all impacted the resulted edited segmentation. Our proposed approach is different from those methods in that we operate on the probabilistic fields instead of the crisp labelling. We leverage the uncertainty encoded into the segmentation results to highlight those problematic regions that need more attention from the user. For the first time, in this work, we also incorporate the uncertainty information from the disagreement with the learned prior to guide the user in the editing process.

3 METHOD

Our framework consists of two main stages: atlas construction and segmentation uncertainty interaction. In the offline (i.e. no user interaction needed) atlas construction stage, we construct two 2D log histograms: (i) likelihood versus shape prior and (ii) likelihood versus appearance prior using the information from expert-segmented images of a training population. In the second stage, the user can select voxels of those histograms satisfying some meaningful pattern (as we will detail later). In order to localize the effect of the analysis in the spatial domain, the user specifies an axis-aligned region of interest (ROI) using sliders. The interactive brushing and linking technique [5] is used to connect the information represented into the two 2D histograms. The selected voxels are automatically and interactively highlighted in 2D and 3D DVR. Furthermore, the user may correct any misclassification by editing the probability field. This exploration process is iterative, where the user can highlight or edit different regions once or multiple times. Fig 1 shows an overview of the proposed workflow.

Our framework is constructed based on the Bayesian decision theory, a fundamental statistical approach to the problem of pattern classification [12]. This probabilistic approach is based on quantifying the tradeoffs between various classification decisions and the costs that accompany such decisions. Bayes' formula relates the likelihood of a particular decision (or classification) given the current observed features with the prior information about the classes involved in the decision. In the context of image segmentation, Bayes' theorem is formulated as follows

$$P(C_i|X,F) = \frac{P(X,F|C_i)P(C_i)}{\sum_{i=1}^{K} P(X,F|C_i)P(C_i)}. \quad (1)$$

where X is a spatial location in $\Re^3$ (i.e. $X = [x\ y\ z]$) and F is a feature vector associated with X that can be constructed from intensity, gradient, time activity curve information, or other data-derived features. $P(X,F|C_i)$ represents the likelihood term of X and F belonging to class C_i out of K classes. $P(C_i)$ represents the prior term, which models our knowledge about class C_i. $P(C_i|X,F)$ represents the posterior probability term, which captures the relationship between the likelihood of a particular class assignment to X given the data-derived features associated with X and the prior terms. $\sum_{i=1}^{K} P(X,F|C_i)P(C_i)$ represents the evidence, which acts as a normalization factor guaranteeing that $\sum_{i=1}^{K} P(C_i|X,F) = 1$. The vector of probabilities $[P(C_1|X,F), P(C_2|X,F), \cdots, P(C_K|X,F)]$ calculated at each voxel X makes up the probabilistic segmentation field. The typical approach for obtaining a crisp classification from a probabilistic result is by applying the maximum a posteriori (MAP) Bayesian principle [12], where we assign a voxel at X to the first best guess (FBG) class with the maximum posterior probability. Formally, the FBG class is

$$\mathrm{FBG}(X) = \underset{i \in K}{\operatorname{argmax}}\, P(C_i|X,F). \quad (2)$$

and the corresponding FBG probability is given by

$$P_{\mathrm{FBG}}(X) = \max_{i \in K} P(C_i|X,F). \quad (3)$$

In the visualization community [27, 30, 46], the conditional likelihood probability $P(X,F|C_i)$ term is the decisive factor in obtaining the final classification (i.e. maximum likelihood estimation). Thus, the prior term is usually assumed to be uniform and identical across classes, i.e. $P(C_i) = \frac{1}{K}$. In this paper, we construct informative priors guided by shape and appearance knowledge learned from expert-segmented images of a training population. $P(C_i;X)$ captures the shape prior parameterized by the spatial location X, i.e. even before the image data is collected, every position X has a particular likelihood of being assigned to each class. These a priori probabilities are typically assigned to a template (e.g. a template brain) forming what we refer to as a probabilistic atlas in Sec. 3.1 [48, 14, 32]. $P(C_i;F)$ represents the appearance prior parameterized by the feature vector F. We assume that the likelihood term $P(X,F|C_i)$ is obtained from a probabilistic segmentation technique, such as Zhang et al. [51] or Grady [17].

In atlas-based segmentation [44], an energy functional is constructed from the likelihood and the prior terms by taking the log of Eq. 1,

$$E(C_i) = \log(P(X,F|C_i)) + \lambda(X)\log(P(C_i)) \quad (4)$$

where $\lambda(X)$ is introduced as a weighting factor between different terms. The denominator term $P(X,F)$ is ignored in the energy functional construction as it does not depend on the unknown class. This typically results in a difficult optimization problem with a complex energy landscape, i.e. numerous local minima. To minimize the number of unknowns to be optimized, the weighting factor is taken to be spatially invariant, i.e. $\lambda(X) = \lambda$, and is usually set empirically based on trial and error, training data, or contextual information [34, 16, 42]. Also, $P(C_i)$ is further decomposed into multiple terms (i.e. $P(C_i;X)$ and $P(C_i;F)$) capturing different prior information and, hence, increasing the complexity of the optimization problem.

In this work, we construct the likelihood and prior terms independently using standard techniques in the medical image analysis literature as shown in the subsequent sections. This step is usually performed offline, where there is no user interaction required, allowing the incorporation of computationally expensive techniques. In the online stage, instead of attempting to automatically solve a complex optimization problem to balance the weights between the likelihood and prior terms, a route plagued with difficulties, we allow the user to visually explore the information encoded into those terms; the chief constituents of the objective function.

3.1 Shape prior modeling

In order to construct the shape prior $P(C_i;X)$ from the training images, a training set of images (e.g. gray-level MRI) containing the object of

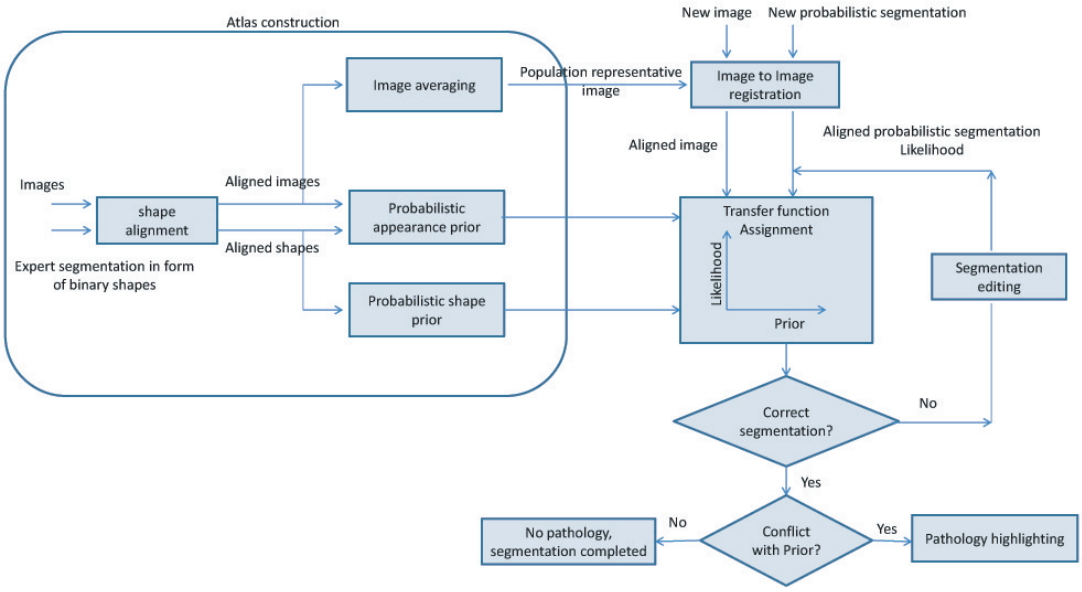


Fig. 1. Shape and appearance probabilistic priors are built from expert-segmented images of a training population. The class likelihood information is obtained from probabilistic segmentation techniques. 2D histogram widgets are used to explore the relationship between the likelihood and prior terms to allow for the identification of regions of disagreement with the prior. Those regions are potentially the result of misclassification, in which case the user can edit the result and provide a satisfactory, correct result. Alternatively, those regions might be suspicious or pathological (e.g. tumors), which must then be highlighted in order to draw the user's attention to them and react appropriately.

interest is collected. In addition, the training set would include data representing the expert segmentation of those images. The segmentations are typically provided in the form of binary images (also called characteristic functions) capturing the shape (and size, location, and rotation) of the object of interest. The binary shape images are then brought into alignment by applying translation, rotation, and scaling mapping them into a common coordinate frame of reference (i.e. removing variability due to pose and retaining only shape variability). A suitable shape model should represent the detailed shape information, be able to model possible shape variations and reflect the shape variability on a local scale (i.e. at each voxel position) which is suitable for the proposed interaction technique. In this work, we take as an input any form of shape prior as long as it provides probabilistic results. The Active Shape Model [7] provides only a global view of the shape variability through PCA and requires a correspondence between landmarks in training images. Hence, we adopt the method proposed recently by Hamarneh and Li [19], who introduced a modification of the idea of a probabilistic atlas by incorporating additional information derived from the distance transform. The probability value at each voxel position X is proportional to the number of shapes containing (or enclosing) X after alignment. The distance transform accounts for possible shape variability unseen in the training images by giving the voxels outside the training shapes non-zero probability. First, a shape histogram $SH(X)$ is obtained by adding the values (0 or 1) of corresponding voxels of the set of aligned binary shape images. $SH(X)$ is then normalized to have a maximum value of 1. We use the shape histogram SH and the distance transform to build a probabilistic shape prior $P(C_i;X)$ as follows,

$$P(C_i;X) = \left\{ \begin{array}{ll} s_2 + s_1 \cdot SH(X) & \text{if } X \in W \\ s_2 \cdot (1 - DIST(X)) & \text{otherwise} \end{array} \right\} \quad (5)$$

where W is the set of nonzero pixels in the shape histogram SH. $DIST(X)$ is the normalized Euclidean distance transform for the binary image having W as foreground. The two scalars, s_1 and s_2, satisfy $s_1 + s_2 = 1$ and determine the weight of the shape information when constructing the probability map. Throughout the paper, we used $s_1 = 0.9$ and $s_2 = 0.1$.

3.2 Appearance prior modeling

In addition to capturing shape variation information from the expert-segmented binary images, we model the appearance knowledge from the intensity patch of the object of interest in the training set of images (e.g. the gray level intensity values in an MR image, the Hounsfield units in a CT scan, or the time activity curves (TAC) in a dPET or dSPECT image). In medical images, anatomical structures and functional regions have characteristic intensity features that can be utilized when locating similar structures in new images.

Given the feature vector F of length n from all the training images, we need to build a probabilistic appearance model. Unlike the Active Appearance Model [8], which provides a global view of the whole object appearance, we need to model the probability distribution function of the feature vectors at every position within the training images. In this work we model this distribution using a multivariate Gaussian. Despite its simplicity, it is fast to compute and requires a few number of model parameters to be stored during the user interaction. Our framework can accommodate more elaborate but computationally expensive techniques such as non-parametric kernel density estimation [12] or support vector machine for density estimation [35]. We construct a multivariate Gaussian distribution as follows

$$P(C_i;F) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma_i)^{\frac{1}{2}}} \exp(-\frac{1}{2}(F-\mu_\mathbf{i})^T \Sigma_i^{-1}(F-\mu_\mathbf{i})), \quad (6)$$

where μ_i and Σ_i are the $n \times 1$ sample mean vector and the $n \times n$ sample covariance matrix of the observed features of class C_i in the training set, respectively. The superscript ^{T} and det(.) denote matrix transposition and determinant, respectively.

3.3 Atlas to new image registration

Given a new image to be analyzed, we first perform an affine registration to bring this image into the coordinate frame of the atlas (i.e. the common frame of reference of all the aligned training images). Then, we average all the aligned training images (e.g. the gray level MRIs) to obtain a population representative image. The resulting transformation from registering the population representative image to the new image is applied to the shape prior $P(C_i;X)$. This is an important step

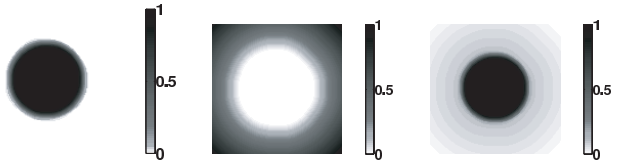
Fig. 2. Synthetic example. a) piecewise constant image, b) blurred with a Gaussian kernel, c) Gaussian noise added.

performed to ensure that the new image coordinates are mapped onto those of the atlas. Note that this step is applied offline (i.e. no user intervention required) while the image is stored on a data server which is common in a hospital setting.

3.4 Algorithm demonstration using a synthetic example

We introduce a synthetic example that is simple enough to convey the main ideas of our approach while still comprehensive enough to simulate the common image generation models and degradation in medical imaging [4]. A grey-level image of size 128×128 consists of one disk region representing one material in addition to the background. We start with a piece-wise constant model in each region with grey level value 30 while the background receives the grey level of zero (Fig 2(a)). We blur the image with a rotationally symmetric Gaussian lowpass filter of size 30×30 pixels with standard deviation 3 pixels (Fig 2(b)). Gaussian noise is then added with a variance of 4 (Fig 2(c)).

We generate a population of 100 images similar to Fig 2(c) but with different noise realizations. The disk center for each image has been randomly translated in the x and y directions with a random amount of up to ± 2 pixels using a normal distribution with mean 0 and standard deviation 2. Fig 3(a) shows the normalized histogram of the population. Fig 3(b) shows the normalized distance transform having W (cf. Eq. 5) as foreground. Fig 3(c) shows the shape prior after applying Eq. 5.



(a) (b) (c)

Fig. 3. Synthetic example of shape prior modeling. a) Normalized shape histogram SH where the intensity represents the number of shapes constructed by adding the values (0 or 1) of corresponding voxels of the set of aligned binary shape images, b) Normalized distance transform image having W as the foreground. The intensity represents the Euclidean distance from the boundary, c) Shape prior map where the intensity represents $P(C_i;X)$ (Eq. 6).

In order to model the appearance of the training set, we consider only grey-level intensity, i.e. $n = 1$ (*cf.* Eq. 6). Fig 4 shows the Gaussian fitting of the intensity histogram for the image population. Fig 5 shows different segmentation cases that vary from normal shape and appearance case (Fig 5(a)): An abnormal shape case (Fig 5(b) and Fig 5(c)), an abnormal appearance case (Fig 5(d)) and both abnormal shape and appearance case (Fig 5(e)). We segment all the images shown in Fig 5 with a mixture of two Gaussians centered at the known means of the two main regions with variance of 56 to account for the PVE. The mixture of Gaussians assumption is supported by the intensity histogram of Fig 6(a), which clearly shows a mixture of Gaussians. Fig 6(b) shows the unnormalized mixture of Gaussians used in the segmentation process, where the x-axis represents the grey level value and the y-axis is the probability for each region marked with a distinct color. Fig 6(c) shows the normalized version of the mixture when the probabilities sum to 1. The normalized probabilities will be the actual output of a probabilistic segmentation algorithm which is the basis of the likelihood term. Note that the intensity range of the suspicious region representing the appearance abnormality in Fig 5(d) and Fig 5(e) falls in-between the intensity ranges of the two main materials.

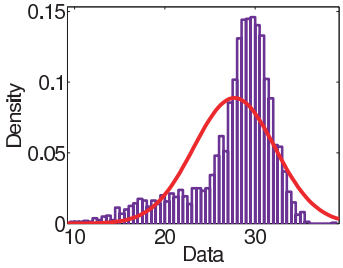


Fig. 4. Synthetic example appearance prior modeling. It shows the intensity histogram of the pixel contained inside the object of interest. The fitted Gaussian is shown in solid red.

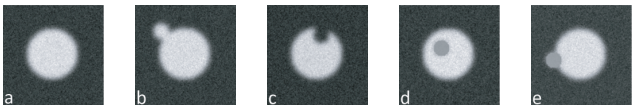
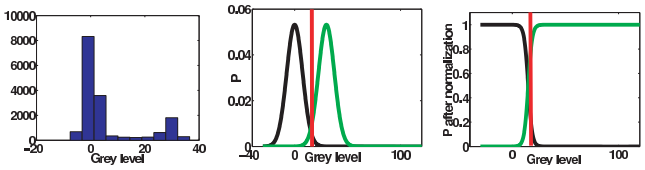


Fig. 5. Synthetic example segmentation cases. a) normal case. b) abnormal shape case. c) abnormal shape case. d) abnormal appearance case. e) abnormal shape and appearance.



(a) (b) (c)

Fig. 6. a) Grey level histogram shows a mixture of Gaussians, b) unnormalized mixture of Gaussians used to segment the images in Fig 5, c) normalized mixture of Gaussians represents the likelihood term $P(X,F|C_i)$. The red bar represent the intensity value of the a small suspicious, tumor-like region shown in Fig 5(d) and Fig 5(e).

Fig 7 shows the analysis of the proposed histograms with respect to the segmentation result of the image shown in Fig 5(a). It demonstrates the expected patterns in the histogram in the normal cases where the likelihood information mostly agrees with the prior information. Fig 7(c) shows the maximum likelihood estimate results for the normal case. Fig 7(a) shows the likelihood vs the shape prior. There is a strong correlation between the likelihood and shape prior where the likelihood result Fig 7(c) agrees with the shape prior map Fig 3(c). Fig 7(b) shows the likelihood vs the appearance prior. There is an agreement between the two but not as strong as the shape prior due to the additive noise nature and partial volume effect along the object boundaries that will be demonstrated in the third and forth rows of Fig 7. The second row of Fig 7 shows the selection of the voxels with high shape prior while selecting the whole appearance prior range. It shows that the whole disk is captured. The third row of Fig 7 shows the selection of a high shape prior in combination with a high appearance prior capturing the core of the organ as shown in Fig 7(i). The last row of Fig 7 shows the selection of a high shape prior in combination with a low appearance prior capturing the boundary voxels which suffer mainly from the partial volume effect as well as some noisy pixels in the center of the object.

Fig 8 shows the analysis of the proposed histograms with respect to the segmentation result of the image shown in Fig 5(b). A strong pattern is observed in the likelihood vs shape prior histogram. It shows a number of voxels with a high likelihood belonging to the object of interest while disagreeing with the shape prior information. Selecting those voxels in Fig 8(b) reveals a shape abnormality not observed in the training set (Fig 8(d)).

Fig 9 shows the analysis of the proposed histograms with respect to the segmentation result of the image shown in Fig 5(c). An abnormal pattern is observed in the likelihood vs shape prior histogram. It shows a number of voxels with low likelihood belonging to the object of interest while strongly agreeing with the shape prior information. Selecting those voxels in Fig 9(b) reveals a shape abnormality not observed in the training set (Fig 9(d)).

Fig 10 shows the analysis of the proposed histograms with respect to the segmentation result of the image shown in Fig 5(d). A strong abnormal cluster is observed in the likelihood vs appearance prior histogram while the shape prior looks normal with strong agreement with the likelihood. It shows a number of voxels with high likelihood belonging to the object of interest while disagreeing with the appearance prior information. Selecting those voxels in Fig 10(c) reveals an appearance abnormality not captured by the likelihood information only.

Fig 11 shows the analysis of the proposed histograms with respect to the segmentation result of the image shown in Fig 5(e). A strong abnormal cluster is observed in both histograms. It shows a number of voxels with high likelihood belonging to the object of interest while strongly disagreeing with the shape and appearance prior information. Selecting those voxels in Fig 11(b) and Fig 11(c) reveals an abnormal behavior in both shape and appearance.

4 Results

In this section, we demonstrate the effectiveness of our framework in analyzing and visualizing different probabilistic segmentations of static and time-varying medical imaging datasets. We will show how we can highlight suspicious regions according to the mismatch with the shape and appearance prior learned from a training set.

4.1 Brain dynamic PET study

In dPET imaging, a series of 3D images are reconstructed from list-mode data obtained by Gamma coincidence detectors. Kinetic modeling is the process of applying mathematical models to analyze the temporal tracer activity, in order to extract clinically or experimentally relevant information. We will analyze the probabilistic segmentation result obtained from the application of a kinetic modeling based probabilistic K-means algorithm [47]. The $4D$ $[^{11}C]$ Raclopride dPET dataset size is $128 \times 128 \times 63$ with 26 time steps and has a voxel size of $2.11 \times 2.11 \times 2.42$ mm^3. The dataset is segmented into six regions ($K = 6$): background, skull, white matter, grey matter, cerebellum and putamen. It is clear from Fig 12(a) that the dataset is suffering from low SNR as well as severe PVE in contrast with a typical structural MRI image (Fig 12(b)). These degradations cause K-means clustering to overestimate the putamen (Fig 12(c) and Fig 12(d)) where a large number of voxels from the surrounding structures have been misclassified as putamen. Fig 12(e) shows the ground truth segmentation with anatomical labels of the putamen. Accurate segmentation of the putamen is crucial for the diagnosis of Parkinson's disease. Fourteen expert segmented dPET datasets with the putamen delineated have been used in the training stage. Fig 13 shows different putamen structures of 4 different patients from the training set. Fig 14 shows how the interaction with the likelihood vs the shape prior histogram can reveal the misclassification results witnessed in Fig 12(d). The upper row shows the voxels corresponding to the selection on the bottom row histograms. The histograms in the bottom row show a large number of voxels disagreeing with the shape prior (i.e. having a low shape prior value). Fig 14(e), (f) and (g) include all the voxels having the putamen as the FBG whereas Fig 14(h) includes all the voxels. Selecting the voxels as shown in Fig 14(h) demonstrates that we can recover finer details that were not possible in Fig 14(c) (i.e. caudate nucleus) or using ProbExplorer [46].

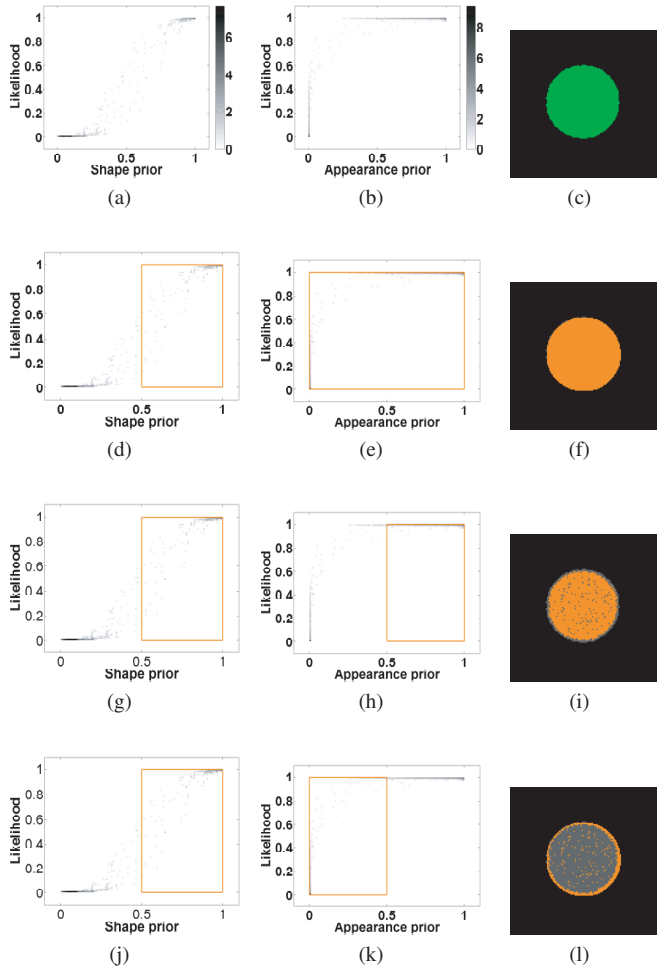


Fig. 7. Analysis of the normal case segmentation of the image shown in Fig 5(a). It demonstrates the patterns observed in the histogram when analyzing normal behavior.

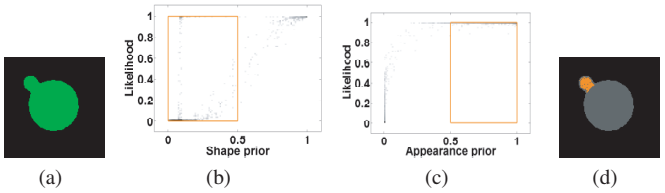


Fig. 8. Analysis of the abnormal shape case segmentation of the image shown in Fig 5(b). It shows a number of voxels with high likelihood belonging to the object of interest while disagreeing with the shape prior information. Selecting those voxels reveals a shape abnormality not seen in the training set.

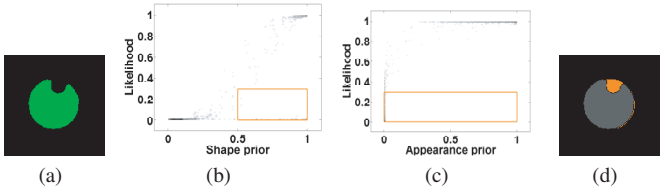
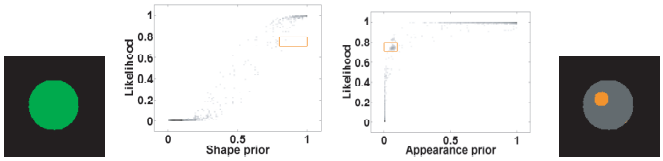
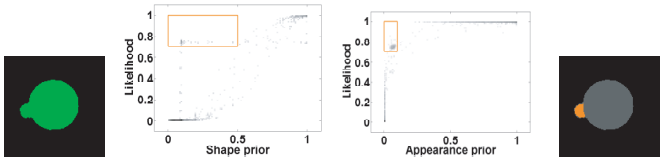


Fig. 9. Analysis of the abnormal shape case segmentation of the image shown in Fig 5(c). It shows a number of voxels with low likelihood belonging to the object of interest while strongly agreeing with the shape prior information. Selecting those voxels reveals a shape abnormality not seen in the training set.



(a) (b) (c) (d)

Fig. 10. Analysis of the abnormal appearance case segmentation of the image shown in Fig 5(d). It shows a number of voxels with high likelihood belonging to the object of interest while disagreeing with the appearance prior information. Selecting those voxels reveals an appearance abnormality not captured by the likelihood information only.



(a) (b) (c) (d)

Fig. 11. Analysis of the abnormal shape and appearance case segmentation of the image shown in Fig 5(e). It shows a number of voxels with high likelihood belonging to the object of interest while strongly disagreeing with the shape and appearance prior information. Selecting those voxels reveals a shape and appearance abnormality.

In order to quantitatively evaluate the segmentation accuracy given the availability of the ground truth segmentation, we calculated the Dice similarity coefficient [11] to measure the overlap between a segmented region and ground truth, with a value of 1 corresponding to a perfect overlap. We were able to increase the Dice similarity coefficient from 0.32 in Fig 14(a) to 0.65 in Fig 14(c) and 0.75 in Fig 14(d).

The results shown in Fig 15 are obtained in a fashion similar to Fig 14. However, here a more accurate but more computationally expensive probabilistic segmentation algorithm was used [45]. We note that the original segmentation in Fig 15(a) is better than that in Fig 14(a). Nevertheless, we can improve it further without rerunning the segmentation algorithm. The Dice similarity coefficient increases from 0.6 in Fig 15(a) to 0.75 in Fig 15(c).

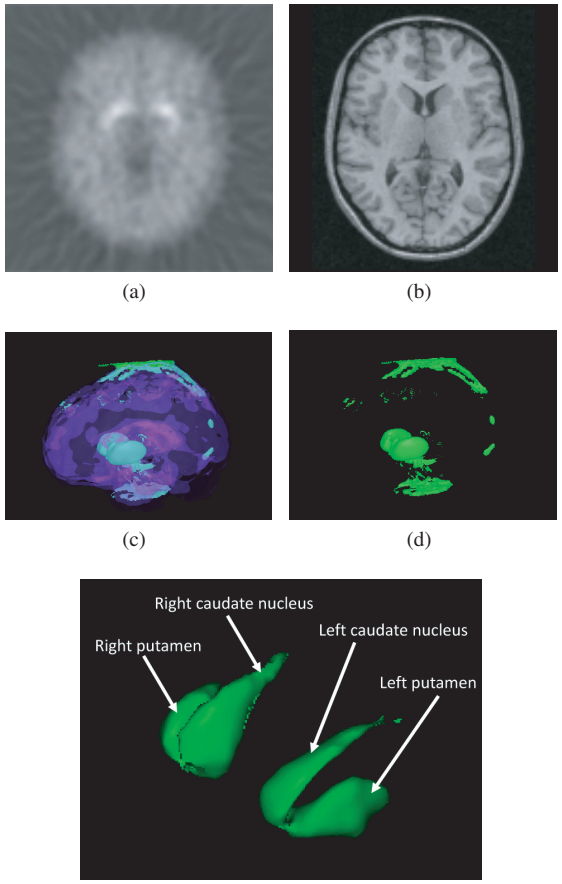
4.2 Brain MRI study

The following study shows a simulated probabilistic field from the BrainWeb database with 20 MR brain images of normal subjects[1]. The probabilistic segmentation fields were used to obtain the shape prior probability. A novel MR image containing a brain tumor (obtained according to [41]) is segmented producing the likelihood probability field. Fig 16 shows the analysis result of the white matter structure. Fig 16(a) shows the isosurface of the maximum likelihood where the tumor is not immediately visible. Fig 16(b) shows the likelihood vs the shape prior where a significant number of voxels show high shape prior while having low likelihood probability. Selecting and highlighting those voxels reveal the pathological and clinically-verified tumor voxels. Fig 17 shows the results of analyzing the grey matter instead of the white matter to reveal the tumor which was not possible using only the maximum likelihood information in Fig 17(a).

4.3 Renal dynamic SPECT study

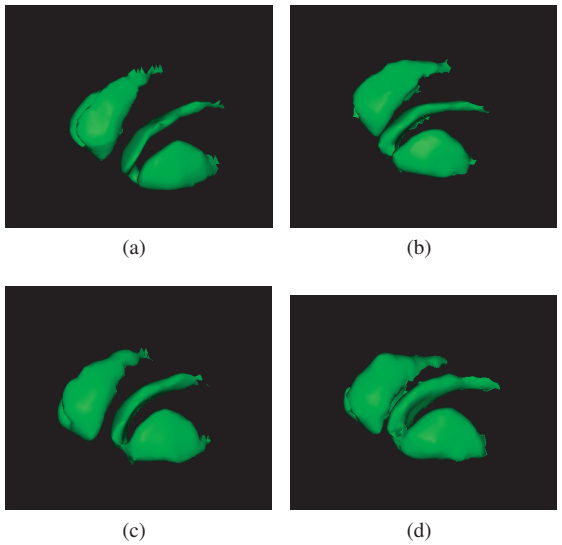
A renal dynamic SPECT study results in a time series of 3D images. From these images, the clinician's goal is to extract information about the spatio-temporal behavior of a radioactive tracer (e.g. $^{99m}Tc-DTPA$), which in turn is used to assess the renal system function. We used a 4D image of size $64 \times 64 \times 64$ with 48 time steps with isotropic voxel size of 2 mm. A 2D coronal slice is shown in Fig 18(a).

[1] http://www.bic.mni.mcgill.ca/brainweb/



(a) (b) (c) (d) (e)

Fig. 12. Dynamic PET case study showing misclassification results for the putamen structure. a) 2D axial slice out of the 4D dataset showing low SNR and PVE, b) Typical structural MRI image, c) DVR, d) overestimated putamen, e) ground truth segmentation with anatomical labeling.



(a) (b) (c) (d)

Fig. 13. Dynamic PET training shapes with ground truth segmentation of the putamen structure. a) patient2, b) patient6, c) patient10 and d) patient15.

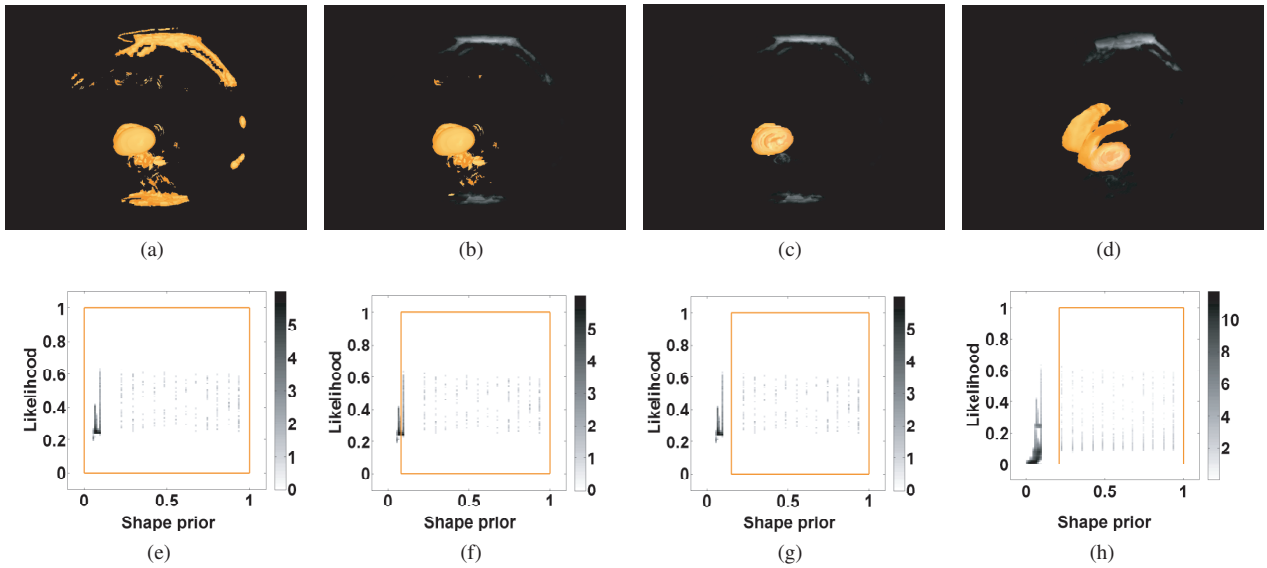


Fig. 14. User-interaction with the proposed likelihood vs shape 2D histograms for a dPET data. It shows how the emphasis on the shape prior, by including only the voxels with high shape prior, from left to right allows the recovery of the correct putamen from the misclassification results. The upper row shows the voxels corresponding to the selection on the bottom row histograms. The histograms in the bottom row show a large number of voxels disagreeing with the shape prior (i.e. having a low shape prior value). e), f) and g) include all the voxels having the putamen as the FBG, whereas h) includes all the voxels. It shows that we can recover finer details that were not possible in c) (i.e. caudate nucleus) or using ProbExplorer [46].

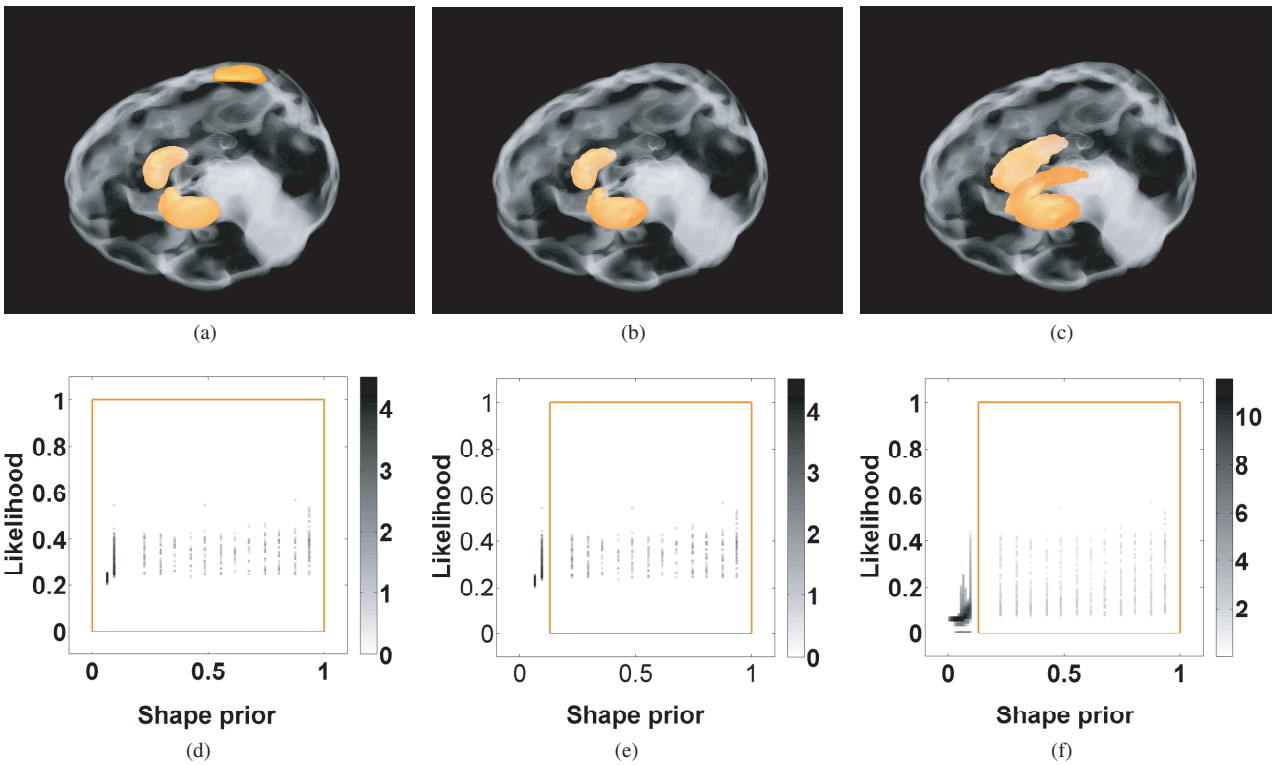


Fig. 15. The results shown here are obtained in a fashion similar to Fig 14. However, here a more accurate but more computationally expensive probabilistic segmentation algorithm was used [45]. We note that the original segmentation in Fig 15(a) is better than that in Fig 14(a). Nevertheless, we can improve it further without rerunning the segmentation algorithm.

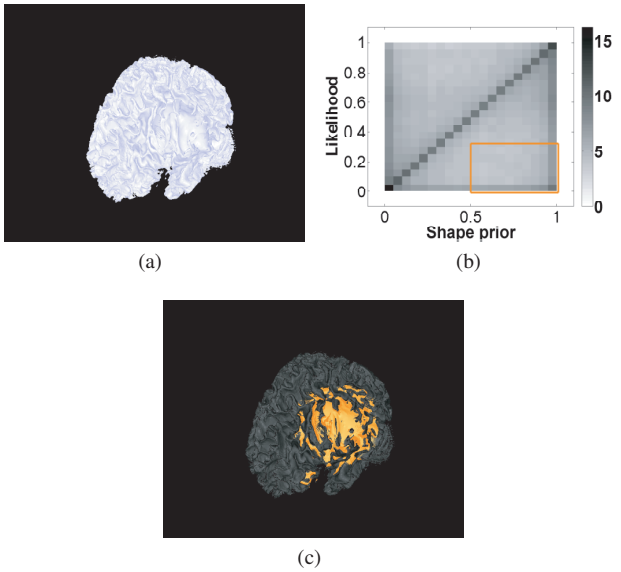


Fig. 16. BrainWeb white matter analysis. a) The isosurface of the maximum likelihood, b) likelihood vs shape prior and c) highlighting the voxels with disagreement between the shape prior and likelihood reveals the tumor location.

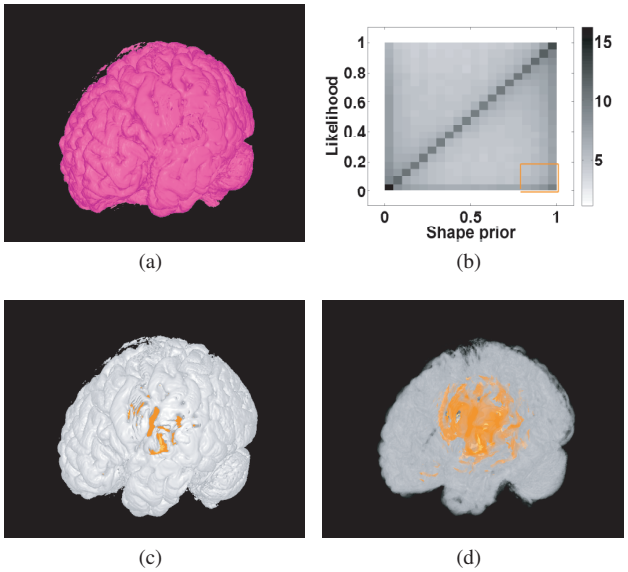


Fig. 17. BrainWeb grey matter analysis. It demonstrates the tumor detection process by highlighting the voxels with disagreement between the likelihood and the shape prior. a) Isosurface of the maximum likelihood, b) voxel selection, c) DVR, d) changing the opacity to highlight the tumor voxels.

In dSPECT, a segmentation of the the kidney shape is easily obtained [20]. The key factor of discovering abnormalities is the appearance of the TACs inside the kidney. Thus, we analyze the likelihood vs. the appearance prior of the TACs unlike the likelihood vs. the shape prior (Sec. 4.2 and Sec. 4.1). The appearance prior was calculated based on kidney TACs obtained from normal subjects using Eq. 6 where F is taken to be the TAC at each voxel. In order to construct the likelihood probability, we blur a crisp manual segmentation of the kidney with a Gaussian kernel to obtain a probabilistic field. Fig 18(b) shows an isosurface rendering of the maximum likelihood. By analyzing the likelihood vs appearance prior histogram (Fig 18(c)), we can see that there are some voxels with disagreement between the appearance prior and the likelihood. Highlighting those voxels reveals an abnormal behavior of the lower third part of the left kidney. The small abnormality in the right kidney was verified to be a result of the PVE which results in low appearance prior probabilities. These results were clinically verified [20].

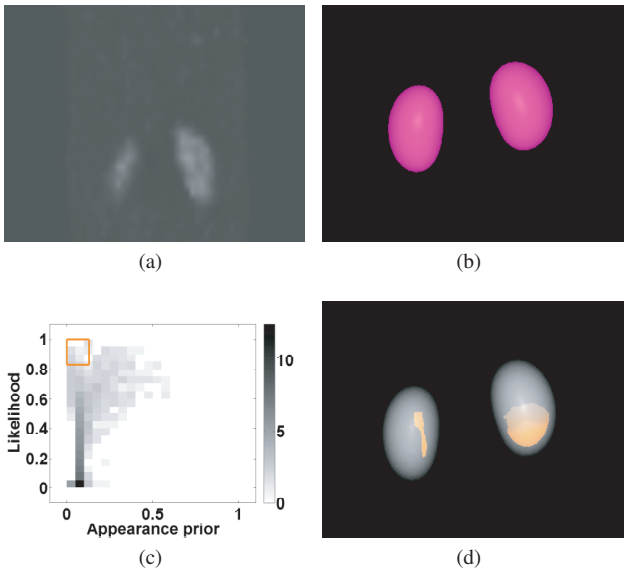


Fig. 18. dSPECT case study with abnormal renal behavior in the lower third part of the left kidney. a) 2D coronal slice out of the 4D dataset, b) the isosurface of the maximum likelihood, c) likelihood vs appearance prior histogram, d) selection of disagreement between the likelihood and appearance reveals abnormality in the lower third part of the kidney. The small abnormality in the right kidney was verified to be a result of the PVE which results in low appearance prior.

5 Conclusion and Discussion

In this paper, we presented a framework for the analysis and visualization of probabilistic segmentation results guided by shape and appearance knowledge learned from expert-segmented images of a training population. We showed how the disagreement between the likelihood terms and the shape and appearance priors can reveal regions of misclassification and abnormal behavior. We demonstrated the efficiency of the algorithm in the context of segmenting multiple simulated and real medical image datasets.

Our focus was based on a single structure atlas, we plan to investigate the effect of multiple atlases modeling multiple structures. This will allow us to take the structures' relationship into account when analyzing the segmentation results.

6 Acknowledgements

This work has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). We would like to thank Prof. Vesna Sossi, director of the Positron Emission Tomography Imaging Group at the University of British Columbia, Canada, Prof. Anna Celler, director of the Medical Imaging Research Group, Department of Radiology, University of British Columbia, Canada, Thomas Humphries and Prof. Manfred Trummer with the Applied and Computational Mathematics Department, Simon Fraser University, Canada for their valuable comments and suggestions.

References

[1] C. Baillard, P. Hellier, and C. Barillot. Segmentation of brain 3D MR images using level sets and dense registration. *Medical Image Analysis*, 5(3):185–194, 2001.

[2] W. A. Barrett and E. N. Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331–341, 1997.

[3] J. Blaas, C. P. Botha, and F. H. Post. Interactive visualization of multi-field medical data using linked physical and feature-space views. In *Proceedings of the joint Eurographics / IEEE VGTC Symposium on Visualization*, pages 123–130, 2007.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimiza-

tion via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[5] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Academic Press, London, 1999.

[6] D. L. Collins and A. C. Evans. ANIMAL: validation and applications of nonlinear registration-based segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(8):1271–1294, 1997.

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding Journal*, 61(1):38–59, 1995.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[9] C. Correa and K.-L. Ma. The occlusion spectrum for volume classification and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1465–1472, 2009.

[10] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford Shah based segmentation. In *Proceedings of the European Conference on Computer Vision*, pages 93–108, 2002.

[11] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[12] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[13] K. Engel, M. Hadwiger, J. M. Kniss, C. Rezk-Salama, and D. Weiskopf. *Real-Time Volume Graphics*. A. K. Peters, Natick, MA, USA, 2006.

[14] A. Evans, M. Kamber, D. Collins, and D. MacDonald. *Magnetic Resonance Scanning and Epilepsy*, chapter An MRI-based probabilistic atlas of neuroanatomy, pages 263–274. Plenum Press, 1994.

[15] Z. Fang, T. Möller, G. Hamarneh, and A. Celler. Visualization and exploration of time-varying medical image data sets. In *Proceedings of Graphics Interface 2007*, pages 281–288, 2007.

[16] T. Finley and T. Joachims. Training structural SVMs when exact inference is intractable. In *Proceedings of the 25th International Conference on Machine Learning*, pages 304–311, 2008.

[17] L. Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

[18] L. Grady and G. Funka-Lea. An energy minimization approach to the data driven editing of presegmented images/volumes. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 888–895, 2006.

[19] G. Hamarneh and X. Li. Watershed segmentation using prior shape and appearance knowledge. *Image and Vision Computing*, 27(1):59–68, 2009.

[20] T. Humphries, A. Saad, A. Celler, T. Möller, and M. Trummer. Segmentation-based regularization of dynamic SPECT reconstructions. In *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 2849–2852, 2009.

[21] T. Jankun-Kelly and K.-L. Ma. A study of transfer function generation for time-varying volume data. In *Proceedings Volume Graphics Workshop*, pages 51–65, 2001.

[22] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, 2004.

[23] Y. Kang, K. Engelke, and W. A. Kalender. Interactive 3D editing tools for image segmentation. *Medical Image Analysis*, 8(1):35–46, 2004.

[24] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[25] G. L. Kindlmann, R. T. Whitaker, T. Tasdizen, and T. Möller. Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Proceedings of IEEE Visualization*, pages 513–520, 2003.

[26] J. Kniss, G. Kindlmann, and C. Hansen. Multi-dimensional transfer functions for interactive volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):270–285, 2002.

[27] J. M. Kniss, R. V. Uitert, A. Stephens, G.-S. Li, T. Tasdizen, and C. Hansen. Statistically quantitative volume visualization. In *Proceedings of IEEE Visualization*, pages 287–294, 2005.

[28] P. Kohli and P. Torr. Measuring uncertainty in graph cut solutions. *Computer Vision and Image Understanding Journal*, 112(1):30–38, 2008.

[29] M. E. Leventon, W. Eric, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 316–323, 2000.

[30] C. Lundström, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1648–1655, 2007.

[31] C. Lundström, P. Ljung, and A. Ynnerman. Local histograms for design of transfer functions in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1570–1579, 2006.

[32] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: Theory and rationale for its development. *NeuroImage*, 2(2, Part 1):89–101, 1995.

[33] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[34] C. Mcintosh and G. Hamarneh. Optimal weights for convex functionals in medical image segmentation. In *Proceedings of the 5th International Symposium on Advances in Visual Computing*, pages 1079–1088, 2009.

[35] R. M. Mohamed and A. A. Farag. Mean field theory for density estimation using support vector. In *Proceedings of the International Conference on Information Fusion*, pages 495–501, 2004.

[36] E. N. Mortensen and W. A. Barrett. Interactive segmentation with intelligent scissors. *Graphics Models and Image Processing*, 60(5):349–384, 1998.

[37] S. Olabarriaga and A. Smeulders. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 5(2):127–142, 2001.

[38] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

[39] M. Poon, G. Hamarneh, and R. Abugharbieh. Efficient interactive 3D livewire segmentation of objects with arbitrarily topologies. *Computerized Medical Imaging and Graphics*, 32:639–650, 2008.

[40] J.-S. Praßni, T. Ropinski, J. Mensmann, and K. H. Hinrichs. Shape-based transfer functions for volume visualization. In *Proceedings of the IEEE Pacific Visualization Symposium*, pages 9–16, 2010.

[41] M. Prastawa, E. Bullitt, and G. Gerig. Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Medical Image Analysis*, 13(2):297–311, 2009.

[42] J. Rao, R. Abugharbieh, and G. Hamarneh. Adaptive regularization for image segmentation using local image curvature cues. In *Proceedings of the European Conference on Computer Vision, in press*, 2010.

[43] S. Roettger, M. Bauer, and M. Stamminger. Spatialized transfer functions. In *Proceedings of IEEE/Eurographics Symposium on Visualization*, pages 271–278, 2005.

[44] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer. *Quo vadis, atlas-based segmentation? In:The Handbook of Medical Image Analysis: Segmentation and Registration Models*. Kluwer Academic / Plenum Publishers, New York, NY, 2005.

[45] A. Saad, G. Hamarneh, T. Möller, and B. Smith. Kinetic modeling based probabilistic segmentation for molecular images. In *Lecture Notes in Computer Science, (MICCAI)*, pages 244–252, 2008.

[46] A. Saad, T. Möller, and G. Hamarneh. ProbExplorer: Uncertainty-guided exploration and editing of probabilistic medical image segmentation. *Computer Graphics Forum*, 29(3):1113–1120, 2010.

[47] A. Saad, B. Smith, G. Hamarneh, and T. Möller. Simultaneous segmentation, kinetic parameter estimation, and uncertainty visualization of dynamic PET images. In *Lecture Notes in Computer Science, (MICCAI)*, pages 726–733, 2007.

[48] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System : An Approach to Cerebral Imaging*. Thieme Medical Publishers, January 1988.

[49] F.-Y. Tzeng, E. Lum, and K.-L. Ma. An intelligent system approach to higher-dimensional classification of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 11(3):273–284, 2005.

[50] F.-Y. Tzeng and K.-L. Ma. A cluster-space visual interface for arbitrary dimensional classification of volume data. In *Proceedings of Eurographics-IEEE TVCG Symposium on Visualization*, pages 17–24, 2004.

[51] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transaction of Medical Imaging*, 20(1):45–57, 2001.