10
Image Segmentation Techniques in Nuclear Medicine Imaging

A.O. Boudraa* and H. Zaidi†

1. Introduction

Recent advances in nuclear medicine imaging systems design has resulted in significant improvements in the areas of anatomical, functional, and dynamic imaging procedures. With these developments, computer-aided diagnosis is becoming a reality. These computer-based tools allow physicians to understand and diagnose human disease through virtual interaction. The role of medical imaging is not limited to visualization and inspection of anatomic structures, but goes beyond that to patient diagnosis, advanced surgical planning, simulation, and radiotherapy planning. In order to identify and quantify the information gathered from the images, regions (also called regions of interest – RoIs) must be precisely delineated and separated out for processing. This process, called *image segmentation*, is the most critical step for automated analysis in medical imaging.[1-3]

Segmentation is the first essential and important step of low level vision. This process partitions the image I into non-intersecting subregions, $R_1, R_2, \ldots, R_C$ such that each region is homogeneous and the union of two non-adjacent regions is homogeneous. A point x_k in a region R_i is connected to x_l if there is a sequence $\{x_k, \ldots, x_l\}$ such that x_j and x_{j+1} are connected and all the points are in R_i. R_i is a connected region if the set of points in R_i has the property that every pair of points is connected. Formally segmentation can be defined as follows:

(a) $\bigcup_{i=1}^{C} R_i = I$;
(b) R_i is a connected region, $i = 1, 2, \ldots, C$;
(c) $R_i \bigcap R_j = \emptyset \ \forall i, j, i \neq j$;
(d) $P(R_i) = \text{TRUE}$ for $i = 1, 2, \ldots, C$;
(e) $P(R_i \cup R_j) = \text{FALSE}$ for $i \neq j$.

*Dr A.O. Boudaa, Ecole Navale, Département Signal, Lanvéoc Poulmic BP600, F 29240 Brest, France
†PD Dr H. Zaidi, Geneva University Hospital, Division of Nuclear Medicine, 1211 Geneva, Switzerland

where P ( ) is a uniformity (homogeneity) predicate and C is the number of regions. A set of regions satisfying conditions (a) through (e) is known as a *partition*. Condition (a) indicates that the segmentation must be complete; i.e. every pixel must be in a region. Condition (b) requires that points in a region must be connected. Condition (c) indicates that the regions must be disjoint. Condition (d) deals with the properties that must be satisfied by the pixels in a segmented region; for example $P(R_i) = \text{TRUE}$ if all pixels in R_i have the same grey level. Finally, condition (e) indicates that regions R_i and R_j are different in the sense of predicate P ( ). In the standard approach, most segmentation algorithms assume that a given pixel or voxel belongs to a single cluster or tissue type (condition (c)). However, the complexity of tissue boundaries causes many pixels to contain a mixture of surrounding tissues (partial volume effect). Thus, image ambiguity within pixels is due to the possible multi-valued levels of brightness in the image. This indeterminacy is due to inherent vagueness rather than randomness. Since image regions (particularly nuclear medicine images) are not always crisply defined, it becomes more convenient and appropriate to view them as fuzzy subsets of the image.[4-7] Thus, the condition (c) is in general not verified:

$$(c') \bigcap_i R_i \neq \emptyset \tag{1}$$

The fuzzy subsets are characterized by the fuzzy membership degree associated to each pixel to belong to one of them. One popular method for assigning multi-subset membership value to pixels, for segmentation or other image processing tasks is the Fuzzy C-Means (FCM).[8]

2. Image Segmentation in Medical Imaging

Medical image segmentation is becoming an increasingly important image processing step for a number of clinical applications including: (a) identification of RoIs such as lesions to measure their volume and thus assess response to therapy;[5] (b) detection of the left ventricle (LV) cavity to determine the ejection fraction;[4,6] (c) volume visualization and quantification of organ uptake[9,10] or uptake defect of the tracer in the myocardium;[7] (d) study of motion or conduction abnormalities of the heart;[11] and (e) attenuation correction in emission tomographic imaging.[12,13] All subsequent interpretation tasks like feature extraction, object recognition, and classification depend largely on the quality of the segmentation output. The level to which the segmentation is carried depends on the problem being solved. That is, segmentation should stop when the regions of interest for a specific application have been isolated. For example, in automated detection of brain lesions using PET images,[5] interest lies in accurate delineation of the lesions with the objective of assessing the true size and the tracer uptake of these lesions. In this case, brain structures such as grey matter and white

matter are oversegmented and thus not well delineated. Segmentation in medical imaging is generally considered as a very difficult problem. This difficulty mainly arises due to the sheer size of the data sets coupled with the complexity and variability of the anatomic organs. The situation is worsened by the shortcomings of these imaging modalities, such as sampling artefacts, noise, and low contrast which may cause the boundaries of anatomical structures to be indistinct and disconnected. Thus, the main challenge of the segmentation process is to accurately extract the contours of the organ or RoI and separate it out from the rest of the data sets.

3. Segmentation Techniques in Nuclear Medicine Imaging

PET and SPECT images give information about the volume distribution of biologically significant radiotracers. The interest lies in the diagnosis of abnormalities in the biochemical process of interest, e.g. metabolic activity. Unfortunately, these images are inherently noisy and provide less quantitative information about the qualitative features discerned by experienced observers. In order to identify and quantify nuclear medicine image information, image regions must be delineated.[14] Compared to MR images, where brain structures such as CSF or grey matter can relatively be easily identified,[15] segmentation of nuclear medicine images is more difficult. Indeed, in addition to the inherently poor spatial resolution and SNR, there is additional bias introduced by physical degrading effects such as scatter, attenuation, partial volume effect, and patient motion during scanning. Several approaches have been proposed for image segmentation,[16,17] but only a few of them can satisfactorily be applied to nuclear medicine images. Image segmentation algorithms generally are based on one of two basic properties of intensity values: *discontinuity* and *similarity*.[17] In the first category, the basic approach is to partition an image based on abrupt changes in intensity, e.g. edges.[18] The main approaches in the second category are based on partitioning an image into regions that are similar according to a set of predefined criteria.[17] Thresholding, clustering, region growing are examples of methods belonging to this category. Segmentation is based on sets of feature that can be extracted from the images such as pixel intensities, which in turn can be used to calculate other features such as edges, and texture. Many segmentation approaches of nuclear images use the grey scale values of the pixels. In this case, pixel intensity input come from a single image. Single image segmentation methods are described in the following sections.

3.1 Thresholding Approaches

Thresholding is one of the simplest and most popular techniques for image segmentation. It can be performed based on global information (e.g. grey level histogram of the entire image) or using local information (e.g. co-occurrence matrix) of the image.[16] Automatic thresholding is important in applications where the speed or the physical conditions prevent interactive human selection of thresholds.[19] If only one threshold is used for the entire image, the process is called *global thresholding*. On the other hand, when the image is partitioned into several subregions and a threshold determined for each of the subregions, it is referred to as *local thresholding*. Techniques of thresholding can also be classified as *bilevel thresholding* and *multithresholding*. In *bilevel thresholding*, the histogram of the image is usually assumed to have one valley between two peaks, respectively. There are various solutions to the problem of locating the intensity threshold that ideally represents the bottom of this sometimes elusive histogram valley. The following strategy is an example of a thresholding method to obtain automatically the optimum threshold value:

1. Select an initial estimate for threshold T;
2. Segment the image using T. This will provide two regions of pixels: R_1 consisting of all pixels with grey level values $>$T and R_2 consisting of pixels with values $<$T;
3. Compute the average grey level values μ_1 and μ_2 for the pixels regions R_1 and R_2;
4. Compute a new threshold value;

$$T = (\mu_1 + \mu_2)/2 \tag{2}$$

5. Repeat Step 2 through 4 until the mean values μ_1 and μ_2 in successive iterations do not change.

Note that the threshold T can be specified using a heuristic technique based on visual inspection of the histogram but this approach is operator-dependent. If the image is noisy, the selection of the threshold is not trivial. Thus, more sophisticated methods have been proposed. Let $p(t) = n_p/N_I$ be the normalized histogram – brightness probability density function (*pdf*). N_I is the number of pixels in the observed image, I, and $n_p(t)$ is the number of pixels of I having grey level value t. We assume that the distribution of grey values in each region of I follows a Gaussian distribution. The observed histogram is a mixture of grey values of pixels of the object(s) and pixels of the background. The histogram may be considered as an estimate of the brightness *pdf*, say, p(t). Then p (t) is a mixture *pdf* given by $p(t) = P_1 \times p_1(t) + P_2 \times p_2(t)$ where $p_1(t) = \sum_{i=0}^{t} p(i)$ and $p_2(t) = 1 - p_1(t)$

are the *pdf*s of the *object* and the *background*, respectively, and P_1 and P_2 are their *a priori* probabilities such that $P_1 + P_2 = 1$. For a Gaussian case, p (t) is written as follows:

$$p(t) = \sum_{i=1}^{2} \frac{P_i}{\sigma_i \sqrt{2\pi}} \exp\left[\frac{(t-\mu_i)^2}{2\sigma_i^2}\right] \tag{3}$$

where μ_i and σ_i are the *a priori* mean values and standard deviations of the two distribution probabilities ($\mu_1 < \mu_2$). Thus, the mixture has five unknown parameters. If all the parameters are known, the optimal threshold is easily determined. A threshold, T, may be defined so that all pixels with grey values below T are considered *object* and those above T are considered *background* pixels. The probability of erroneously classifying a *background* point as an *object* point is $Er_1(T) = \int_{-\infty}^{T} p_2(z)dz$. Similarly, the probability of classifying an *object* point as *background* point is $Er_2(T) = \int_{T}^{+\infty} p_1(z)dz$. The overall probability of erroneously classifying a pixel from one of the two distributions as a pixel from the other distribution, is simply: $Er(T) = P_1 \times Er_1(T) + P_2 \times Er_2(T)$. The threshold for which $Er(T)$ is minimum is given by:

$$\begin{aligned} &dEr(T)/dT = 0 \Rightarrow P_1 \times p_1(T) = P_2 \times p_2(T) \Rightarrow \alpha T^2 + \beta T + \gamma = 0 \\ &\alpha = \sigma_1^2 - \sigma_2^2, \beta = 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2), \gamma = \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 + 2\sigma_1^2\sigma_2^2 \log\left(\frac{\sigma_2 P_1}{\sigma_1 P_2}\right) \end{aligned} \tag{4}$$

The determination of parameters is not a simple matter and often requires numerical solution of a number of simultaneous non-linear equations. If the variances are assumed to be equal, T is given by: $T = (\mu_1 + \mu_2)/2 + \sigma_2/(\mu_1 - \mu_2)\log(P_2/P_1)$. If $P_1 = P_2$, the optimal threshold is simply the average of the means: $T = (\mu_1 + \mu_2)/2$. In this case, the above proposed strategy is found. A well known method for *global* thresholding using the histogram is Otsu's method.[20] This method chooses the optimal threshold by maximizing the *a posterior* between-class variance: $\sigma_B^2(t) = p_1(t)[\mu_1(t) - \mu_0]^2 + p_2(t)[\mu_2(t) - \mu_0]^2 = \frac{[\mu_0 p_1(t) - \mu_1(t)]^2}{p_1(t)[1 - p_1(t)]}$. The optimal threshold T is found by a sequential search for the maximum of $\sigma_B^2(t)$ for values of t where $0 < p_1(t) < 1$. Otsu's method gives satisfactory results when the number of pixels in classes are similar. Reddi *et al.*[21] proposed a method based on the same assumption by selecting a threshold T so that the *a posterior* between-class variance of *object* and *background* regions is maximized. They have shown that the between-class variance $\sigma_B^2(t)$ has a unique maximum and established the condition for this maximum.[21] Thus, $\sigma_B^2(t)$ can be written as follows:

$$\sigma_B^2 = \left(\sum_{z=0}^{t} zp(z)\right)^2 \times (p_1(t))^{-1} + \left(\sum_{z=t+1}^{N-1} zp(z)\right)^2 \times (p_2(t))^{-1} - \mu_0^2 \tag{5}$$

Setting $\partial\sigma_B^2(t)/\partial t = 0$, the following relation is found: $\mu_1(T) + \mu_2(T) = T$, where $\mu_1(T)$ and $\mu_2(T)$ are the mean values below and above the threshold

T. μ_0 is the average brightness. Finding the value of T that satisfies the above relation can be accomplished by an exhaustive sequential search, giving the same result as Otsu's method. Starting with threshold $T = \mu_0$, fast convergence is obtained. Kittler and Illingworth[22] proposed a method that minimizes the Kullback-Leibler (KL) distance which measures the discrepancy D between the probability distribution of the histogram and its approximation:

$$D = (1 + \log 2\pi)/2 - \sum_{i=1}^{2} p_i(t) \log p_i(t) + \left(p_1(t) \log \sigma_1^2 + p_2(t) \log \sigma_2^2\right)/2 \quad (6)$$

$p_i(T)$ is the estimated probability that pixel value falls in the i^{th} region. The threshold T should be chosen to minimize the information measure H. The method of Pun[23,24] is based on the maximum entropy criterion and has been examined in detail by Kapur *et al.*[25] It relies on the assumption that the threshold T corresponds to the maximum value of the sum of the two within-region entropies: $T = \arg\max_t [H_1(t) + H_2(t)]$ where $H_i(t)$ is the Shannon entropy of the i^{th} region. The $H_i(t)$ are given by:

$$H_1(t) = -\sum_{g=0}^{t} \frac{p(g)}{p_1(t)} \log \frac{p(g)}{p_1(t)} \text{ and } H_2(t) = -\sum_{g=t+1}^{N-1} \frac{p(g)}{p_2(t)} \log \frac{p(g)}{p_2(t)} \quad (7)$$

Thresholding is the most common approach used in segmenting SPECT images.[9,26-31] Figure 1(a) shows a transverse PET cardiac slice through the LV of a resting patient. The thresholding result using the method proposed by Reddi *et al.*[21] (T = 104) is shown in Fig. 1(b).

When the image is composed of several regions (clusters) one needs several thresholds for segmentation. Let the image be described by a

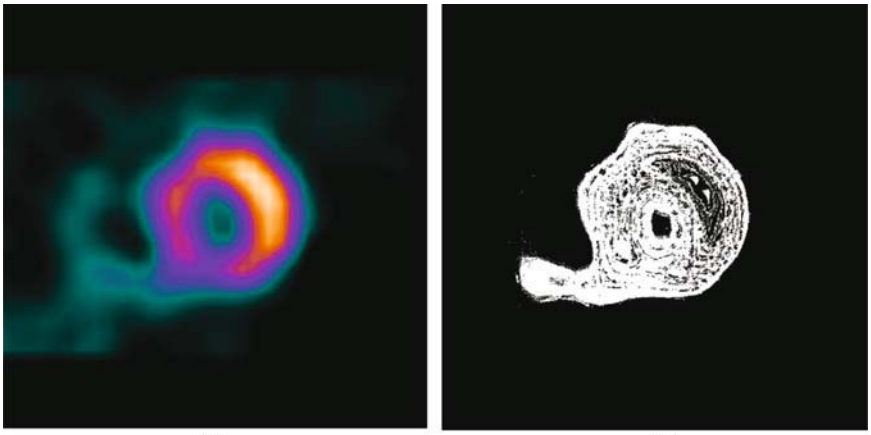

FIGURE 1. Illustration of Reddi *et al.*[21] thresholding method. (a) PET cardiac image. (b) Thresholded image.

two-dimensional function I (x, y), where (x, y) denotes the spatial coordinates and I (x, y) the feature value of pixel at location (x,y). In *multithresholding* one tries to get a set of thresholds $\{t_1, t_2, \ldots, t_K\}$ such that all pixels with $I(x, y) \in [t_i, t_{i+1}], i = 0, 1, \ldots, K$ constitute the i^{th} region type $t_0 = 0$, $t_{K+1} = G - 1$. G is the number of grey levels. There are various methods for *multithresholding*. The method of Reddi *et al.*[21] is fast and stands for a version extended to *multithresholding* of Otsu's method.[20] Wang and Haralick[32] presented a recursive technique for automatic multiple threshold selection. The local properties of the image (edges) are used to find the thresholds. The method of Carlotto[33] determines thresholds by handling the information derived from the changes of zero-crossing in the second derivatives. The method of Papamarkos and Gatos[34] is based on a combination of hill-clustering algorithm and an appropriate linear programming approximate technique. Recently, Fleming and Alaamer[31] have proposed case-specific volume-of-interest segmentation rules. The iterative algorithm calculates local context sensitive thresholds along radial profiles from the centre of gravity of the object.

Figure 2 shows an illustration of the multithresholding method of Reddi *et al.*[21] Figure 2(a) is obtained with two thresholds. Note that the image is not well segmented because the pattern of tracer uptake is not discriminated. Figure 2(b) shows the segmentation result of five thresholds where the pattern of tracer uptake is detected. Figure 2(c) shows the result of eight threshold values corresponding to an over-segmentation of the image.

3.2 Region Growing Approaches

The main drawback of histogram-based region segmentation is that the histogram provides no spatial information. Region growing approaches exploit the important fact that pixels which are close to each other have

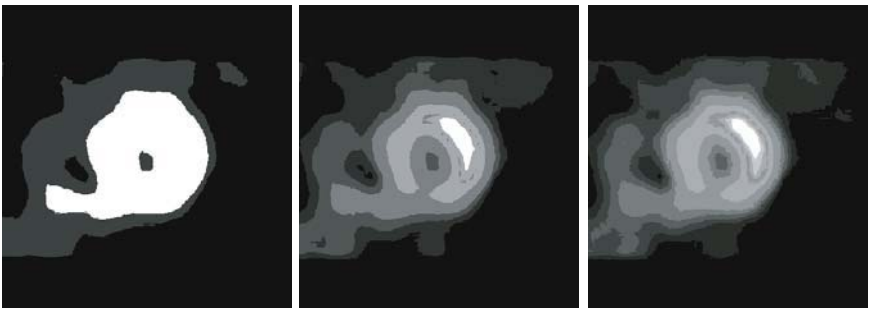

FIGURE 2. Illustration of the Reddi *et al.*[21] multithresholding method. Segmentation of the PET cardiac image shown in Fig. 1(a) with (a) two, (b) five and (c) eight thresholds, respectively.

similar grey values. Region growing is a bottom-up procedure which starts with a set of seed pixels. The aim is to grow a uniform, connected region from each seed. The user selects a pixel or a set of pixels and a region grows out from that (these) seed(s) until some criteria for stopping growth are met. A pixel is added to a region if and only if:

- it has not been assigned to another region;
- it is a neighbour of that region;
- the new region created by addition of the new pixel is still uniform.

The algorithm can be summarised by the following pseudo-code:

Let I be an image, and $R_1, R_2, \ldots, R_C$ a set of regions each consisting of a single seed pixel

Repeat
  For $l = 1, \ldots C$
    For each pixel k at the border R_l
      For all neighbours of pixel k
        Let x,y be the neighbour's coordinates
          Let μ_l be the mean grey level of pixels in R_l
            If neighbour unassigned and $|I(x,y) - \mu_l| \leq \varepsilon$
              Add neighbour to R_l
            Update μ_l
Until no more pixels are being assigned to regions.

The main assumption of the region growing approach is that regions are nearly constant in image intensity. Region growing has been used by Van Laere *et al.*[35] in the anatomic standardization and comparison technique with normal templates of brain SPECT studies. An automated interpretation of myocardial perfusion SPECT using region growing has been reported by Slomka *et al.*[36] The study results show that template-based region growing is a robust technique for volumetric quantification and localization of abnormal regions.

3.3 Classifiers

During the segmentation process, labels are affected to regions. Classification is a pattern recognition technique that deals with association of classes with tissue types, e.g. white matter, grey matter, CSF, ... etc. This step is also called *labelling*. Labels association can be performed by an operator or by an interpreting physician. Thus, pre-segmented images called *training data* are required. Classification seeks to partition *feature space* derived from the image using known labels. The feature space is the range of p-dimensional feature vectors formed at each pixel. The *features* could include pixel intensity, the gradient at a given pixel and so on. The simplest form of classifiers is the Nearest-Neighbour Classifier (NNC) where each pixel is classified in the

same class as the training datum with closest intensity. Examples of classifiers are the k-Nearest Neighbour (k-NN), the Nearest Mean Classifier (NMC),[37] Fisher Linear Discriminant (FLD),[38] Parzen classifier[39] and Support Vector Machine (SVM).[40] The k-NN classifier is the generalization of the NNC, where the pixel is classified according to the majority of the k closest training data. In Parzen window, classification is made according to the majority vote within a predefined window of the feature space centred at the unlabeled pixel (mapped to feature space). SVM is a supervised learning method which is suitable for high dimensional data.[40] Given training examples labelled either ''yes'' or ''no'', a maximum margin hyperplane splits the ''yes'' and ''no'' training examples such that the distance from the closest examples (the margin) to the hyperplane is maximized. The use of the maximum margin hyperplane is motivated by statistical learning theory, which provides a probabilistic test error bound that is minimized when the margin is maximized. If there exists no hyperplane that can split the ''yes'' and ''no'' examples, an SVM will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. NMC calculates the mean of the training vectors for each class while searching and classification are done only against the mean vector for each class. FLD classifier projects high-dimensional data onto a line and performs classification in this one-dimensional space. The projection maximizes the distance between the means of the two classes while minimizing the variance within each class. FLD and NMC separate the object (image) with one hyperplane in feature space into two classes. The discriminant function of NMC is defined as follows:

$$f_{NMC}(x) = (x - \mu_2)^t(x - \mu_2) - (x - \mu_1)^t(x - \mu_1) \tag{8}$$

where x is the object to be classified and μ_1 and μ_2 are the means of the feature vectors in the training set for the classes ''1'' and ''2'', respectively. It spans up an equidistant hyperplane between both class means. Contrarily to NMC, the FLD classifier takes both class means and covariance, A, into account (assumed to be common to both classes):

$$f_{FLD}(x) = [x - (\mu_1 + \mu_2)/2]^t A^{-1}(\mu_1 - \mu_2) \tag{9}$$

When A is the identity matrix, NMC and FLD are equivalent. If the number of training samples is smaller than p, A is singular and pseudo FLD (PFLD) is formed by replacing A by its pseudo inverse, $(A^tA)^{-1}A^t$. Stoeckel *et al.*[41] proposed a method for classifying a single SPECT HMPAO image as being an image of a normal subject or a probable Alzheimer patient. The aim is to provide a tool to assist the interpretation of SPECT images for the diagnosis of dementia of Alzheimer type (DAT). PFLD and NMC are used to assign one of the possible class labels (DAT or normal) to a SPECT image given as an input.[41] The success rates of classification is estimated using a classification error, *leave one out*.[42] As reported by the authors, PFLD outperforms

NMC, which can be explained by the fact that PFLD takes the covariance into account, and thus the shape of the classes in feature space.[41] The differences in success rates between normal and DAT images might be explained by the presence of nearly two times more normal than DAT images.

3.4 Clustering-based Approaches

The limitation of thresholding techniques is that they only apply to a single-band image, such as greyscale image or a single band of a multi-band image. Data clustering is another effective technique for image segmentation. The advantage of clustering is that it can be applied to a multi-band image such as a colour image or image composed of multiple feature layers. The main disadvantage of clustering techniques is that the appropriate number of clusters should be determined beforehand.[43-45] Clustering algorithms essentially perform the same function as classifiers without the use of training data. Thus, they are called *unsupervised* methods. In order to compensate for the lack of training data, clustering methods iterate between segmenting the image and characterizing the properties of each class. In a sense, clustering methods train themselves using the available data. *Unsupervised* methods include the C-means,[46] the FCM algorithm,[8] and the expectation-maximization (EM) algorithm.[47] *Unsupervised* clustering has shown promise for tumour detection in PET imaging, but algorithm initialization is an important step for reliable clustering results and for reduction of computation time.[5] FCM and C-means are the most commonly used algorithms for nuclear medicine[4-6,48] and transmission[13] image segmentation.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite data set and $C \geq 2$ an integer; n is the number of data points; x_k is a p-dimensional feature vector. R^{Cxn} denote the set of all real $C \times n$ matrices. A fuzzy C-partition of X is represented by a matrix $U = [\mu_{ik}] \in R^{Cxn}$, the entries of which satisfy

$$\begin{aligned} &\mu_{ik} \in [0,1]\ 1 \leq i \leq C;\ 1 \leq k \leq n \\ &\sum_{i=1}^{C} \mu_{ik} = 1;\ 1 \leq k \leq n \\ &0 < \sum_{k=1}^{n} \mu_{ik} < n;\ 1 \leq i \leq C \end{aligned} \tag{10}$$

U can be used to describe the cluster structure of X by interpreting μ_{ik} as the degree of membership of x_k to cluster i. Good partitions U of X may be defined by the minimization of the FCM objective functional:[8]

$$J_m(U,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{C} (\mu_{ik})^m \, \| x_k - v_i \|_A^2 \tag{11}$$

where $m \in [1, +\infty)$ is a weighting exponent called the fuzzifier, $V = (v_1, v_2, \ldots v_c)$ is the vector of the cluster centres. $\| x \|_A = \sqrt{x^T Ax}$ is any inner product norm where A is any positive definite matrix. Approximate optimization of J_m by the FCM algorithm is based on iteration through the following necessary conditions for its local extrema.

FCM theorem.[8] Assume $m \geq 1$ and

$\| x_k - v_i \|_A^2 > 0$; $1 \leq i \leq C$; $1 \leq k \leq n$, (U, V) may minimize J_m only if:

$$\mu_{ik} = \left[\sum_{j=1}^{C} \left(\frac{\| x_k - v_i \|_A}{\| x_k - v_j \|_A} \right)^{\frac{2}{(m-1)}} \right]^{-1} \tag{12}$$

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m} \tag{13}$$

The FCM algorithm consists of iterations alternating between Eqs. (12) and (13). This algorithm converges to either a local minimum or a saddle point of J_m.[8] Boudraa *et al.*[4] have segmented nuclear cardiac images with FCM using as features the spatial information (x, y) and grey level value I (x, y) (in this case $p = 3$). Recently, a fast version of the FCM was proposed.[49] This version is based on one dimensional attribute such as the grey-level. Let H_S be the histogram of I with G-levels. Each pixel has a feature that lies in the discrete set X. In the new formulation, FCM minimizes the following functional, which is very similar to that of Bezdek:[8]

$$J_m(U,V;X) = \sum_{g=0}^{N-1} \sum_{i=1}^{C} \left(\mu_{ig} \right)^m . H_S(g) . \| g - v_i \|_A^2 \tag{14}$$

The FCM only operates on the histogram and hence is faster than the conventional version which processes the whole data set. Thus, the computation of the membership degrees of $H_S(g)$ pixels is reduced to that of only one pixel with g as grey level value. The algorithm is outlined in the following steps:

⟨*FCM*0⟩ Find the maximum and the minimum values of H_S: G_{min} and G_{max}

⟨*FCM*1⟩ Fix the number of clusters C, $2 \leq C \leq G_{max}$, and the threshold ε;

⟨*FCM*2⟩ Find the number of occurrences, $H_S(g)$, of the level g; $g = G_{min}, \ldots, G_{max}$

⟨*FCM*3⟩ Initialize the membership degrees μ_{il} using the $(G_{max} - G_{min} + 1)$ grey levels such that: $\sum_{i=1}^{c} \mu_{ig} = 1$, $g = G_{min}, \ldots, G_{max}$

⟨*FCM*4⟩ Compute the centroid v_i as follows:

$$v_i = \frac{\sum_{g=G_{min}}^{G_{max}} \left(\mu_{ig}\right)^m . H_S(g).g}{\sum_{g=G_{min}}^{G_{max}} \left(\mu_{ig}\right)^m . H_S(g)} \quad i = 1, 2, \ldots, C \tag{15}$$

⟨*FCM5*⟩ Update the membership degrees

$$\tilde{\mu}_{ig} = \left[\sum_{j=1}^{c} \left(\frac{\| g - v_i \|_A}{\| g - v_j \|_A}\right)^{\frac{2}{(m-1)}}\right]^{-1} \tag{16}$$

⟨*FCM6*⟩ Compute the defect measure

$$E = \sum_{i=1}^{C} \sum_{g=G_{min}}^{G_{max}} \left|\tilde{\mu}_{ig} - \mu_{ig}\right| \tag{17}$$

If ($E > \varepsilon$)
{
$\mu_{ig} \leftarrow \tilde{\mu}_{ig}$
goto <*FCM*4>
}
⟨*FCM7*⟩ Defuzzification process.

Figure 3 illustrates an automated outlining of the LV contour in gated radionuclide ventriculography. The FCM is run at the left anterior oblique (LAO) projection providing the best separation of the left and right (RV) ventricles. The number of clusters is estimated during the clustering process using a cluster validity index.[6,11] The clustering is performed with ε set to 10^{-3} (Fig. 3b). The cluster formed by the LV and the RT (RT = RV + Atria) is separated from the background (Fig. 3c) followed by a labelling analysis of the connected components LV and RT. Figure 3d) shows that the LV is well separated. Since the LV cluster has the greatest area of the cardiac

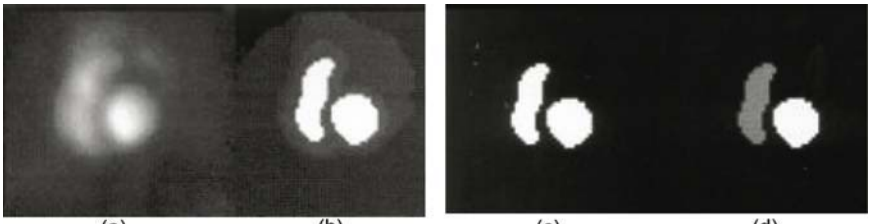

FIGURE 3. LV edge detection by FCM in LAO projection. (a) ED frame. (b) Fuzzy clustering result of the ED frame. (c) Separation of the LV and RT clusters from the background. (d) Labelling of the connected components LV and RT. (Reprinted with permission from ref.[6]).

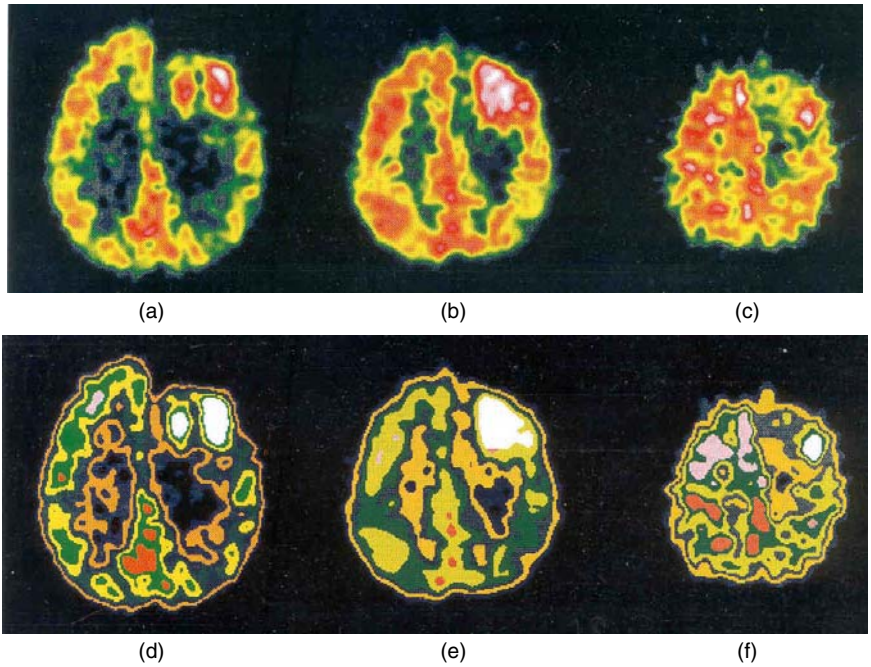

FIGURE 4. Segmentation of FDG-PET images of a patient obtained two weeks after surgery. (a)–(c) are the three slices where the tumour is evidenced by the CT. (d)–(f) are the corresponding segmented images. (Reprinted with permission from ref.[5]).

images sequence in End-diastolic (ED) phase, a framing operation is performed to obtain automatically, a ''box'' enclosing the LV cluster. Thus, the fuzzy clustering of the remaining images of the cardiac cycle is narrowed to this ''box'' with C set to 2 (LV and background).[6]

Figure 4 shows an application of the FCM to localization and quantitation of brain lesions using PET imaging.[5] The CT scans have shown that the brain tumour is only evidenced in three slices (Figs. 4a–c). The FCM is started with an overspecified number of clusters ($C = 20$), believed by the experts to make up the image being segmented. This value leads to an oversegmentation of soft tissues but reduces the chance that the lesion(s) is(are) clustered into classes that contain soft tissues. The fuzzy clustering is followed by a merging process.[5] The brain tumour is hypermetabolic. A careful examination of PET images (Figs. 4a–c) and the corresponding segmented images (Figs. 4d–f) shows that there is a good agreement between the two sets of images and that the tumour is correctly delineated. By comparing the segmented image and the original one, one may easily compute the area (in pixels) of the selected RoI and its total count using the corresponding label affected by the FCM. It has been noticed that the larger is the abnormality, the easier is its detection, even in slices of poor contrast. Nevertheless, a

small tumour which is not readily discerned from its surrounding region (because it has almost the same glucose metabolism) necessitates for its detection, a clustering with high precision.

It should be mentioned that FCM is sensitive to noise and outliers. To handle the problem of noise, the Possibilistic C-Means (PCM)[50] or Maximum Entropy Principle-based Fuzzy Clustering (MEPFC)[51] algorithms can be used. PCM assumes that the membership function of a point in a fuzzy set (or cluster) is absolute, i.e. it is an evaluation of a degree of typicality not depending on the membership values of the same point in other clusters. By contrast, clustering approaches including FCM and MEPFC impose a probabilistic constraint, according to which the sum of the membership values of a point in all the clusters must be equal to one. The PCM algorithm treats each cluster independently of the rest and thus can provide several identical prototypes while completely failing to detect some of other clusters. Furthermore PCM is heavily dependent on initialisation and its ''bandwidth parameter'', η, is difficult to estimate. It was also reported that this algorithm has the undesirable tendency to produce coincidental clusters.[52] Also, MEPFC depends heavily on initialisation and often generates coincidental clusters since the objective function of this algorithm is separable. In addition the ''scale parameter'', β, depends on the data set and usually has to be determined experimentally.[12] Even if FCM is more sensitive to noise than PCM and MEPFC, it does not require parameters such as η and β to run. Furthermore, FCM is less sensitive to initialisation than both PCM and MEPFC. For slightly noisy data, the performance of FCM is not disturbed. However, for high noisy data the performance of FCM is affected.

3.5 Edge Detection

Segmentation can be performed through edge detection of various image regions. Edges are formed at intersection of two regions where there are abrupt changes in grey level intensity values. Edge detection works well on images with good contrast between regions. However, the detection is limited in regions with low contrast. Furthermore, it is difficult to find correlation between the detected edges of the RoIs. Since edges are local features, they are determined based on local information. There are different types of differential operators such as Roberts gradient, Sobel gradient, Prewitt gradient and the Laplacian operator.[17] While the first operators are called first difference operator, Laplacian is a second difference operator. An edge is marked if a significant spatial change occurs in the second derivative. The Laplacian of a 2-D function I(x, y) is defined by $G_r(x, y) = -\nabla^2\{I(x, y)\}$ where the Laplacian operator is defined as $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$.$G_r(x, y)$ exhibits a sign change at the point of inflection of I(x,y). The zero crossing of $G_r(x, y)$ indicates the presence of an edge. The digital second derivative form encountered most frequently in practice is given by:

$$\begin{aligned} \partial^2 I(x, y)/\partial x^2 &= I(x+1, y) + I(x-1, y) - 2I(x, y) \\ \partial^2 I(x, y)/\partial y^2 &= I(x, y+1) + I(x, y-1) - 2I(x, y) \end{aligned} \tag{18}$$

The digital implementation of the 2-D Laplacian is obtained by summing the two components:

$$\nabla^2 I(x, y) = I(x+1, y) + I(x-1, y) + I(x, y+1) + I(x, y-1) - 4I(x, y) \tag{19}$$

The Laplacian generally is not used in its original form for edge detection for several reasons:[17] it is not orientation invariant, and typically is unacceptably sensitive to noise owing to the small size of its template. The magnitude of the Laplacian produces doubles, an undesirable effect because it complicates segmentation. A good edge detector should be a filter with the following two features. First, it should be a differential operator, taking either a first or second partial derivative of the image. Second, it should be capable of being tuned to act at any desired scale, so that large filters can be used to detect blurry shadow edges, and small ones to detect sharply focused details.[16] An operator that satisfies the above conditions is the Laplacian of Gaussian (LoG) in which Gaussian-shaped smoothing is performed prior to application of the Laplacian.[53] Its formulation can be expressed in the continuous domain as follows:

$$G_r(x, y) = -\nabla^2 (I(x, y) \otimes H_\sigma(x, y)) = I(x, y) \otimes \left(-\nabla^2 (H_\sigma(x, y))\right) \tag{20}$$

where $H(x, y) = -\nabla^2 (H_\sigma(x, y)) = \frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right] \exp\left\{\frac{x^2+y^2}{2\sigma^2}\right\}$ is the impulse function of the operator. The position of the edges is present at zero-crossing in $G_r(x, y)$. $H_\sigma(x, y)$ is a Gaussian smoothing filter where σ is the spread of the Gaussian and controls the degree of smoothing. A discrete domain version of the LoG operator can be obtained by sampling the continuous domain of the impulse function $H(x, y)$ over the $W_F \times W_F$ window. In order to avoid deleterious truncation effects, the size of the array should be set such that $W_F = 3b$ or greater, where $b = 2\sqrt{2\sigma}$ is the width of the central excitatory region of the LoG operator. In practice, $G_r(x, y)$ is calculated by multiplying the original image by a template generated by sampling the impulse function $H(x, y)$. For the computation of $G_r(x, y)$, many researchers proposed approximation schemes to speed up the procedure.[53,54] As a result of various sources of error, it has been shown that zero crossing do not always lie at pixel sample points. Sub-pixel accuracy of LoG detector was discussed in ref.[54] To make LoG edge detection robust, zero-crossing can be combined with other measures to judge whether an edge exist or not.

Another popular algorithm is the Canny edge detector.[18] This operator was designed to be an optimal edge detector. Canny's approach is based on optimizing the trade-off between two performance criteria:

- good edge detection: the probabilities of failing to mark real edge points and marking false edge points should be low;

- good edge localization: the positions of edge points marked by the edge detector should be as close as possible to the real edge.

The optimization can be formulated by maximizing a function that is expressed in terms of:

- the signal-to-noise ratio of the image;
- the localization of the edges;
- a probability that the edge detector only produces a single response to each actual edge in an image.

The Canny edge detection algorithm is based on the following steps:

1) smooth the image with a Gaussian filter;
2) compute the gradient magnitude and orientation using finite-difference approximations for the partial derivatives;
3) apply nonmaxima suppression to the gradient magnitude;
4) perform hysteresis thresholding algorithm to detect and link edges.

The Canny edge detector approximates the operator that optimizes the product of signal-to-noise ratio and localization.

3.5.1 Smoothing

Let I(x, y) denote the image. Convolution of I(x, y) with $H_\sigma(x, y)$ gives an array of smoothed data: $S(x, y) = H_\sigma(x, y)I(x, y)$.

3.5.2 Gradient Calculation

Firstly, the gradient of the smoothed array S(x, y) is used to produce the x and y partial derivatives P(x, y) and Q(x, y), respectively:

$$\begin{aligned} P(x, y) &= [S(x, y+1) - S(x, y) + S(x+1, y+1) - S(x+1, y)]/2 \\ Q(x, y) &= [S(x, y) = [S(x, y) - S(x+1, y) + S(x, y+1) - S(x+1, y+1)]/2 \end{aligned} \tag{21}$$

The partial derivatives are computed by averaging the finite differences over the 2×2 square. From the standard formulas for rectangular-to-polar conversion, the magnitude and orientation of the gradient can be computed as:

$$M(x, y) = \sqrt{P^2(x, y) + Q^2(x, y)},\ \theta(x, y) = \tan^{-1}(Q(x, y), P(x, y)) \tag{22}$$

3.5.3 Nonmaxima Suppression (edge thinning)

Once the rate of intensity change at each point in the image is found, edges must be placed at the points of maxima; or rather non-maxima must be suppressed. A local maximum occurs at a peak in the gradient function M(x, y), or alternatively where the derivative of the gradient function is set to zero. However, in this case we wish to suppress non-maxima perpendicular

to the edge direction, rather than parallel to (along) the edge direction, since we expect continuity of edge strength along an extended contour. This assumption creates a problem at corners. Instead of performing an explicit differentiation perpendicular to each edge, another approximation is often used. Each pixel in turn forms the centre of a nine pixel neighbourhood. By interpolation of the surrounding discrete grid values, the gradient magnitudes are calculated at the neighbourhood boundary in both directions perpendicular to the centre pixel. If the pixel under consideration is not greater than these two values (i.e. non-maximum), it is suppressed.

3.5.4 Edge Thresholding

The Canny operator works in a multi-stage process. First of all the image is smoothed by Gaussian convolution Then a simple 2-D first derivative operator (somewhat like the Roberts Cross) is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as *non-maximal suppression*. The tracking process exhibits hysteresis controlled by two thresholds: T_1 and T_2, with $T_1 > T_2$. These two thresholds can be calculated based on noise estimates in the image, but they often are set by hand. Tracking can only begin at a point on a ridge higher than T_1. Tracking then continues in both directions out from that point until the height of the ridge falls below T_2. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments. The effect of the Canny operator is determined by three parameters: σ, T_1, and T_2. Increasing σ reduces the detector's sensitivity to noise at the expense of losing some of the finer details in the image. The error localization in the detected edges also increases slightly as σ is increased.

Figure 5 shows an example of Canny detector applied to blood pool cardiac image. Note that the LV edges are well delineated (Fig. 5f). Edge detection methods have been proposed to extract the edges of the LV in gated cardiac nuclear medicine images,[55,56] and for volume quantitation in SPECT.[28,57,58] Edge detection of gated SPECT images has been performed to estimate LV volume and left myocardial volume in hypertrophic cardiomyopathy patients.[59] Canny detector is used to segment SPECT images in order to analyze abnormal functional asymmetry of the prefrontal boys with attention deficit hyperactivity disorder.[60] Since nuclear medicine images are inherently noisy, the performance of edge detection algorithms is affected. In most cases, these algorithms are not used on their own for segmentation, but coupled with other segmentation algorithms to solve a particular segmentation problem. For example, Dai *et al.*[56] combined the Laplacian operator with mean field annealing method to delineate the LV edges.

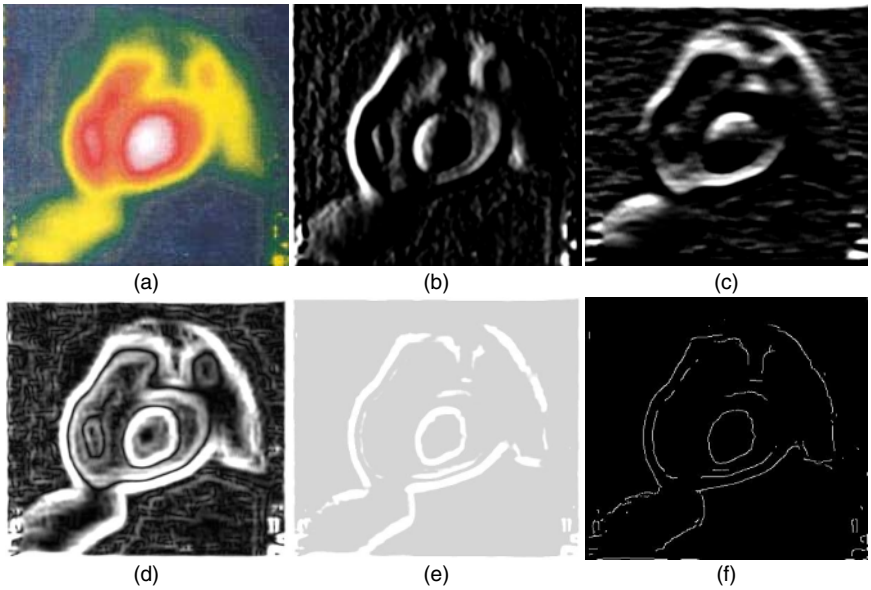

FIGURE 5. Canny edge detector applied to blood pool cardiac image. (a) Initial image. (b) Edge detection along X-axis direction. (c) Edge detection along Y-axis. (d) Norm of the gradient. (e) The norm of the gradient after thresholding. (f) Thinning.

3.6 Markov Random Field Models

Thresholding methods or clustering techniques such as FCM and C-means work well on noise-free images, with slow spatial variation in intensity, and non textured images. These methods typically do not take into account the spatial characteristics of the image, which render them sensitive to noise and thus particularly less reliable. Furthermore, spatial feature is an important piece of information for segmentation. It is advantageous to statistically model the noise and any texture which is random by nature. One may also take advantage of two-dimensional spatial ergodicity to average the effects of noise.[61] If a region is spatially ergodic then a pixel and its neighbours will have similar statistical properties. A natural way to incorporate spatial correlation into a segmentation process is to use Markov random fields (MRFs). MRFs are a class of statistical models that describe contextual constraints. They can be interpreted as a generalization of Markov chain models, which describe (unidirectional) temporal constraints. Assume that an image is defined on $M \times N$ lattice Ω indexed by (i, j) so that $\Omega = \{(i, j); 1 \leq i \leq M, 1 \leq j \leq N\}$. Each element of Ω is called *site*. The sites in Ω are related one to another via a *neighbourhood system*, which is defined as $N_v = \{N_{ij}, (i, j) \in \Omega\}$, where $N_{ij} \subseteq \Omega$ is the set of neighbouring sites (i, j), $(i, j) \in N_{ij}$ and $(i, j) \in N_{i'j'} \Leftrightarrow (i', j') \in N_{ij}$. The sites together with N_v

form an undirected graph, which is used to define the contextual constraints between site labellings. A site often represents a point or a region in the Euclidean space such as an image pixel or an image feature such as a corner point or a line segment. Let $Y = \{Y_{ij} = y_{ij};\, (i,j) \in \Omega\}$ be the observed grey level image where pixels take values from the set $\Lambda = \{1, 2, \cdots, L-1\}$. The observation image itself can be denoted by y. The image segmentation problem involves assigning to each pixel a class label taking a value from the set $B = \{1, 2, \cdots, C\}$ where C is the number of classes. Let $X = \{X_{ij};\, (i,j) \in \Omega\}$ be a family of random variables defined on the set Ω, in which each random variable X_{ij}, associated to the site $(i,j) \in \Omega$, takes a label value x_{ij} in B. Thus a label is an event that may happen to a site. The family X is called random field (or labels field). The notation $X_{ij} = x_{ij}$ denotes that the event takes the value x_{ij} and the notation $(X_{11} = x_{11}, \cdots, X_{MN} = x_{MN})$ denotes the joint event and is also a labelling of the sites in Ω in terms of the labels in B. A *configuration* of X, corresponding to a realization of the field, is noted by $x = \{x_{11}, x_{12}, \cdots, x_{MN}\}$. The probability that a random variable X_{ij} takes the value x_{ij} is denoted $P(X_{ij} = x_{ij})$. A random field X is said to be a MRF on Ω with respect to N_{ij} if

$$P(x_{ij}) > 0 \;\forall\; (i,j) \in \Omega \text{ and } P\left(x_{ij} | X_{\Omega - \{(i,j)\}}\right) = P\left(x_{ij} | X_{N_{ij}}\right) \tag{23}$$

where $X_{\Omega-\{(i,j)\}} = \{x_{kl};\, k = 1, 2, \cdots, i-1, i+1, \cdots, M;\, l = 1, 2, \cdots, j-1, j+1, \cdots, N\}$. $\Omega - \{(i,j)\}$ denotes the set of all sites excluding (i,j). $X_{\Omega-\{(i,j)\}}$ and $X_{N_{ij}}$ denote the random variables associated with the sites $\Omega - \{(i,j)\}$ and N_{ij}, respectively. A first order neighbourhood consists of the four nearest pixel sites; a second order neighbourhood consists of the eight nearest pixel sites, and so on. A clique is a set of one or more sites such that each site in the clique is a neighbourhood of all other sites in the clique. The segmentation problem can now be simply as follows. Given the corrupted data y, find an estimate $\tilde{x}$ of the true labelling configuration x^* (observed under ideal conditions). This inverse problem is ill-posed. Prior information is often very useful in formulating a regularized process so that optimal results can be obtained by solving the regularized problem.[62] A well known approach to this problem is the Bayesian estimation that incorporates prior information through an *a priori* distribution of the random field X, P (X). The Bayesian estimation can be formulated as a maximum *a posteriori* (MAP) estimation that maximizes the posterior probability or likelihood of $\tilde{x}$ given y. In particular if P ( ) is an appropriate probability measure, then one would like to find the estimate $\tilde{x}$ which maximizes $P(X = \tilde{x}|Y = y)$. The Bayes theorem is given by the following relation: $P(X = \tilde{x}|\, Y = y) = P(Y = y|X = \tilde{x}).P(X = \tilde{x})/P(Y = y)$ where $P(Y = y|X = \tilde{x})$ is the conditional probability of Y given the field X. Since $P(Y = y)$ is independent of $\tilde{x}$, we can maximize the relation $P(X = \tilde{x}|Y = y) \propto P(Y = y|X = \tilde{x}).\, P(X = \tilde{x})$ or $\log[(X = \tilde{x}|Y = y)] \propto \log[(Y = y|X = \tilde{x})] + \log[(X = \tilde{x})] + \log[(X = \tilde{x})]$. It

should be pointed out that the difficulty in maximizing this equation lies in the fact that it is a joint log-likelihood for all the image data. It does not simply describe the likelihood of a single pixel. In particular, the maximizing $\tilde{x}$ is one of $C^{M\times N}$ possibilities. The solution is given by

$$\hat{x} = \arg\max_{x\in X} \{\log P[(y|\tilde{x})] + \log[P(\tilde{x})]\} \tag{24}$$

The first term of Eq. 24 is the likelihood, telling how the data are obtained from the labels while the second is the prior probability of a given labelling. In practice, the first term forces fidelity to the data while the second penalizes unlikely trough labelling. Note that, without the *priori* distribution P(x), the estimation scheme becomes a maximum likelihood estimation (MLE). MLE is often used as the initial estimate in the iterative MAP estimation. For practical use, a means to specify the conditional probabilities is required. From the above equation, computation of the prior probability of the class and the likelihood probability of the observation is needed. Since X is assumed as a realization of a MRF, according to the Hammerseley-Clifford theorem,[63] the probability density of X is given by a Gibbs distribution having the form:[62,64]

$$P(X) = Z^{-1}\exp(-U(X)/T) = Z^{-1}\exp\left(-\sum_{c\in C_L} V_c(X)/T\right) \tag{25}$$

where Z is a normalizing constant, called the partition coefficient. T stands for "temperature" and controls the degree of peaking in the probability density, i.e. the larger the value, the larger is the peaking. U ( ) in Eq. 25 is an energy function, composed of potentials of the cliques $V_c(\ )$, which specifies the degree of penalty imposed on the neighbours. A clique potential specifies the admissibility of a specific labelling of the nodes of a clique. The value of $V_c(\ )$ depends on the local configuration of clique c. The set of cliques C_L comprises the completely connected subgraphs of the MRF graph. An example of potential functions is given by:

$$V_c(X) = \begin{cases} \beta & \text{if } x_{ij} = x_{i'j'}\ (i,j),\ (i',j') \in C_L \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

Assuming that the noise in the image, is additive, Gaussian, independent, and with zero mean and variance σ^2, $y_i = x_i + \eta$, the conditional density probability of $(y_i - x_i)$ is simply the probability of the noise

$$P(y_{ij}|x_{ij}) = \left(1/\sigma\sqrt{2\pi}\right)\exp\left(-(y_{ij} - x_{ij})^2/\sigma^2\right) \tag{27}$$

Assuming the noise is independent, we find for the entire image $P(y|x) = \Pi_{(i,j)\in\Omega} P(y_{ij}|x_{ij})$. Thus,

$$P(y|x) = \left(\Pi_{(i,j)\in\Omega}\sigma\sqrt{2\pi}\right)^{-1}\exp(-U(y|x)) \tag{28}$$

where $U(y|x) = \Sigma_{(i,j)\in\Omega}(y_i - x_i)^2/(2\sigma^2)$ is the likelihood energy. It is easy to show that $\log[P(x|y)] \propto -U(x|y)$ where $U(x|y) = U(y|x) + U(x) + \text{Const}$ (where Const is a constant) is the posterior energy. Finally, the MAP estimation is equivalent to minimizing the posterior energy function:

$$\hat{x} = \arg\min_{x\in X} \{U(y|x) + U(x)\} \tag{29}$$

Although mathematically simple, finding a global minimum for MAP estimation, given all the configurations x is a difficult task. Therefore, optimal solutions are usually computed using iterative optimization methods. A well known method is the iterated conditional modes (ICM) algorithm of Besag,[63] which uses the "greedy" strategy in the iterative local minimization and convergence is guaranteed after only a few iterations. Given the image y and labels $x^k_{\Omega-\{(i,j)\}}$ ICM sequentially updates each $x^{(k)}_{ij}$ into $x^{(k+1)}_{ij}$ by minimizing $U\left(x_{ij}|y, x_{\Omega-\{(i,j)\}}\right)$, the conditional posterior probability with respect to x_{ij} (k is the iteration number).

The MRF model has been applied to segmentation of both PET and SPECT images as a pre-process for classification and restoration.[65-69] For example, MRF model was used to segment brain tissues, such as striatum, GM and WM, in dynamic FDOPA-PET studies. The goal is to classify the tissues according to their physiological functions.[69] Markovian segmentation of 3D brain SPECT images was also performed to find the support of the objects to be restored using a 3D blind deconvolution technique.[67] The aim of this restoration is to improve the spatial and inter-slice of SPECT volumes for easy clinical interpretation. In another study, 3D SPECT brain images were segmented into three classes (CSF, WM, GM) using 3D MRF.[68] Feature vectors are extracted from the segmented SPECT volumes to classify brains into two classes, namely "healthy brains" and "diseased brains" (i.e., brains with possible cerebrovascular disease).

3.7 Artificial Neural Networks

Classical segmentation methods such as edge detection or deformable models discussed in section 3.8 often require considerable user expertise. For any artificial vision application, one desires to achieve robustness of the system with respect to random noise and failure of process. Moreover, a system can probably be made artificially intelligent if it is able to emulate some aspects of the human information processing.[16] Another important requirement is to have the output in real time. Artificial Neural Networks (ANNs) are attempts to achieve these goals.

ANNs are massively connected and parallel networks of elementary processors.[70-81] Each processor simulating biological learning is capable of performing elementary computation. Learning is achieved through the adaptation of weights (synaptic coefficients) assigned to the connections between processors. The massive connectionist architecture usually makes

the system robust while the parallel processing enables the system to produce output in real time.[16] ANNs are widely used in segmentation as a classifier. The weights are determined using training data, and the ANN is then used to segment new data. ANNs can also be used in an unsupervised way as a clustering method. Different neural models have been developed for image segmentation, particularly Hopfield, Kohonen, radial basis functions and Pulse-Coupled Neural Networks (PCNNs), and MultiLayer Perceptrons (MLPs).[75-77,79-81] These methods work well in a noisy environment and consequently are interesting for processing nuclear medicine images, which are inherently noisy. For example, Keller and Mckinnon[79] used PCNNs for segmentation of nuclear ventilation/perfusion images of the lungs. A PCNN is physiologically motivated information processing model based on the mammalian visual cortex.[72] In this model, each neuron in the processing layer is directly tied to an image pixel or set of neighbouring input image pixels. Each neuron iteratively processes signals feeding from these nearby image pixels (i.e., feeding inputs) and linking from nearby neurons (i.e. linking inputs) to produce a pulse train.[79] There is no training involved for the standard PCNN. Ghosh *et al.*[77] presented an original neuro-fuzzy approach to extract objects from noisy images. This method takes into account the neighbourhood of the pixels and ensures the compactness of the extracted objects by extracting spatially compact regions through the process of self-organization using only one noise corrupted image. A MLP-type of neural network is used and trained using the well-known back-propagation (BP) algorithm. The advantage of this approach is that there is no supervised learning. The error of the proposed system is computed using concepts of fuzzy sets. Based on this approach, Behloul *et al.*[7] proposed a neural system to myocardium extraction in PET images. The noisy background of these images makes myocardium extraction and tracer uptake quantification a very difficult task.

Myocardium extraction is required for a good visualization and interpretation. Indeed, analysis of ^{18}F-FDG PET heart images performed by simple visual evaluation slice by slice is tedious and suffers from observer bias. As an alternative to this visual evaluation, information can be represented using a polar-map technique. This polar map can be thought of as the image that would be obtained if one looks at the LV as a 3D cone shaped object and projected onto a single plane. This form of display has the advantage that all areas of myocardium are represented in a single image, but it is less easy to assess the relationship between the defected region and coronary arterial territories.[80] Both slice by slice visualization and polar map do not allow an easy assessment of the extent of tracer uptake defect on the LV.[7] Thus, a 3D visualization method presenting accurate localization and extent of the disease is desirable.

Cardiac ^{18}F-FDG PET images are generally involved in viable myocardium assessment studies. The quantification of the viable tissue for a patient presenting a CAD can help to decide whether or not the patient will benefit

from a revascularisation procedure. In general, 3 levels of ^{18}F-FDG uptake are distinguished by the clinicians: low, intermediate and high. Normal or remote myocardium is expected to have a high uptake rate that is more than 70% of the maximal value in the image (assuming that the maximum value corresponds to a normal tissue). Seriously damaged or infracted tissues have a low uptake rate (less than 50%).[80] However, it is very difficult to diagnose ischemic processes involved in the studied dysfunctional myocardium based only on ^{18}F-FDG PET images and especially when ^{18}F-FDG uptake is intermediate (50 to 70 %). A medium (intermediate) uptake region may correspond to a partially infracted tissue (heterogeneous) or a homogeneous reduced metabolism function. Because of the doubt, some clinicians have rather classified these regions into the viable category. Ideally, a threshold value equal to 50% (or in the range of 50 to 60%) of the max value should allow the visualization of viable myocardium.[80] However, due to the noisy nature of PET images, simple thresholding techniques, generally provided in commercial software supplied to end-users, does not seem to be suitable for a 3D extraction and visualization of the viable myocardium.

Indeed, in some images, some pathological regions were not visible and thus a higher threshold value was needed. However, an increased threshold value tends to overestimate the extent of already detected defects and thins down the myocardium. There is obviously a need for more sophisticated methods to extract accurately viable myocardium from the background. To this end, Behloul *et al.*[7] proposed a Radial Basis Function Network (RBFN) to extract myocardium through a self-organization process using only one PET image. The error of the RBFN architecture is computed using the concept of fuzzy sets and fuzziness measures. RBFNs are mainly much easier to train than MLP since they establish the Radial Basis Function (RBF) parameters directly from the data, and training is primarily on the output layer.[80] The RBFN is a feed forward neural network which accomplishes an input-output non-linear mapping by linear combination of nonlinearly transformed inputs according to:

$$o_j = \sum_{i=1}^{M_f} w_{ij}\varphi_i(x) \tag{30}$$

where x is the input vector, o_j the output of the j^{th} output node and w_{ij} are the output linear combining weights. The $\varphi_i(x)$ are RBFs and M_f is the number of RBFs. Each component of the input vector x feeds forward to M_f RBF node whose output are linearly combined with weights into the network output node. An output of the network is a simple linear combination of the hidden neuron outputs. The most important feature of RBFs is that they are local i.e. they give a significant response only in a neighbourhood near a central point. Their response decreases monotonically with distance from a central point. RBF parameters are its center, shape, and width.

A typical local RBF is a Gaussian function centred at c_x and of width (or radius) r, having the form:

$$G(x, c_x, r) = \exp\left(\frac{\|x - c_x\|_R^2}{2r^2}\right) \tag{31}$$

where R is a positive definite matrix. RBFNs have traditionally been associated with radial function networks in a single hidden layer. The hidden and output layers are generally trained sequentially: the RBF parameters are first fixed and the optimal linear combining weights are then computed.[82,83] Once the number and parameters of the RBFs defined, the hidden layer performs a fixed nonlinear transformation; it maps the input space into a new space. The output layer then implements a linear combiner on this new space. The only output layer parameters to adjust are the weights of this linear combiner. In general, the output weights of a RBFN can be determined by a pseudo-inverse matrix. When applied to supervised learning with linear models, the Least Square (LS) principle leads to a particularly easy optimization problem. However, this approach can be computationally demanding when the training set is large. In this case, delta-rule type of BP is preferred since it is a less demanding technique. Behloul *et al.*[7] used the BP algorithm to train the output layer. The procedure for learning the correct set of weights consists in varying the weights so that the error (E) is reduced. In supervised training, a set of input-output couples is supplied. The error of the network is generally computed by:

$$E = 0.5 \times \sum_{j=1}^{N} \left(t_j - o_j\right)^2 \tag{32}$$

where o_j is the spontaneous output of node j and t_j is the expected output of node j, N is the total number of output units. The network is supposed to extract the heart from the noisy image based only on the contrast between myocardium and background. As human eyes would do, the network will have to focus on the object in the noisy image until it becomes "clear". Thus, a measure of fuzziness of the image can be considered as the error of the network, to be reduced.[81]

A fuzzy set F is characterized by a membership function $\mu_F(x)$ that associates to each point x in a space of points X a real number in the interval [0, 1]. The value $\mu_F(x)$ represents the grade of membership of x in F. $\mu_F(x) = 1$ indicates a strict containment of x in F and $\mu_F(x) = 0$ means that x does not belong to F. Any intermediate value would indicate the degree to which x is an element of F. Formally, a fuzzy set F that has a finite number of elements $x_1, x_2, \cdots, x_n$ is defined as a collection of pairs

$$F = \{(\mu_F(x_i), x_i), i = 1, 2, \cdots, n\} \tag{33}$$

A measure of fuzziness estimates the average ambiguity in a fuzzy set. Intuitively, the fuzziness of a crisp set using any measure should be zero (or minimum), as there is no ambiguity about whether an element belongs to the set or not. If the set is maximally ambiguous, then the fuzziness measure should be maximum. When the membership value approaches either 0 or 1, the ambiguity of the set decreases. Thus, a fuzzy set is most ambiguous when $\mu_F(x) = 0.5 \forall x$. The degree of fuzziness, I(F), of a fuzzy set F represents the amount of ambiguity in making the decision whether a point belongs to F or not. Such measures have been proposed by several authors.[84–88] Two mathematical models of fuzziness measures are used: the index of fuzziness and the fuzzy entropy.[77] Both measures presented below lie in the interval [0, 1]. The index of fuzziness of F is defined by $v(F) = \frac{2}{n^k} d(F, \underline{F})$, where $\underline{F}$ is the nearest ordinary set to F given by

$$\mu_{\underline{F}}(x) = \begin{cases} 0 & \text{if } \mu_F(x) \leq 0.5 \\ 1 & \text{if } \mu_F(x) > 0.5 \end{cases} \tag{34}$$

$d(F, \underline{F})$ is the distance between the two sets F and $\underline{F}$. The value of k depends on the type of the distance used. $k = 1$ corresponds to the generalized hamming distance and $k = 0.5$ to the Euclidean distance. The corresponding index of fuzziness is called the quadratic index of fuzziness $v_q^2(F)$.

$$v_q(F) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{\left(\mu_F(x_i) - \mu_{\underline{F}}(x_i)\right)^2} \tag{35}$$

The index of fuzziness reflects the ambiguity of an image by measuring the distance between its fuzzy property plane and the nearest ordinary plane. Pal and Pal[86] proposed an exponential entropy given by

$$H_z(F) = \frac{1}{n(\sqrt{e} - 1)} \sum_{i=1}^{n} \left(S_n(\mu_F(x_i)) - 1\right) \tag{36}$$

with

$$S_n(\mu_F(x_i)) = \mu_F(x_i) e^{1-\mu_F(x_i)} + (1 - \mu_F(x_i)) \mu_F(x_i) e^{\mu_F(x_i)}$$

$H_z(F)$ gives a measure of the average amount of difficulty in taking a decision on any pixel. Both terms (index and entropy) give an idea of "indefiniteness" of an image. They may be considered as measures of average intrinsic information that is received when one has to make a decision in order to classify the pixels described by a fuzzy set. Thus, these measures of fuzziness are used to compute the error of the SRBFN.

The architecture of SRBFN, derived form the approach of Ghosh *et al.*[77], presents three layers: the input layer, one hidden (RBF) layer and the output layer (Fig. 6). In every layer, there are $L_x \times L_y$ neurons (for an $L_x \times L_y$ image) each neuron corresponds to a single pixel. Neurons in the same layer are not connected to each other (Fig. 6b). Each neuron is connected to the corresponding neuron and its neighbours of the next layer. The

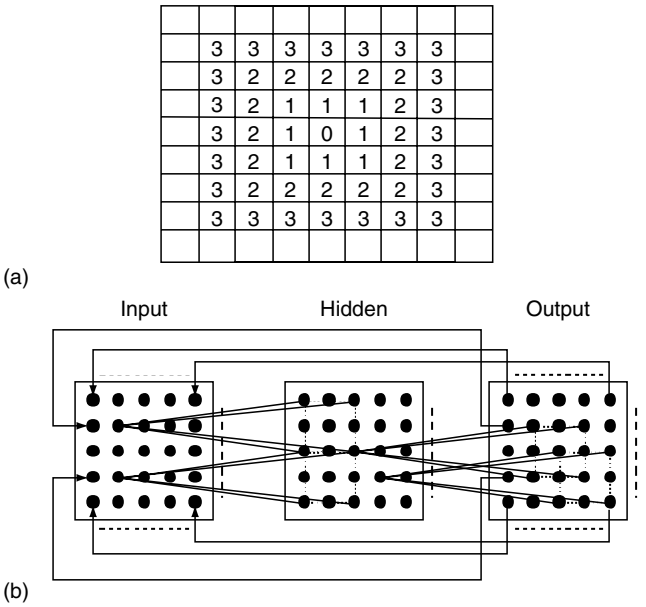

FIGURE 6. Illustration of the architecture of self organized RBFN. (a) Neighbourhood system. (b) RBFN topology. (Reprinted with permission from ref.[7]).

neighbourhood considered here corresponds to the $8(3 \times 3 - 1)$ nearest neighbours (N) or the $24(5 \times 5 - 1)$ nearest neighbours $\left(N^2\right)$ and so on (Fig. 6a).

In the SRBFN, the layers are partially connected and the neurons have spatial positions corresponding to the pixels in the image. The feedback connections from the output layer to the inputs make the network recurrent. The output is considered as input for the next iteration. Behloul[80] proposed a sigmoid function f for the activation of the output nodes $o_j = f\left(I_j\right)$ where $I_j = \sum_{i=1}^{M_f} w_{ij}\varphi_i(x)$ and

$$f(x) = 1/\left(1 + e^{-(x-\theta)}\right) \tag{37}$$

where θ is a bias value. The output of each neuron of the hidden and output layers lies in [0, 1]. It represents the degree of brightness of the corresponding pixel in the image. The measure of fuzziness of the fuzzy set bright is viewed as the error of the network. Once the weights have been adjusted according to the error measure, the neurons in the output layer are feed back to the corresponding neurons in the input layer. The output values of the output nodes which lie in [0, 1] are considered to be the brightness degrees of the corresponding pixels (0 for black, 1 for white, and any intermediate value is a grey level). The grey levels corresponding to the output values are considered as the input of the next iteration. The BP algorithm is used to train the output layer. The change in weights, such that the error E is reduced, is given

by $\Delta w_{ij} = \eta(-\partial E/\partial o_j) f'(I_j) o_i$ where η is a learning rate. The sigmoid function is used because of its simple derivative function: $f'(I_j) = \partial o_j/\partial I_j = o_j(1 - o_j)$. The mathematical derivations of the BP for quadratic index of fuzziness v_q^2 and fuzzy entropy H_z are defined as follows:

- *Quadratic index of fuzziness*
The error is defined by $E = v_q^2$ where

$$v_q^2 = (4/n)\left[\sum_j \{\min(o_j, (1 - o_j))\}^2\right] \tag{38}$$

and n is the number of output units. Thus,

$$-\partial E/\partial o_j = \begin{cases} -8o_j/n & \text{if } 0 \le oj \le 0.5 \\ 8(1 - o_j)/n & \text{if } 0.5 < oj \le 1 \end{cases} \tag{39}$$

$$\Delta w_{ij} = \begin{cases} \eta_1(-o_j) f'(I_j) o_j & \text{if } o \le oj \le 0.5 \\ \eta_1(1 - o_j) f'(I_j) o_j & \text{if } 0.5 < oj \le 1 \end{cases} \tag{40}$$

where $\eta = \eta_1 \times (4/n) \times 2$

- *Fuzzy entropy*
The error is defined by $E = H_z$ where

$$H_z = \frac{1}{n(\sqrt{e} - 1)} \sum_{j=1}^{n} \left(o_j e^{1-o_j} + (1 - o_j) e^{o_j}\right) \tag{41}$$

$$\frac{\partial H_z}{\partial o_j} = \frac{1}{n(\sqrt{e} - 1)} \left((1 - o_j) e^{1-o_j} - o_j e^{o_j}\right) \tag{42}$$

To make weight correction value minimum when the membership values of the elements are 0 or 1 and maximum when they are all 0.5, we take

$$\Delta w_{ij} = \eta_2 \frac{1}{\frac{\partial E}{\partial o_j}} f'(I_j) o_i \tag{43}$$

$$\Delta w_{ij} = -\eta \frac{1}{(1 - o_j) e^{1-o_j} - o_j e^{o_j}} f'(I_j) o_i \tag{44}$$

where $\eta = \eta_2 \times n(\sqrt{e} - 1)$

$$\Delta w_{ij} = \begin{cases} -\eta \frac{1}{(1-o_j) e^{1-o_j} - o_j e^{o_j}} f'(I_j) o_i & \text{if } 0 \le oj \le 0.5 \\ \eta \frac{1}{o_j e^{o_j} - (1-o_j) e^{1-o_j}} f'(I_j) o_i & \text{if } 0.5 \le oj \le 1 \end{cases} \tag{45}$$

At the first iteration, the input nodes receive the grey levels of the corresponding pixel. For the hidden layer, each RBF centre is localized at the

value of the corresponding pixel in the image. The total input to any node of the output layer lies in $[0, N_1]$ where N_1 is the number of links that a neuron has. The parameter θ of the sigmoids in the output nodes is set to $N_1/2$, since this value corresponds to the middle of the total input range.[7] All initial weights (between the hidden and output layers) are set to 1. The SRBFN presented by Behloul[80] aimed to extract a compact object from one noisy 2D image. It is easy to extend the architecture to process 3D images; however the computation complexity grows dramatically. It turned out that the result obtained when considering the 3D image as a set of 2D slices and applying the SRBFN to each slice was comparable (in terms of size and shape of the extracted object) to that when considering the total 3D image. The 2D images are processed independently in a sequential manner. Behloul considered the 3D PET volume as a set of 2D-images where two orthogonal views are presented: 63 (128×128)-slices and 128 (63×128)-slices; the intersection of the two independently extracted objects constitutes the final result of the myocardium extraction process.

Different sizes of neighbourhood have been tested and N^2 turned out to be the best compromise between processing time and accuracy of shape and width of the extracted myocardium. As pointed out by Ghosh *et al.*[77] and Behloul *et al.*[7], the index of fuzziness is better than the entropy measure for maintaining the compactness of the extracted objects. However, the shapes of objects are better preserved by the entropy measure. For the index of fuzziness, the learning rate is lower than that of the entropy measure. Low learning rate smoothes out noise and creates compact regions, while entropy measure enables the network to preserve object boundaries as learning rate is very high around the most ambiguous region ($o_j \approx 0.5$).[77] Figure 7 shows the results obtained using the index of fuzziness versus those obtained using the entropy measure. The diseased area appears as a hole in the myocardium (Fig. 8).

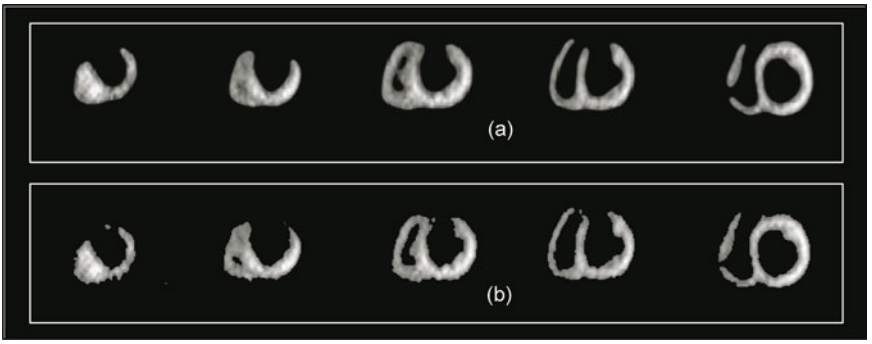

FIGURE 7. Object extraction results. (a) Index of fuzziness. (b) Fuzzy entropy. (Reprinted with permission from ref.[7]).

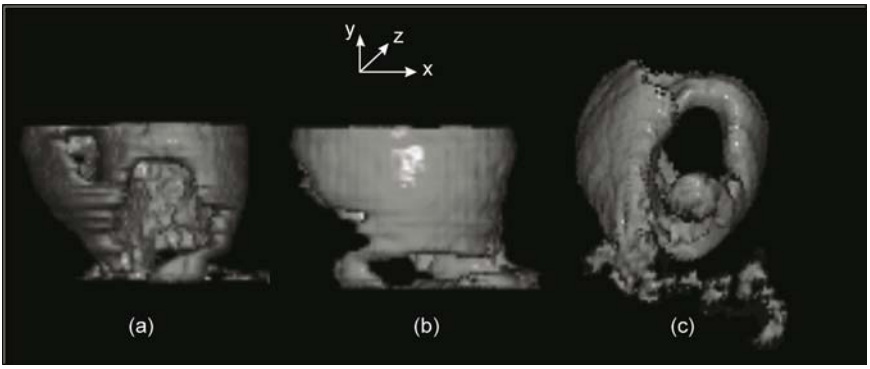

FIGURE 8. Volume rendered visualization of the entropy extracted myocardium: the diseased region appears as a hole in the myocardium. (a) Orthogonal view of the heart showing a defect in almost all the posterior wall of the LV. (b) 90° y-axis rotation of (a) showing the extent of the disease in the apical anterior wall of the LV. (c) 45° x-axis rotation of (a) showing both defects (posterior and apical anterior) and the RV. (Reprinted with permission from ref.[7]).

For the index of fuzziness (Fig. 7a), the diseased region and the myocardium width are overestimated (over-smoothed), while for entropy (Fig. 7b), the results are very much closer to manually delineated regions. Moreover, using the index of fuzziness, the network requires more time to stabilize. To take advantage of the main positive features of both measures in the same time: "smoothing and compactness" for index of fuzziness, and "good shape estimation" for the fuzzy entropy, both measures were combined. For the first iterations of the self-organizing process (2 to 3 iterations), index of fuzziness is used. Because of its low learning rate, this measure produces a contrast enhancement on the original images (Fig. 9a). The entropy measure is then used; it ensures a good shape extraction (Fig. 9b).

The sequential application of the two fuzziness measures presented the best results. Indeed, when using separately the fuzziness measures, the network was not able to detect some regions of the myocardium where there is a low ^{18}F-FDG uptake (the human eye was still able to allow delineation of such regions). Only the combination of both measures made the network successful in retaining these regions in the extracted myocardium (see arrows in Fig. 9b). Thus, it seems that the two error models complement each other perfectly for an accurate delineation of the myocardium in PET images.

3.8 Deformable Models

Deformable models are curves, surfaces or solids defined with an image or volume domain. These models deform under the influence of external and internal forces. The original deformable model, called *snake* model, was

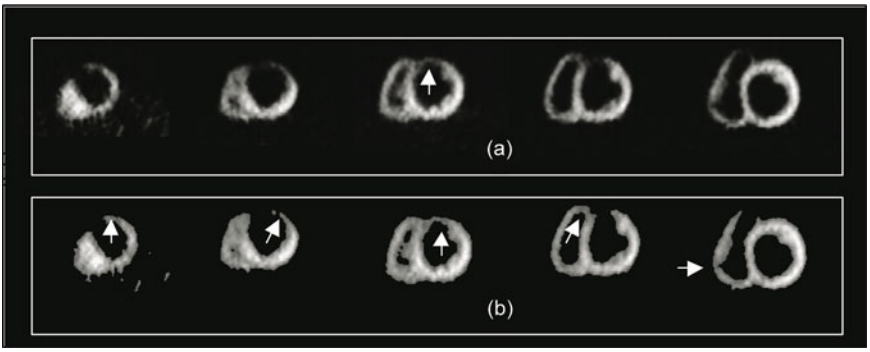

FIGURE 9. Combination of both fuzzy measures. (a) Images obtained after 2 iterations using only the index of fuzziness measure. (b) Final result using the entropy measure on the images shown in (a). (Reprinted with permission from ref.[7]).

introduced by Kass *et al.*[89] as an active spline reacting with image features. Basically, the snake model is an energy-minimizing spline controlled by external image forces such as lines and edges, and internal spline forces which impose a piecewise smoothness constraint. In practice, the user initializes the deformable model near the RoI and allows it to deform into place. User could manually fine-tune the filling by using interactive capabilities of the model. The snake deforms itself into conformity with the nearest salient contour by minimizing the snake energy, which pushes the snake toward salient image features and internal forces, which are responsible for smoothness of the curve. As the algorithm advances, the terms of snake functional energy can be adjusted to obtain a local minimum. The main advantages of deformable models are their ability to directly generate close parametric curves or surfaces from the image and their incorporation of a smoothness constraint that provides robustness to noise and spurious edges. A disadvantage is that they require manual interaction to place an initial model and choose appropriate parameters.[3]

The original formulation of the snake is for a continuous spline and image. The snake geometry is represented by a curve V (s, t) with parameters s (spatial index) and t (time index) defined on given open intervals Ω and T_m, respectively. By permitting the snake to have two deformational degrees of freedom in the plane, that is, the x and y coordinates, the active model, embedded in the plane image $(x, y) \in \mathcal{R}^2$, is represented as a time varying parametric contour by

$$\begin{aligned} V(s, t) = (x(s, t), y(s, t)) : s \in \Omega,\ t \in T_m \\ V : \Omega = [0, 1] \rightarrow \Re^2 \end{aligned} \tag{46}$$

Thus, at time $t = t_0$ the snake or the closed contour C is represented by its pair of coordinate functions $\{x(s, t_0), y(s, t_0)\}$, where $s \in [0, 1]$ is the

normalized arc length along C. Let A_d be a space of admissible deformations. The shape of the contour subject to an image I(x, y) can be dictated by its energy function E_{snake} as follows

$$\begin{aligned} &E_{snake} : A_d \rightarrow \Re \\ &E_{snake} = \int_{\Omega} \left[E_{int}(V(s,\, t)) + E_{image}(V(s,t)) + E_{con}(V(s,\, t)) \right] ds \end{aligned} \tag{47}$$

where E_{int}, E_{image}, and E_{con} refer to the internal, image and external constraint forces, respectively. The energy terms are defined cleverly in a way such that the final position of the contour will have a minimum energy (E_{min}).E_{int} represents the internal energy term of the snake which imposes the regularity on the curve by bending and stretching, and is given by

$$E_{int} = \frac{1}{2} \left\{ \alpha(s) \underbrace{\left\| \frac{\partial [V(s,\, t)]}{\partial s} \right\|^2}_{E_1} + \beta(s) \underbrace{\left\| \frac{\partial^2 [V(s,\, t)]}{\partial s^2} \right\|^2}_{E_2} \right\} \tag{48}$$

E_1 and E_2 represent the continuity and smoothness (curvature) terms, respectively. The contour is treated as an elastic rubber band possessing elastic energy. It discourages stretching by introducing a tension:

$$E_{elastic} = \frac{1}{2} \int_{\Omega} \alpha(s) \left\| \frac{\partial [V(s,\, t)]}{\partial s} \right\|^2 ds \tag{49}$$

$E_{elastic}$ has larger values where there is a gap in the curve. The weight $\alpha(s)$ allows to control elastic energy along different parts of the contour. The $\alpha(s)$ value determine the extent to witch the contour is allowed to stretch. If $\alpha(s) = 0$, a discontinuity can occur. For most applications, $\alpha(s)$ is assumed to be constant throughout the curve. In addition to being considered as an elastic band, the snake is also considered to behave like a thin metal strip giving rise to bending energy. This second term discourages bending. This energy is defined as sum of squared curvature of the contour:

$$E_{bending} = \frac{1}{2} \int_{\Omega} \beta(s) \left\| \frac{\partial^2 [V(s,\, t)]}{\partial s^2} \right\|^2 ds \tag{50}$$

Sharp corners or points of high curvature are characterized as high frequencies and bending energy is more sensitive for contours having such corners because the second derivative will be very high for such contours. $E_{elastic}$ has larger values when the curve is bending rapidly. The bending energy is minimum for a circle. The weight $\beta(s)$ value determine the extent to which the contour is allowed to bend at point s. Setting $\beta(s) = 0$ at point s means that we are relaxing our condition and allowing that *s* develops a corner. The total internal energy of the snake is the sum of elastic and

bending energies. Adjusting the weights $\alpha(s)$ and $\beta(s)$ controls the relative importance of the elastic and bending terms and therefore the internal energy of the contours.

E_{image} is derived from the image data over which the snake lies (object of interest). The three important features the snake can be attracted to are line, edge and termination functions. The total image energy can be expressed as a weighted combination of the three:

$$E_{image} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \tag{51}$$

The simplest useful image functional is the image intensity $E_{line} = I(x, y)$. Depending on w_{line} the snake is attracted to dark or light lines. The edge-based functional, E_{edge}, attracts the snake to contours with image gradients, that is, to locations of string edges. A way to define E_{edge} is as follows:

$$\begin{aligned} E_{edge}(V(s)) &= -\gamma(s)E_3(V(s)) \\ E_3(V(s)) &= \|\nabla I(V(s))\|^2 \\ E_3(V(s)) &= \|\nabla(G_\sigma(V(s)) \otimes I(V(s)))\|^2 \end{aligned} \tag{52}$$

$\gamma(s)$ is a parameter that controls the weight of image attraction. $\nabla I(V(s))$ denotes the gradient of the image I and $G_\sigma(V(s))$ is a 2D Gaussian function with standard derivation σ (the image is pre-processed with a low pass filter). It can be seen that larger σ will increase the blurring of the edges thereby increasing the capture range of the active contour. The termination functional, E_{term}, can be obtained by a function checking the curvature of level lines in a slightly smoothed image. Let $C(x, y) = G_\sigma(x, y)^* I(x, y)$ be the smoothed version of I(x,y) and $\theta = \tan^{-1}\left(\frac{C_y}{C_x}\right)$ the gradient angle. Curvature of contour in $C(x, y)$ can be written as:

$$\begin{aligned} E_{term} &= \frac{\partial\theta}{\partial R_\perp} \\ E_{term} &= \frac{\partial^2 C/\partial R_\perp^2}{\partial C/\partial R} \\ E_{term} &= \frac{C_{yy}C_x^2 - 2C_{xy}C_xC_y + C_{xx}C_y^2}{\left(C_x^2 + C_y^2\right)^{3/2}} \\ E_{term} &= \frac{\left(\partial^2 C/\partial y^2\right)(\partial C/\partial x)^2 - 2\left(\partial^2 C/\partial x\partial y\right)(\partial C/\partial x)(\partial C/\partial y) + \left(\partial^2 C/\partial x^2\right)(\partial C/\partial y)^2}{\left((\partial C/\partial x)^2 + (\partial C/\partial y)^2\right)^{3/2}} \end{aligned} \tag{53}$$

where $R = (\cos\theta,\ \sin\theta)$ is a unit vector in direction of the gradient and $R_\perp = (-\sin\theta,\ \cos\theta)$ is a vector perpendicular to gradient direction.

E_{con} represents the external constraint energy term that is responsible for guiding the snake towards the desired local minimum. It can be defined by the user, constructed from physical phenomena,[90] or derived form high-level interpretation.[89] For example, the user connects a spring between any point

p_1 in the spatial domain and a point p_2 on the snake. A snake trying to reduce its constraint energy is pulled towards the point p_1 defined by the user. Constraint energy can also be interpreted as a kind of prior for the system.

A contour is defined to lie in the position in which the snake reaches a local energy minimum. Before the minimization process, the snake formulation is discretized both in time and space domains. The spatial discretization is done by sampling the active contour V into N points $V(v_i, i = 1, 2, \ldots, N)$. For an initial contour (i.e., at $t = 0$), the N points are chosen to be evenly spaced on the contour. The time discretization is defined to start at time $t = 0$ with a constant time step Δt. The discretization of E_{int} is done by approximating the first and second derivatives by finite differences as follows:

$$\begin{aligned} E_1(i) &= \|v_i - v_{i-1}\|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \\ E_2(i) &= \|v_{i-1} - 2v_i - v_{i+1}\|^2 = (x_{i-1} - 2x_i - x_{i+1})^2 + (y_{i-1} - 2y_i - y_{i+1})^2 \end{aligned} \tag{54}$$

where x_i and y_i are the x and y coordinates of the v_i^{th} contour-point, respectively. Equation 54 is formulated to minimize the distance between neighbouring.[91] For simplicity we suppose that $w_{line} = w_{term} = 0$. Thus, the boundary is detected by iteratively deforming an initial approximation of the boundary by minimizing the reduced snake energy function:

$$E_{snake} = \sum_{i=1}^{N} \{\alpha(i)E_1(i) + \beta(i)E_2(i) + \gamma(i)E_3(i)\} \tag{55}$$

$$E_{min} = \min\{E_{snake}\} \tag{56}$$

Equation 56 was originally solved by using techniques of variational calculus and applying a finite difference method.[89] It was improved later on by Cohen and Cohen[92] using a finite element method. Another solution was suggested by Amini *et al.*[93] using dynamic programming that promotes stability of snake and hard constraints to be enclosed at cost of expensive computation. Williams and Shah[91] proposed another approach based on the Greedy algorithm, which is faster than the method of Amini *et al.*[93] Suppose that the contour has N points (Fig. 10), which are allowed to move to any point in a neighbourhood of size M at each iteration. The Greedy algorithm is described as follows:

Initialize α_i, β_i, and γ_i to 1 for all i
Do
/* Loop to move points to new locations */
For i = 1 to N /*N: number of points in contour*/

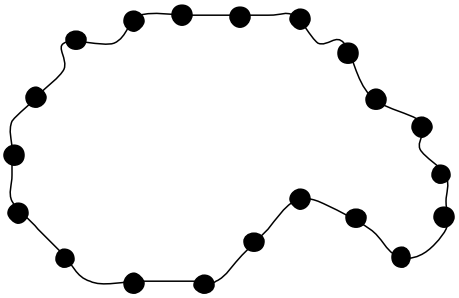

FIGURE 10. Curve representation with a set of N points $(v_i(x_i,y_i)$, $i = 1, \cdots, N)$.

$E_{min} \leftarrow \infty$

For $j = 1$ to N

$E(j) \leftarrow \alpha(i)E_1(j) + \beta(i)E_2(j) + \gamma(i)E_3(j)$

If $E_j < E_{min}$

$E_{min} \leftarrow E(j)$

$j_{min} \leftarrow j$

Move point v_i to location j_{min}

If j_{min} is not current location, ptsmoved ++

Until ptsmoved < Th_Mov

where ptsmoved is the number of points which have been moved to the new locations. Th_Mov is a threshold to decide whether we have converged according to the number of points that moved in each iteration. Figure 11 shows the result of the snake applied to a gated blood pool image. Fig. 11(a) shows a cardiac image acquired in the LAO projection providing the septal separation of the two ventricles. The aim is to delineate correctly the LV contour in order to estimate the EF. The initial contour drawn by the user is

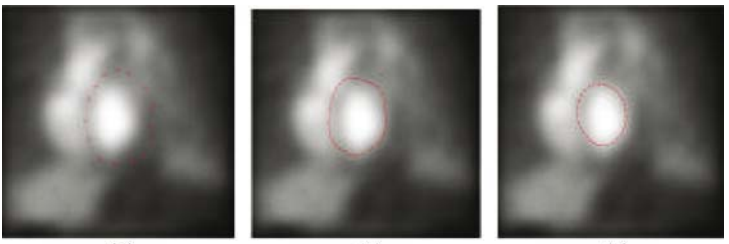

FIGURE 11. Snake applied to blood pool cardiac image. (a) Initial contour. (b) Contour obtained after 20 iterations. (c) The LV enclosed by the snake.

shown in Fig. 11(a). Contour obtained after 20 iterations with Th_Mov set to 6 is represented in Fig. 11(b). As shown in Fig. 11(c), the LV is well delimited. Results of delineation of the cortical structure from PET brain images are reported in ref.[94] These results show the interest to use deformable models for repeatable brain regional extraction applicable for regional quantification of tracer uptake.

3.9 Atlas-guided Approaches

Atlas-guided medical image segmentation could be a robust and efficient approach when a standard atlas or "template" for the particular organ and imaging modality being investigated is available. The atlas is produced by gathering detailed information on the anatomy of the structure to be segmented for a particular patient population. The template is then used as a reference frame for segmenting other images.[3] It is worth emphasizing that atlas-guided approaches are conceptually similar to classifiers described in section 3.3 keeping in mind that they are implemented in the spatial domain rather than in a feature space.

The standard atlas-guided approach considers segmentation as a registration problem (see chapter 9 for a detailed survey of registration techniques). The algorithm starts with a process called *atlas warping*, that is an estimate is made of the transformation matrix that matches a pre-segmented atlas image to the target image to be segmented. This transformation can be linear,[95] non-linear[96,97] or a combination of both[98] to handle efficiently anatomical variability. It is worth emphasizing that this issue is rather difficult to solve and might result in inaccurate segmentation of most dissimilar complexly-shaped structures exhibiting large variability in size and shape among subjects. Another important conceptual limitation for this technique, namely the existence of patient-specific anomalies that are obviously not modelled in an atlas obtained from a single or even an average representation of the population. It is therefore well accepted that atlas-guided approaches are more appropriate for segmentation of structures that do not vary significantly among subjects. The use of probabilistic atlases has proven to be useful in modelling anatomical variability at the expense of additional complexity and increase in computation time.[99] Atlas-guided approaches have been developed and applied mainly in MR brain imaging. Some of them are, however, suitable for an extension towards segmentation of nuclear medical images.

3.10 Analysis of Time Activity Curves

Gated blood pool scintigraphy (SPECT and PET) can measure the functional changes in tissues.[11,100] Theses modalities generate images demonstrating temporal changes in radioactive tracer distribution. For example, gated blood pool scintigraphy is a reliable non-invasive method for detecting

abnormalities of contraction and conduction of the heart. Dynamic imaging provides a time sequence of images taken throughout the whole cardiac cycle. The quantitative information contained in the images describes the physiological behaviour of the imaged structures such as the LV and RV.

In routine dynamic studies, RoI delineation is required for quantitative analysis. A common approach is to identify the anatomic structures by placing RoIs directly on the images, and the underlying tissue time-activity curves (TACs) are then extracted for subsequent analysis. This RoI analysis approach, although widely used in clinical and research settings, is operator-dependent and prone to reproducibility errors. A semi-or fully-automated RoI delineation (or segmentation) method offers advantages by reducing operator error/subjectivity, thereby improving reproducibility.[100] Automatic segmentation may be an alternative to manual RoI delineation. Cluster analysis has been investigated to partition sets of tissue TACs into a set of homogeneous TACs (clusters). It was found that a linear combination of those homogeneous TACs using a mixture model can accurately represent the original data while providing quantitative parametric images.[101,102] Wong *et al.*[100] used a non parametric C-means clustering technique to segment dynamic PET data into C characteristic curves automatically. However, large-scale and multi-centric studies for different applications are necessary to demonstrate the potential of automatic segmentation for replacing manual RoI delineation.

Conventional analysis of image sequences is usually performed by visual evaluation of differences from image to image in order to obtain qualitative information for example about the cardiac contraction kinetics or tissue perfusion. This is not the optimal way to interpret the changes appearing in the images along the series. This operation is tedious and suffers from observer bias. Furthermore, a sequence of images contains also spatially differentiated quantitative information describing the behaviour of the imaged structure, which is difficult to extract by visual evaluation. As an alternative to visual evaluation, the required information can be obtained from an image sequence using parametric analysis methods. Based on the first harmonic fit to the pixel time activity curve, Fourier analysis[103] has also been used to detect and describe wall motion abnormalities in gated blood pool scintigraphy.[104,105] The primary assumption of Fourier analysis is that the data are periodic and the second is that the transition between the first and the last frame of the study must be smooth. This analysis generates two parametric images, the amplitude image corresponding to the maximal change in counts within the image series regardless of the time these changes occur and the phase image corresponding to the time of maximal contraction (ED). In a first approximation, the amplitude image is proportional to the stroke volume. The phase image allows good separation of the ventricular regions (left or right) from the atria and this processing is useful to delineate automatically the LV contours in order to estimate the EF value from ventricular regions of interest.[4]

Let Vol(i, j, k) be the value of the pixel (i, j) of the k^{th} cardiac image. Two images, respectively, I_{cos} and I_{sin} are calculated using the following equations:

$$\begin{aligned} I_{cos}(i,j) &= \sum_{k=1}^{T_C} \cos\left[\frac{2\pi}{T_C}(k-1)\right] \times Vol(i,j,k) \\ I_{sin}(i,j) &= \sum_{k=1}^{T_C} \sin\left[\frac{2\pi}{T_C}(k-1)\right] \times Vol(i,j,k) \end{aligned} \tag{57}$$

where T_C is the number of images in the series corresponding to the cardiac cycle duration. The phase image is given by

$$I_{phase}(i,j) = \arctan\left(\frac{I_{sin}(i,j)}{I_{cos}(i,j)}\right) \tag{58}$$

and the amplitude by:

$$I_{amp}(i,j) = \sqrt{(I_{cos}(i,j))^2 + (I_{sin}(i,j))^2} \tag{59}$$

For cardiac studies, the amplitude image is a picture of pixel-by-pixel stroke volume. The rationale for using the phase image rests on the observation that ischemia reduces the velocity of myocardial contraction and thus local abnormalities of wall motion are frequently associated with changes in the time of wall movement.

In spite of the apparent mathematical difficulties, Factor Analysis of Dynamic Structures (FADS) has gained clinical acceptance for cardiac studies at equilibrium with abnormalities of both contraction and conduction.[106–108] This analysis assumes that any time activity curve is a weighted sum of a limited number, q, of pure time activity evolutions, called physiological components.[107] These components correspond to regions of similar temporal behaviour.

Assuming that image sequence has m frames each containing n pixels, this can be expressed as a *q* -factor problem:[108]

$$X = \hat{X} + E;\ \hat{X} = FA \tag{60}$$

where $X, \hat{X}$, E, F and A are m × n,m × n, m × n, m × q and q × n matrices, respectively. X is the data matrix (noise corrupted observation) whose columns contain TACs of each pixel; $\hat{X}$ is the relevant part of the observation; E is an error matrix; F contains the q temporal factors in its columns and A contains the q factor images in its rows. The FADS solution involves choosing the number of factors q, minimizing the size of E in a suitable sense and determining the composition of $\hat{X}$ into FA by appropriate constraints.[108] For dynamic studies in nuclear medicine a Principal Component Analysis (PCA) may be performed on the population of pixels (TAC for each pixel) in the image to extract the first p components. A Singular Value

Decomposition (SVD) of the row space of $X(m < n)$ gives $X = PL^{0.5}Q^T$. P and Q are $m \times m$ and $n \times m$ column orthogonal matrices and L is the $m \times m$ diagonal matrix with non-negative eigenvalues in decreasing order. Let P_p and Q_p represent the first p columns of P and Q; L_P the upper $p \times p$ submatrix of L; and $D_p = P_p L_p^{0.5} Q_p^T$ the $m \times n$ matrix. Then $X = X_p + \varepsilon_p = C_e W_e + \varepsilon_p$ where $C_e = P_p$; $W = L_p^{0.5} Q_p^T$. C_e is an $m \times p$ matrix containing the orthogonal components in its columns; W_e is a $p \times n$ matrix containing the pixel coefficients for each component in its corresponding rows; and ε_p is the residual error matrix. It is now assumed that $p = q$. Since PCA minimizes residual variance for a given number of components, it is natural to estimate E by ε_p and thus estimate $\hat{\mathbf{X}} = C_e W_e$. Transforming a PC solution to p-factor solution involves finding a non-singular $p \times p$ transformation matrix V such that $F = C_e V$; $A = V^{-1} W_e$. The columns of W (one per pixel) will plot as points q-dimensional feature space, as will the columns of V (one per structure). The problem reduces to locating the points in feature space corresponding to the columns of V. Since nuclear medicine data are in principle non-negative, the elements of F and A should be non-negative.[108]

The estimated physiological components given by FASD are used to reconstruct functional images (factor images). This method has an obvious similarity with Fourier analysis: while Fourier analysis is based on sine and cosine functions, no such restriction is placed on the shape of the PCs. Although factor analysis provides a valid representation of wall motion and conduction abnormalities, some limitations must be considered. The quality of the results and the number of significant factors depend on the signal to noise ratio of the dynamic series.[107]

Recently, a new technique designed to analyze nuclear cardiac image series was proposed.[11] Compared to Fourier analysis, this method does not assume that the data are periodic and no restriction is placed on the shape of the pixel TAC. This method segments the time series of the cardiac images in regions of similar temporal behaviour (components) but, contrary to PCA, knowledge of their number is not needed to run the algorithm.[11] The aim of this technique is to capture the essential of the sequence while reducing the amount of image data presented to the clinician for diagnostic interpretation. This method is based on the measure of similarity between the temporal response of pixels in a reference RoI and the temporal response of the other pixels in the image series. The calculated similarity provides quantitative information about the degree of local similarity in comparison with the RoI taken as a reference. Template matching is perhaps the most straightforward way of locating the presence of an object of interest within an image field. The template is a duplicate of the sought object. If the template matches an unknown object and if it is sufficiently close, the unknown object is labelled as the template object. In nuclear cardiac studies, the template is a fixed reference TAC describing the temporal evolution of a

cardiac RoI during the cardiac cycle. The image field to be searched is the time series of cardiac images. For dynamic analysis, template matching may be viewed as a temporal one. For example, the template may therefore be a TAC of a subregion of the LV or the whole ventricle. Similarity measures provide quantitative means of determining the degree of temporal match between the time sequence of images and the template. The similarity used is the covariance function. The covariance value of the $(i, j)^{th}$ pixel is given by:

$$\mathrm{Cov}(i, j) = \frac{1}{T_C} \sum_{t=1}^{T_C} (\mathrm{Vol}(i, j, t) - \mu_{AC}(i, j))(R(t) - \mu_R) \tag{61}$$

where

$$\begin{aligned} R(t) &= \frac{1}{M_C} \sum_{(i,j)\in \mathrm{RoI}} \mathrm{Vol}(i, j, t) \\ \mu_R &= \frac{1}{T_C} \sum_{t=1}^{T_C} R(t),\ \mu_{AC}(i, j) = \frac{1}{T_C} \sum_{t=1}^{T_C} \mathrm{Vol}(i, j, t) \end{aligned} \tag{62}$$

$R(t)$ is the reference series, μ_R the mean value of the reference series and $\mu_{AC}(i, j)$ the mean value of the time activity curve of the $(i, j)^{th}$ pixel. M_C is the total number of pixels in the RoI.

There is no major problem of spatial registration for this kind of images because the images of the series correspond to the same morphologic projection even if its shape changes with time. The method is based on the computation, pixel by pixel, of the covariance (Eq. 61) between two time series representing the TAC of any pixel and a reference time series. The generated similarity map is an image where the value of each pixel represents the degree of temporal similarity to the reference. This similarity measure is less than the absolute value of $\sigma_{AC}\sigma_R$, where σ_{AC} and σ_R are the standard deviations of the TAC and the reference, respectively. A positive value of the covariance indicates that the TAC and the reference variables vary in the same sense. While, a negative value indicates that these two variables vary in opposite sense. It is important to keep in mind that all the pixels that match or mismatch the reference are equally important to describe or to interpret the information contained in the image series.

Gated cardiac blood pool images are analyzed by the covariance technique. The imaging is performed in LAO projection. Figure 12 shows the placement of four RoI references. For covariance analysis, only one reference is used. Each reference time series is calculated in, for example, a 5×5 pixels window size of the corresponding RoI. Once a covariance map is calculated for all pixels, with respect to the chosen reference RoI, its values are mapped into 0 through 255. The map is displayed with 256 colours look-up table and a colour coded image represents the distribution of the temporal degree of similarity to be assessed. The red colour corresponds to a maximum covariance value while the blue colour corresponds to minimum

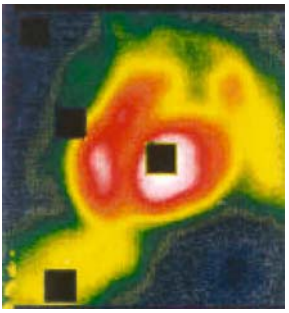

FIGURE 12. End-diastolic image with the reference RoIs of the background (top left), the liver (bottom left), the LV (middle right) and the atria (middle left). (Reprinted with permission from ref.[11]).

covariance value. Like in the phase analysis, the strength of the covariance image (similarity map) lies in its relative and not in its absolute values. For a ventricular reference, the similarity maps of the ventricles are displayed in red and the atria in blue. This is expected since the atrial pixels are out of phase with the ventricular ones. Figure 13 shows the results obtained in four pathological patients (from top to bottom: patients 1–4) with RoI reference placed in the LV region. The left and the middle columns of figure 13 represent the end diastolic and systolic images of the four patients, respectively. The corresponding similarity maps are shown in the right column. In patient 1, the similarity map shows the septal portion of the LV with a delayed temporal response compared to the remaining segments of the ventricle. Although the LV has a normal morphology, it presents an anteroseptal and hypokinesia. Patient 2 has a largely dilated LV. Note the opposition of temporal responses of the ventricular and the atria are well demonstrated. As the LV, the RV is dilated but its anterobasal portion is hypokinetic. Patient 3 also has a dilated LV. The covariance map reveals, in the LV, an akinesis of the inferoapical segment and dyskinesis of the distal septum. In patient 4, the map shows that there is an akinesis of the inferior part of the RV with a limited dyskinetic region. One may also note a hypokinesis of the septal segment of the LV.

4. Clinical and Research Applications of Image Segmentation

Image segmentation has found numerous clinical and research applications both in diagnostic and therapeutic nuclear medicine. This includes estimation of organ (e.g. liver, spleen, heart, thyroid . . .) or tumour volumes as well as target treatment volumes definition in radiation therapy, extraction of parameters of clinical relevance such as left ventricular region in nuclear

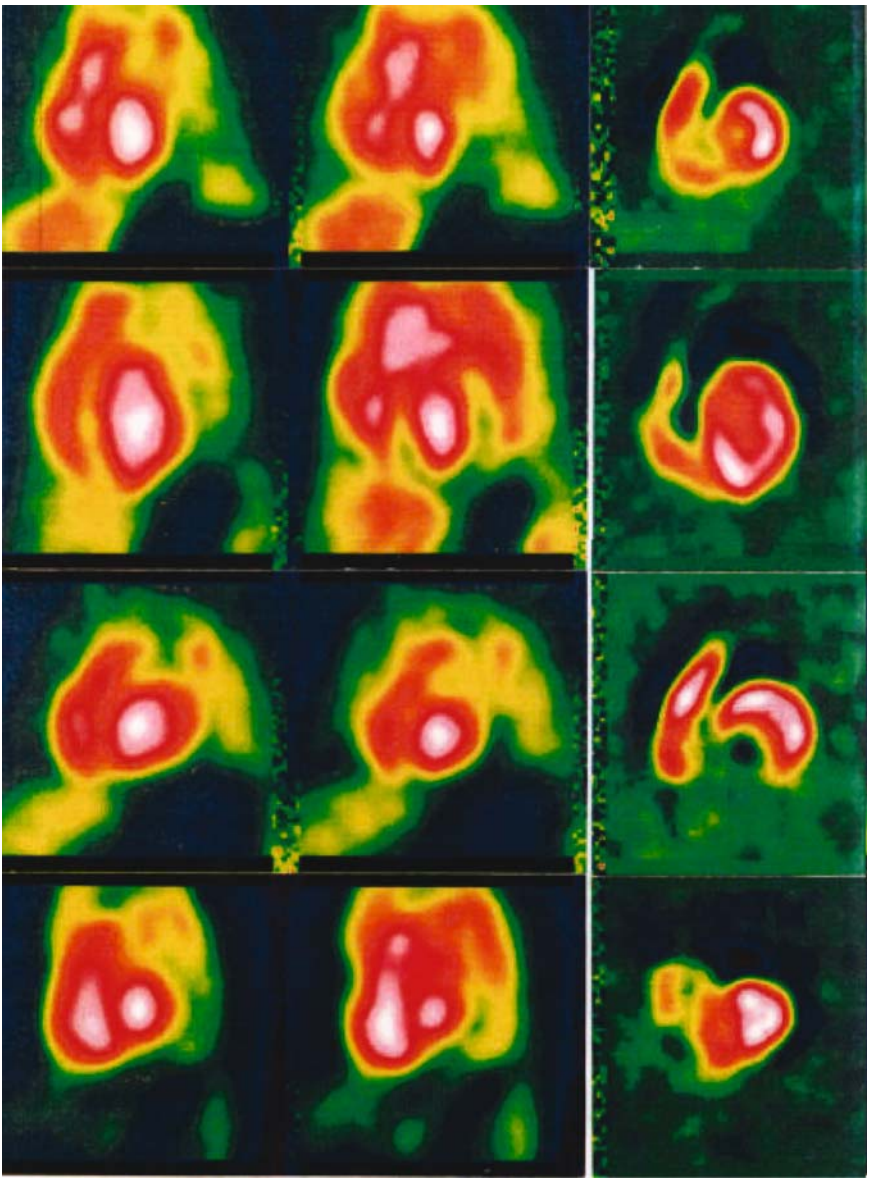

FIGURE 13. Study of four patients showing end-diastolic images (left column), end-systolic images (middle column) and the covariance map of the LV RoI reference (right column). (Reprinted with permission from ref.[11]).

cardiology, automated RoI delineation for structures of interest in dynamic functional imaging, determination of the attenuation map in emission tomography and construction of voxel-based anthropomorphic phantoms based on high resolution anatomical images. For the latter, the interested reader is refereed to a recent book describing the development of such computational models in connection with Monte Carlo modelling tools in nuclear imaging.[109]

In diagnostic and oncological nuclear medicine imaging, if the intention is to measure the volume and uptake in a specific organ or lesion, it may be convenient to make preliminary phantom studies on an object of similar shape and size with the assumption that the size, shape, location, and density of the simulated object and surrounding tissues are comparable to that of the patient to be measured. Absolute quantitation by SPECT and PET is feasible now and has been shown to be a clinically reliable and useful technique. For example, the accurate determination of thyroid volume may be an important factor in the calculation of the most appropriate dose of therapeutic radioiodine.[9,110] The volume measurement of different organs and lesions using SPECT and PET has been widely applied in the clinical field.[10,29,111–113] The major problems encountered in functional volume quantitation are image segmentation and imperfect system transfer function. The difficulty in image segmentation is compounded by the low spatial resolution and high noise characteristics of nuclear images.[114] Long *et al.*[28] have evaluated the relative performance of a number of image segmentation methods for volume quantitation on simulated SPECT images. This included: (i) manual methods in which the operator defines the surface boundary of an image structure by delineating its cross-sectional edges in the contiguous slices incorporating it, (ii) algorithms based on simple thresholding techniques to define the edge of the organ where voxels are flagged as part of the volume of interest if they contain counts greater than some pre-defined threshold value, (iii) the adaptive thresholding method which uses a nonparametric procedure of automatic threshold selection formulated by Otsu,[20] and (iv) two-dimensional (2-D) and three-dimensional (3-D) implementations of a multifeature determination of edge location based on calculation of the gradient (magnitude and direction of the greatest rate of change in counts). They conclude that the 3-D gradient-based method of image segmentation is the one requiring minimal operator intervention while providing the most accurate and consistent estimates of object volume across changes in object size and contrast. A survey of different methods proposed for evaluation of image segmentation algorithms including the criteria for choosing figures of merit is given elsewhere.[115]

Segmentation of radionuclide attenuation images for the purpose of reducing noise propagation in short transmission scanning or X-ray CT and MRI for identifying tissues with different attenuation properties is also an appealing research area.[116] In particular, Pan *et al.*[117] reported on an accurate segmentation method to determine the attenuation map from Compton

scatter and photopeak window SPECT projection data. During the last decade, techniques using transmission image segmentation and tissue classification have been proposed to minimize the acquisition time and increase the accuracy of the attenuation correction process, while preserving or even reducing the noise level. The reconstructed transmission image pixels are segmented into populations of uniform attenuation. In a clinical setting, segmentation algorithms must be designed to balance image quality and computational time of the emission tomograms. The majority of segmentation methods used for attenuation correction fall into one of the following two classes (see chapter 6): histogram-based thresholding techniques[118–120] and fuzzy-clustering based segmentation techniques.[12,13] Other interesting approaches to segment noisy transmission data include the use of active contour models,[121] neural networks[122] morphological segmentation[123] and hidden Markov modelling.[124] An alternative to segmentation of transmission images with the goal of reducing noise in PET transmission measurements is the use of coregistered segmented MRI data in functional brain imaging.[125] The T_1-weighted brain MR images were segmented by means of a fuzzy clustering technique which identifies tissues of significantly different density and composition. A promising technique for segmentation of skull in human T_1-weighted MR images, which could be useful for more robust implementation of the method mentioned above, consists in performing skull segmentation using a sequence of mathematical morphological operations.[126]

5. Summary

It is gratifying to see in overview the progress that image segmentation has made in the last ten years, from operator-dependent manual delineation of structures, through simple thresholding, the use of classifiers and fuzzy clustering, and more recently atlas-guided approaches incorporating prior information. Recent developments have been enormous particularly in the last ten years, the main opportunities striving towards improving the accuracy, precision, and computational speed through efficient implementation in conjunction with decreasing the amount of operator interaction. The application of medical image segmentation is well established in research environments and is still limited in clinical settings to institutions with advanced physics and extensive computing support. As the above mentioned challenges are met, and experience is gained, implementation of validated techniques in commercial software packages will be useful to attract the interest of the clinical community and increase the popularity of these tools. It is expected that with the availability of computing power in the near future, more complex and ambitious computer intensive segmentation algorithms will become clinically feasible.

Acknowledgments. HZ was supported in part by the Swiss National Science Foundation under grant SNSF 3152A0-102143 and Geneva University Hospital R&D funds under grant PRD-04-1-08. AB wishes to express his sincere thanks to Dr. F. Behloul, Leiden University Medical Center (Department of Radiology) for her helpful discussions and suggestions. Illustrations presented in *Artificial Neural Networks* section are essentially inspired from her PhD dissertation.

References

1. Suetens P., Bellon, E., Vandermeulen, D. *et al.*, Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *Eur J Radiol* **17:** 14-21 (1993).
2. Deklerck R., Cornelis, J. and Bister, M., Segmentation of medical images. *Image Vision Comput* **11:** 1101-1113 (1993).
3. Pham D. L., Xu, C. and Prince, J. L., A survey of currents methods in medical image segmentation. *Ann Rev Biomed Eng* **2:** 315-337 (2000).
4. Boudraa A. O., Mallet, J.-J., Besson, J.-E. *et al.*, Left ventricle automated detection method in gated isotopic ventriculography using fuzzy clustering. *IEEE Trans Med Imaging* **12:** 451-465 (1993).
5. Boudraa A. O., Champier, J., Cinotti, L. *et al.*, Delineation and quantitation of brain lesions by fuzzy clustering in positron emission tomography. *Comput Med Imaging Graph* **20:** 31-41 (1996).
6. Boudraa A. O., Arzi, M., Sau, J. *et al.*, Automated detection of the left ventricular region in gated nuclear cardiac imaging. *IEEE Trans Biomed Eng* **43:** 430-437 (1996).
7. Behloul F., Boudraa, A.O., Lelieveldt, B. P. *et al.*, Myocardium extraction in positron emission tomography based on soft computing. *Comput Med Imaging Graph* **25:** 277-286 (2001).
8. Bezdek J. C., *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, (1981).
9. Zaidi H., Organ volume estimation using SPECT. *IEEE Trans Nucl Sci* **43:** 2174-2182 (1996).
10. Erdi Y. E., Mawlawi, O., Larson, S. M. *et al.*, Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* **80:** 2505-2509 (1997).
11. Boudraa A. O., Champier, J., Djebali, M. *et al.*, Analysis of dynamic nuclear cardiac images by covariance function. *Comput Med Imaging Graph* **23:** 181-191 (1999).
12. Bettinardi V., Pagani, E., Gilardi, M. *et al.*, An automatic classification technique for attenuation correction in positron emission tomography. *Eur J Nucl Med* **26:** 447-458 (1999).
13. Zaidi H., Diaz-Gomez, M., Boudraa, A. O. *et al.*, Fuzzy clustering-based segmented attenuation correction in whole-body PET imaging. *Phys Med Biol* **47:** 1143-1160 (2002).
14. Peter J., Freyer, R., Smith, M. *et al.*, Nuclear medicine image segmentation using a connective network. *IEEE Trans Nucl Sci* **44:** 1583-1590 (1997).

15. Clarke L. P., Velthuizen, R. P., Camacho, M. A. *et al*., MRI segmentation: methods and applications. *Magn Reson Imaging* **13:** 343-368 (1995).
16. Pal N. R. and Pal, S. K., A review on image segmentation techniques. *Pattern Recognition* **26:** 1277-1294 (1993).
17. Gonzallez R. C. and Woods, R. E., *Digital image processing*., Pearson Education International, 2nd edition, Upper Saddle River, NJ, (2002).
18. Canny J. F., A computational approach to edge detection. *IEEE Trans Pattern Anal Machine Intell* **8:** 679-698 (1986).
19. Albregtsen F., ''Non-parametric histogram thresholding methods. Error versus relative object area'' Proc. 8th Scandinavian Conf. Image Processing, *pp* 273-280 (1993).
20. Otsu N., A thresholding selection method from gray-level histograms. *IEEE Trans Sys Man Cyber* **9:** 62-66 (1979).
21. Reddi S. S., Rudin, S. F. and Keshavan, H. R., An optimal multiple threshold scheme for image segmentation. *IEEE Trans Syst Man and Cybern* **14:** 661-665 (1984).
22. Kittler J. and Illingworth, J., Minimum error thresholding. *Patt Recogn* **19:** 41-47 (1986).
23. Pun T., A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Process* **2:** 223-237 (1980).
24. Pun T., Entropic thresholding: a new approach. *Comput Graph Image Process* **16:** 210-239 (1981).
25. Kapur J. N., Sahoo, P. K. and Wong, A., A new method for gray-level picture thresholding using the entropy of the histogram. *Graph Models Imag Proc* **29:** 273-285 (1985).
26. Mortelmans L., Nuyts, J., Van Pamel, G. *et al*., A new thresholding method for volume determination by SPECT. *Eur J Nucl Med* **12:** 284-290 (1986).
27. Groshar D., Frankel, A., Iosilevsky, G. *et al*., Quantitation of renal uptake of technetium-99m DMSA using SPECT. *J Nucl Med* **30:** 246-250 (1989).
28. Long D. T., King, M. A. and Sheehan, J., Comparative evaluation of image segmentation methods for volume quantitation in SPECT. *Med Phys* **19:** 483-489 (1992).
29. Alaamer A. S., Fleming, J. S. and Perring, S., Evaluation of the factors affecting the accuracy and precision of a technique for quantification of volume and activity in SPECT. *Nucl Med Commun* **15:** 758-771 (1994).
30. Savolainen S., Pohjonen, H., Sipila, O. *et al*., Segmentation methods for volume determination with 111In/99mTc SPET. *Nucl Med Commun* **16:** 370-377 (1995).
31. Fleming J. S. and Alaamer, A. S., A rule based method for context sensitive threshold segmentation in SPECT using simulation. *Phys Med Biol* **43:** 2309-2323 (1998).
32. Wang S. and Haralick, R. M., Automatic multithreshold selection. *Comput Vis Graph Imag Process* **25:** 46-67 (1984).
33. Carlotto J. M., Histogram analysis using scale-space approach. *IEEE Trans Pattern Anal Mach Intell* **9:** 121-129 (1987).
34. Papamarkos N. and Gatos, B., A new approach for multithreshold selection. *Graph Models Imag Proc* **56:** 357-370 (1994).
35. Van Laere K. J., Warwick, J., Versijpt, J. *et al*., Analysis of clinical brain SPECT data based on anatomic standardization and reference to normal data: an ROC-

based comparison of visual, semiquantitative, and voxel-based methods. *J Nucl Med* **43:** 458-469 (2002).
36. Slomka P. J., Hurwitz, G. A., St. Clement, G. *et al.*, Three-dimensional demarcation of perfusion zones corresponding to specific coronary arteries: application for automated interpretation of myocardial SPECT. *J Nucl Med* **36:** 2120-2126 (1995).
37. Skurichina M. and Duin, R. P. W., "Stabilizing classifiers for very small sample sizes" Proc. 13th Int. Conf. Patt. Recogn. Vol. 12; *pp* 891-896 (1996).
38. Raudys S. J. and Duin, R. P. W., Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. *Patt Recogn Letters* **19:** 385-392 (1998).
39. Hamamoto Y., Fujimoto, Y. and Tomita, S., On the estimation of a covariance matrix in designing Parzen classifiers. *Patt Recogn* **29:** 1751-1759 (1996).
40. Vapnik V. N., *Statistical learning theory*, John Wiley & Sons Inc, New York, (1998).
41. Stoeckel J., Malandain, G., Migneco, O. *et al.*, "Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients" in Conf. Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI'01), Utrecht, The Netherlands, Vol. 2208; *pp* 666-674 (2001).
42. Raudys S. J. and Jain, A. K., Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Trans Patt Anal Mach Intell* **13:** 252-264 (1991).
43. Xie X. L. and Beni, G., A validity measure for fuzzy clustering. *IEEE Trans Pattern Anal Machine Intell* **13:** 841-847 (1991).
44. Pal N. R. and Bezdek, J. C., On cluster validity for the fuzzy c-means model. *IEEE Trans Fuzzy Syst* **3:** 370-379 (1995).
45. Boudraa A. O., Dynamic estimation of number of clusters in data sets. *IEEE Electronics Letters* **35:** 1606-1608 (1999).
46. Coleman G. B. and Andrews, H. C., Image segmentation by clustering. *Proc. IEEE* **5:** 773-785 (1979).
47. Dempster A. P., Laird, N. M. and Rubin, D. B., Maximum likelihood estimation from incomplete data via the EM algorithm. *J Royal Stat Soc B* **39:** 1-38 (1977).
48. Acton P. D., Pilowsky, L. S., Kung, H. F. *et al.*, Automatic segmentation of dynamic neuroreceptor single-photon emission tomography images using fuzzy clustering. *Eur J Nucl Med* **26:** 581-590 (1999).
49. Boudraa A. O. and Clarysse, P., "Fast fuzzy grey level image segmentation method" World Congress on Medical Physics and Biomedical Engineering, Nice, France, Vol. 35; *pp* 686 (1997).
50. Krishnapuram R. and Keller, J. M., A possibilistic approach to clustering. *IEEE Trans Fuzzy Syst* **1:** 98-110 (1993).
51. Beni G. and Liu, X., A least biased fuzzy clustering method. *IEEE Trans Pattern Anal Machine Intell* **16:** 954-960 (1994).
52. Barni M., Cappellini, V. and Mecocci, A., Comments on "A possibilistic approach to clustering". *IEEE Trans Fuzzy System* **4:** 393-396 (1996).
53. Marr D. and Hildreth, E., Theory of edge detection. *Proc Roy Soc London* **207:** 187-217 (1980).
54. Huertas A. and Medioni, G., Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks. *IEEE Trans Patt Anal Mach Intell* **8:** 651-664 (1986).

55. Wang C. X., Small, L., Synder, W. E. *et al.*, ''Edge gated cardiac nuclear medicine images'' 7th IEEE Symposium on computer-based medical systems, Winston-Salem, NC, *pp* 28-33 (1994).
56. Dai X., Snyder, W. E., Bilbro, G. L. *et al.*, Left-ventricle boundary detection from nuclear medicine images. *J Digit Imaging* **11:** 10-20 (1998).
57. Long D. T., King, M. A. and Gennert, M. A., Development of a 3D gradient-based method for volume quantitation in SPECT. *IEEE Trans Nucl Sci* **38:** 748-754 (1991).
58. Hillel P. G. and Hastings, D. L., A three-dimensional second-derivative surface-detection algorithm for volume determination on SPECT images. *Phys Med Biol* **38:** 583-600 (1993).
59. Nihsimura Y., Katafuchi, T., Hirase, Y. *et al.*, Measurement of left ventricular chamber and myocardial volume in hypertrophic cardiomyopath patients by ECG-gated myocardial perfusion SPECT: Application of newly developed edge-detection algorithm. *Jpn J Radiol Technol* **58:** 1586-1591 (2002).
60. Langleben D. D., Austin, G., Krikorian, G. *et al.*, Interhemispheric asymmetry of regional cerebral blood flow in prepubescent boys with attention deficit hyper-activity disorder. *Nucl Med Comm* **22:** 1333-1340 (2001).
61. Li S. Z., *Markov Random field modeling in computer vision*, Springer-Verlag, Tokyo, (1995).
62. Geman S. and Geman, D., Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans Patt Anal Mach Intell* **6:** 721-741 (1984).
63. Besag J., On the statistical analysis of dirty pictures. *J Royal Stat Soc B* **48:** 259-302 (1986).
64. Derin H. and Elliott, H., Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans Patt Anal Mach Intell* **9:** 39-55 (1987).
65. Chen J. L., Gunn, S. R., Nixon, M. S. *et al.*, ''Markov random field models for segmentation of PET images'' in: *Proc. Information Processing in Medical Imaging*, edited by and R.M. Leahy M.F. Insana, (2001), *pp* 468-474.
66. Lin K. P., Chen, T. S., Yao, W. J. *et al.*, ''Dynamic PET-FDG image segmentation using Markov Random field method'' Proc SPIE Int Soc Opt Eng, Vol. 3338; *pp* 1198-1204 (1998).
67. Mignotte M. and Meunier, J., ''Unsupervised restoration of brain SPECT volumes'' Vision Interfaces (VI 2000), Montreal, Canada, *pp* 55-60 (2000).
68. Mignotte M., Meunier, J., Soucy, J. P. *et al.*, ''Segmentation and classification of brain SPECT images using 3D MRF and density mixture estimations'' 5th World Multi-Conference on Systemics, Cybernetics and Informatics, Concepts and Applications of Systemics and Informatics, Orlando, FL, USA, Vol. 10; *pp* 239-244 (2001).
69. Lin H. D., Lin, K. P., Chung, B. T. *et al.*, ''Using kinetic parameter analysis of dynamic FDOPA-PET for brain tissue classification'' Proc SPIE Int Soc Opt Eng, Vol. 4683; *pp* 436-449 (2002).
70. Kohonen T., *Self-organization and associative memory*, Springer-Verlag, Berlin, (1989).
71. Pao Y. H., *Adaptive pattern recognition and neural networks*, Addison-Wesley, New York, (1989).

72. Eckhorn R., Reitboeck, H. J., Arendt, M. *et al*., Feature linking via synchronization among distributed assemblies: simulation of results from cat visual cortex. *Neural Computation* **2:** 293-307 (1990).
73. Hertz J., Krogh, A. and Palmer, R. G., *Introduction to the theory of neural computation*, Addison-Wesley, New York, (1991).
74. Kosko B., *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence*, Prentice-Hall, Englewood Cliffs, New Jersey, (1991).
75. Blanz W. E. and Pal, S. L., Image segmentation using a neural network. *Biol. Cybern*. **66:** 151-158 (1991).
76. Ghosh A., Pal, N. R. and Pal, S. K., Object background classification using Hopfield type neural network. *Int J Pattern Recogn Mach Intell* **6:** 989-1008 (1992).
77. Ghosh A., Pal, N. R. and Pal, S. K., Self-organization for object extraction using multilayer neural networks and fuzziness measures. *IEEE Trans Fuzzy Syst* **1:** 54-68 (1993).
78. Haykin S., *Neural networks. A comprehensive foundation*, Macmillan, New York, (1994).
79. Keller P. E. and McKinnon, A. D., "Segmentation of medical imagery with pulse-coupled neural networks." Int Joint Conf Neural Networks, Washington, DC, USA, Paper #130, available on CDROM (1999).
80. Behloul F., "Fusion de données TEP-IRM par méthodes neuro-floues pour l'étude de la viabilité du myocarde" Ph.D Thesis, INSA de Lyon, 1999.
81. Behloul F., Lelieveldt, B. P., Boudraa, A.O. *et al*., Neuro-fuzzy systems for computer-aided myocardial viability assessment. *IEEE Trans Med Imaging* **20:** 1302-1313 (2001).
82. Musavi M. T., Ahmed, W., Chan, K. H. *et al*., On the training of radial basis function classifiers. *Neural Networks* **5:** 595-603 (1992).
83. Hwang Y. S. and Bang, S. Y., An Efficient method to construct a radial basis function neural network classifier. *Neural Networks* **8:** 1495-1503 (1997).
84. Kang H. B. and Walker, E. L., Characterising and controlling approximation in hierarchical perceptual grouping. *Fuzzy Sets Syst* **65:** 187-223 (1994).
85. Kosko B., Fuzzy entropy and condition. *Inform Sci* **40:** 165-174 (1989).
86. Pal N. R. and Pal, S. K., Higher order of fuzzy entropy and hybrid entropy of a set. *Inform Sci* **61:** 211-231 (1992).
87. Deluca A. and Termini, S., A definition of non probabilistic entropy in the setting of fuzzy set theory. *Inform and Control* **20:** 301-312 (1972).
88. Kandel A., *Fuzzy mathematical technique with applications*, Addison-Wesley, New York, (1986).
89. Kass M., Witkin, A. and Terzopoulos, D., Snakes: active contour models. *Int J Comput Vision* **1:** 321-331 (1988).
90. Leymarie F. and Levine, M. D., Tracking deformable objects in the plane using an active contour model. *IEEE Trans Patt Anal Mach Intell* **15:** 617-634 (1993).
91. William D. J. and Shash, M., A fast algorithm for active contours and curvature estimation. *Image Understanding* **55:** 14-26 (1992).
92. Cohen L. D. and Cohen, I., "A finite element method applied to new active contour models and 3D reconstruction from cross-section" Proc. IEEE, 3rd Int. Conf. Computer Vision, Osaka, Japan, *pp* 587-591 (1990).

93. Amini A. A., Tehrani, S. and Weymouth, T. E., "Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints" Proc. IEEE 2rd Int. Conf. Computer Vision, *pp* 855-867 (1988).
94. Mykkanen J. M., Juhola, M. and Ruotsalainen, U., Extracting VOIs from brain PET images. *Int J Med Inf* **58-59:** 51-57 (2000).
95. Lancaster J., Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp* **5:** 238-242 (1997).
96. Ashburner J. and Friston, K. J., Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* **7:** 254-266 (1999).
97. Minoshima S., Koeppe, R. A., Frey, K. A. *et al*., Anatomic standardization: linear scaling and nonlinear warping of functional brain images. *J Nucl Med* **35:** 1528-1537 (1994).
98. Sandor S. and Leahy, R., Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Trans Med Imaging* **16:** (1997).
99. Thompson P. and Toga, A. W., Detection, visualization and animation of abnormal anatomic structure with a probabilistic brain atlas based on random vector field transformations. *Med Image Analysis* **1:** 271-294 (1997).
100. Wong K.-P., Dagan, F., Meikle, S. R. *et al*., Segmentation of dynamic PET images using cluster analysis. *IEEE Trans Nucl Sci* **49:** 200-207 (2002).
101. O' Sullivan F., Imaging radiotracer model parameters in PET: A mixture analysis approach. *IEEE Trans Med Imaging* **12:** 399-412 (1993).
102. Kimura Y., Hsu, H., Toyama, H. *et al*., Improved signal-to-noise ratio in parametric images by cluster analysis. *Neuroimage* **9:** 554-561 (1999).
103. Bossuyt A., Deconinck, F., Lepoudre, R. *et al*., "The temporal Fourier transform applied to functional isotopic imaging" in: *Information processing in medical imaging*, edited by R Di Paola and E Kahn INSERM 88, Paris, (1979), *pp* 397-408.
104. Botvinick E., Dunn, R., Frais, M. *et al*., The phase image: its relationship to patterns of contraction and conduction. *Circulation* **65:** 551-560 (1982).
105. Pavel D., Byrom, E., Lam, W. *et al*., Detection and quantification of regional wall motion abnormalities using phase analysis of equilibrium gated cardiac studies. *Clin Nucl Med* **8:** 315-321 (1983).
106. Barber D. C., The use of principal components in the quantitative analysis of gamma camera dynamic studies. *Phys Med Biol* **25:** 283-292 (1980).
107. Cavailloles F., Bazin, J. P. and Di Paola, R., Factor analysis in gated cardiac studies. *J Nucl Med* **25:** 1067-1079 (1984).
108. Houston A. S. and Sampson, W. F., A quantitative comparison of some FADS methods in renal dynamic studies using simulated and phantom data. *Phys Med Biol* **42:** 199-217 (1997).
109. Zaidi H. and Sgouros, G., *Therapeutic applications of Monte Carlo calculations in nuclear medicine*, Institute of Physics Publishing, Bristol, (2002).
110. Zaidi H., Comparative methods for quantifying thyroid volume using planar imaging and SPECT *J Nucl Med* **37:** 1421-1426 (1996).
111. Erdi Y. E., Wessels, B. W., Loew, M. H. *et al*., Threshold estimation in single photon emission computed tomography and planar imaging for clinical radioimmunotherapy. *Cancer Res* **55:** 5823s-5826s (1995).
112. Erdi Y. E., Humm, J. L., Imbriaco, M. *et al*., Quantitative bone metastases analysis based on image segmentation. *J Nucl Med* **38:** 1401-1406 (1997).

113. Sjorgreen K., Ljungberg, M. and Strand, S. E., Parameters influencing volume and activity quantitation in SPECT. *Acta Oncol* **35:** 323-330 (1996).
114. King M. A., Long, D. T. and Brill, A. B., SPECT volume quantitation: influence of spatial resolution, source size and shape, and voxel size. *Med Phys* **18:** 1016-1024 (1991).
115. Zhang Y. J., A survey on evaluation methods for image segmentation. *Patt Recogn Letters* **29:** 1335-1346 (1996).
116. Zaidi H. and Hasegawa, B. H., Determination of the attenuation map in emission tomography. *J Nucl Med* **44:** 291-315 (2003).
117. Pan T.-S., King, M. A., de Vries, D. J. *et al*., Segmentation of the body and lungs from Compton scatter and photopeak window data in SPECT: a Monte-Carlo investigation. *IEEE Trans Med Imaging* **15:** 13-24 (1996).
118. Meikle S. R., Dahlbom, M. and Cherry, S. R., Attenuation correction using count-limited transmission data in positron emission tomography. *J Nucl Med* **34:** 143-150 (1993).
119. Xu M., Cutler, P. and Luk, W., An adaptive local threshold segmented attenuation correction method for whole-body PET imaging. *IEEE Trans Nucl Sci* **43:** 331-336 (1996).
120. Bilger K., Kupferschlager, J., Muller-Schauenburg, W. *et al*., Threshold calculation for segmented attenuation correction in PET with histogram fitting. *IEEE Trans Nucl Sci* **48:** 43-50 (2001).
121. Tai Y.-C., Lin, K.-P., Dahlbom, M. *et al*., A hybrid attenuation correction technique to compensate for lung density in 3-D total body PET. *IEEE Trans Nucl Sci* **43:** 323-330 (1996).
122. Yu S. K. and Nahmias, C., Segmented attenuation correction using artificial neural networks in positron tomography. *Phys Med Biol* **41:** 2189-2206 (1996).
123. Riddell C., Brigger, P., Carson, R. E. *et al*., The watershed algorithm: a method to segment noisy PET transmission images. *IEEE Trans Nucl Sci* **46:** 713-719 (1999).
124. Anderson J. M. M., Srinivasan, R., Mair, B. A. *et al*., Hidden Markov model based attenuation correction for positron emission tomography. *IEEE Trans Nucl Sci* **49:** 2103-2111 (2002).
125. Zaidi H., Montandon, M.-L. and Slosman, D. O., Magnetic resonance imaging-guided attenuation and scatter corrections in three-dimensional brain positron emission tomography. *Med Phys* **30:** 937-948 (2003).
126. Dogdas B., Shattuck, D. W. and Leahy, R. M., "Segmentation of the skull in 3D human MR images using mathematical morphology" Proceedings of SPIE Medical Imaging Conference, San Diego, USA, Vol. 4684; *pp* 1553-1562 (2002).