Validation of segmentation techniques for positron emission tomodraphy using ex vivo images of oncological surgical specimens


A R T I C L E   I N F O

Article history:
Received 15 April 2013
Accepted 6 June 2013

Keywords:
Positron emission tomodraphy
Surgical specimen
Segmentation
Oncology

A B S T R A C T

Objective: To design a novel ex vivo acquisition technique to establish a common framework to validate different segmentation techniques for oncological PET images. To evaluate several automatic segmentation algorithms on this set of images.

Material and methods: In 15 patients with cancer, ex vivo PET studies of surgical specimens removed during surgery were performed after injection of 18F-FDG. Images were acquired in two scanners: a clinical PET/CT and a high-resolution PET scanner. Real tumor volume was determined in each patient, and a reference image was generated for segmentation of each tumor. Images were segmented with 12 automatic algorithms and with a standard method for PET (relative threshold at 42%) and results were evaluated by quantitative parameters.

Results: It has been possible to demonstrate by segmentation of PET images of surgical specimens that on high resolution PET images, 8 out of 12 evaluated segmentation techniques outperformed the standard method, whose value is 42%. However, none of the algorithms outperformed the standard method when applied on images from the clinical PET/CT. Due to the great interest of this set of PET images, all studies have been published on the Internet in order to provide a common framework for validation and comparison of different segmentation techniques.

Conclusions: We have proposed a novel technique to validate segmentation techniques for oncological PET images, acquiring ex vivo PET studies of surgical specimens. We have demonstrated the usefulness of this set of PET images by evaluating several automatic segmentation algorithms.

© 2013 Elsevier España, S.L. and SEMNIM. All rights reserved.

R E S U M E N

Objetivo: Diseñar una técnica novedosa de adquisición ex vivo para establecer un marco común de validación de diferentes técnicas de segmentación para imágenes PET oncológicas. Evaluar sobre estas imágenes el funcionamiento de varios algoritmos de segmentación automática.

Material y métodos: En 15 pacientes oncológicos se realizaron estudios PET ex vivo de las piezas quirúrgicas extraídas durante la cirugía, previa inyección de 18F-FDG, adquiriéndose imágenes en 2 tomógrafos: un PET/CT clínico y un tomógrafo PET de alta resolución. Se determinó el volumen tumoral real en cada paciente, generándose una imagen de referencia para la segmentación de cada tumor. Las imágenes se segmentaron con 12 algoritmos automáticos y con un método estándar para PET (umbral relativo del 42%) y se evaluaron los resultados mediante parámetros cuantitativos.

Resultados: La segmentación de imágenes PET de piezas quirúrgicas ha demostrado que para imágenes PET de alta resolución 8 de las 12 técnicas de segmentación evaluadas superan al método estándar del 42%. Sin embargo, ninguno de los algoritmos superó al método estándar en las imágenes procedentes del PET/CT clínico. Debido al gran interés de este conjunto de imágenes PET, todos los estudios se han publicado a través de Internet con el fin de servir de marco común de validación y comparación de diferentes técnicas de segmentación.
Introduction

Positron emission tomography (PET) is a diagnostic imaging technique which has demonstrated great utility in the field of oncology. There has recently been great interest in its potential for the planning of radiotherapy treatments. The most critical process for this planning is the precise delimitation of the tumor volume to be treated. At present the delimitation of tumor volume in routine clinical practice is performed manually on anatomic images such as computed tomography or magnetic resonance. However, manual delineation on PET images is quite complex due to the limited spatial resolution of this modality, producing an elevated intra- and inter-operator variability. Thus, in the case of PET a robust, automatic, standardized segmentation technique is required for the determination of tumor volume.

To date, the technique most used for segmenting PET images is the method of threshold value fundamentally due to its simplicity of implementation and application. The threshold is generally selected visually according to the criteria of the specialist in nuclear medicine. However, the selection of the optimal threshold is critical for correct delimitation.

In the past few years a large number of studies have been published in relation to the investigation of different automatic or semiautomatic segmentation techniques applicable for the detection of tumor volume on PET images. The techniques proposed range from the use of a fixed, relative or automatically selected threshold to segmentation using complex statistical models. Nonetheless, there is still no standardized method for this purpose.

For translation to clinical practice, all segmentation techniques must be validated. The purpose of validation is to verify that the method of delineation considered is applicable to a wide variety of cases with reasonable precision. The quality and congruence of the validation depend not only on the quality of the evaluation criteria but also on the group of images used. Indeed, the precision of automatic delineation is directly conditioned by the quality of the image. Thus, the protocol of image acquisition and reconstruction is as important as the delineation technique itself.

The greatest problem for the comparison of segmentation methods is the lack of a common database with images of quality for this purpose. In a recent study Dewalle-Vignion et al. described a series of PET images in patients with lymphoma for validation of segmentation techniques. The greatest limitation of this study was that the image of reference was based on manual segmentation over the PET image.

Within this framework, the present study proposes a novel technique of acquisition of high resolution PET images ex vivo to establish a common framework of validation of different segmentation techniques for oncologic PET images. The images obtained by this procedure are available on Internet with the objective of providing a common framework of validation of any segmentation technique. To demonstrate the utility of this series of images, we evaluated the functioning of several algorithms of segmentation based on the automatic selection of the optimal threshold.

Material and methods

Patients and protocol of acquisition

Oncologic patients who were to undergo surgery were studied after having obtained approval from the Ethical Committee of our institution. The series included 15 patients: 2 with breast cancer, 2 with colorectal cancer and 10 with cancer of the prostate. Each patient provided informed consent. Table 1 describes some of the demographic and clinical data of these patients.

A protocol of ex vivo acquisition was designed performing images of the resected surgical pieces according to the protocol shown in Fig. 1. The ex vivo PET study was carried out on the day of the surgery following the administration of 7 MBq/kg of 18F-FDG prior to the surgical removal of the tumor. The tracer was injected when approximately 60 min of remaining surgery were estimated. On resection of the tumor the surgical piece was transferred to the Department of Nuclear Medicine where 2 PET studies of each tumor were made: the first in the clinical Biograph Duo PET/CT equipment (Siemens, Hoffman Estates, IL, USA) and the second in a high resolution Mosaic PET (Philips, Milpitas, CA, USA) usually dedicated to the study of small animals.

In the Biograph Duo PET/CT the protocol consisted in one CT for correction of attenuation and one static PET image of 3 min in a single bed position. The image was reconstructed using the iterative algorithm OSEM with 2 iterations and 8 subgroups, a 5 mm Gaussian filter and corrections of attenuation and scatter. In the high

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Weight (kg)</th>
<th>Primary tumor</th>
<th>Histological tumor size (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>49</td>
<td>F</td>
<td>66</td>
<td>Breast</td>
<td>2.4</td>
</tr>
<tr>
<td>p4</td>
<td>81</td>
<td>M</td>
<td>70</td>
<td>Colorectal</td>
<td>6.6 x 5.5</td>
</tr>
<tr>
<td>p5</td>
<td>58</td>
<td>F</td>
<td>64</td>
<td>Colorectal</td>
<td>7.5 x 4</td>
</tr>
<tr>
<td>p6</td>
<td>58</td>
<td>M</td>
<td>76</td>
<td>Colorectal</td>
<td>5.5 x 2.2</td>
</tr>
<tr>
<td>p7</td>
<td>61</td>
<td>F</td>
<td>86</td>
<td>Colorectal</td>
<td>2.5 x 1.4</td>
</tr>
<tr>
<td>p8</td>
<td>58</td>
<td>M</td>
<td>63</td>
<td>Prostate</td>
<td>1.5</td>
</tr>
<tr>
<td>p12</td>
<td>61</td>
<td>M</td>
<td>102</td>
<td>Colorectal</td>
<td>7.5 x 5.5</td>
</tr>
<tr>
<td>p16</td>
<td>74</td>
<td>M</td>
<td>84</td>
<td>Prostate</td>
<td>0.9</td>
</tr>
<tr>
<td>p17</td>
<td>57</td>
<td>F</td>
<td>63</td>
<td>Colorectal</td>
<td>2.8 x 2.8</td>
</tr>
<tr>
<td>p18</td>
<td>68</td>
<td>M</td>
<td>67</td>
<td>Colorectal</td>
<td>4 x 3.8</td>
</tr>
<tr>
<td>p19</td>
<td>68</td>
<td>M</td>
<td>83</td>
<td>Colorectal</td>
<td>5 x 3</td>
</tr>
<tr>
<td>p20</td>
<td>78</td>
<td>F</td>
<td>54</td>
<td>Breast</td>
<td>2.5 x 2.5</td>
</tr>
<tr>
<td>p21</td>
<td>87</td>
<td>F</td>
<td>54</td>
<td>Breast</td>
<td>1.9</td>
</tr>
<tr>
<td>p22</td>
<td>78</td>
<td>M</td>
<td>74</td>
<td>Colorectal</td>
<td>5 x 4</td>
</tr>
<tr>
<td>p23</td>
<td>47</td>
<td>F</td>
<td>71</td>
<td>Colorectal</td>
<td>7 x 6</td>
</tr>
</tbody>
</table>

F: female; M: male.
resolution Mosiac PET, one study of transmission with a source of $^{137}$Cs followed by one 20-min emission study was performed. The standard iterative algorithm of this equipment (3 D Ramla with 1 mm Gaussian filter) was used for reconstruction, with corrections of dead time, random events, attenuation and scatter.

**Segmentation of the images**

Prior to segmenting the images 2 preprocessing operations were performed. First the images were cut. This cut is done because the field of vision of both tomographs is large in comparison with the size of the surgical piece and thus, most of the image remains empty. After cutting, the image is centered in the zone of interest and only contains the tumor and surrounding healthy tissue. In addition, all the new images were scaled so that the scale expressed SUV values instead of kBq/cm$^3$. To do this all the images were divided by the factor (activity administered/patient mass) expressed in kBq/g.

The ex vivo images of the surgical pieces obtained in both the PET Mosaic and Biograph Duo PET/CT equipment are then segmented. The images are segmented with 12 segmentation algorithms based on the automatic selection of the optimal threshold (Table 2) previously implemented by our group of investigation$^{11,17}$ available at: [https://www.cun.es/la-clinica/servicios-medicos/departamento/medicina-nuclear/pet-segmentation-tools-0](https://www.cun.es/la-clinica/servicios-medicos/departamento/medicina-nuclear/pet-segmentation-tools-0).

In addition to the 12 automatic segmentation algorithms proposed by Prieto et al.$^{11}$ all the images were segmented with a standard method for PET images. A relative threshold of 42% was specifically used with respect to the maximum value of the image.$^9$

**Evaluation of the results of segmentation**

For the evaluation of the segmentation results the volume of reference delimiting the real size of each tumor must be defined. To do this, the anatomopathological analysis performed in the surgical piece was used as the starting data including the tumor size in 1, 2 or 3 dimensions according to the geometry of the tumor. With these data a spherical volume of interest was drawn on the high resolution PET image (Mosiac PET) centered in the zone of maximum uptake and with a diameter corresponding to the maximum dimension of the tumor according to the histology. Consequently, the reference volume presented an irregular shape due to the isocontour and its dimensions coincided with the histological data.

The following quantitative measures which compare the image obtained with each segmentation algorithm with the reference image were used.$^{11,14}$

- **Volume error (VE):** this measures the difference between the reference and segmented volumes expressed as a percentage with respect to the reference volume.
- **Classification error (CE):** this measures the total quantity of voxels incorrectly assigned to an object or the background in the segmented image with respect to the number of voxels of the object in the reference image.
- **Dice similarity index (DSI):** this is a measure of similarity between 2 objects defined as the intersection between the 2 segmented and reference objects over the mean of the group combined.

To follow the same criteria in the 3 parameters 1-DSI was used instead of DSI. Thus, a low value in any of the 3 measurements represents better segmentation. Although the VE and CE may have values greater than 100%, in the present study the maximum error was 100%.

The segmented volumes from the Mosiac PET images were comparable voxel to voxel with the reference image, allowing the calculation of the 3 parameters previously defined (VE, CE and 1-DSI). The arithmetic mean of these 3 values was calculated to determine a measurement of overall error for each image.

In the case of the clinical PET/CT images, the segmented volume could not be compared voxel to voxel with the reference image because the images were not co-registered. Consequently, the CE and 1-DSI parameters could not be calculated and only the VE measurement was obtained, which was based on the comparison of volumes. Since the quantitative analysis was limited to a single parameter a visual analysis was carried out by the fusion of the segmented image with the reference volume but making the co-registry only with respect to the displacements and rotations without applying the sampling change.

The measurements of error obtained with the different segmentation algorithms were statistically analyzed and the normality of the variables was compared with the Shapiro–Wilk test. Due to the absence of criteria of normality the comparisons paired between methods were evaluated using the sign test. The analyses were carried out using the statistical package SPSS (version 11.0, Chicago, IL, USA). Significance was set at a $p$ value <0.05 for all the comparisons.

**Results**

**Segmentation of the images of the Mosaic PET equipment**

Fig. 2A shows a slice of each ex vivo PET image obtained in the high resolution tomograph, each case superimposing the outline.
PET Mosaic

Fig. 2. Slices representative of each of the 15 surgical specimens analyzed in this study. (A) Images acquired in the Mosaic PET tomograph; the white outlines show the extension of the tumor according to the reference image determined visually. (B) Images acquired in the Biograph Duo PET/CT tomograph.

of the reference volume selected visually by the expert in nuclear medicine. The images of the 15 surgical pieces acquired in the high resolution PET tomograph were segmented with the 12 automatic algorithms evaluated as well as with the standard method for PET images: threshold of 42% of the maximum value of the image. The results of the 3 parameters of evaluation are presented in Fig. 3A in which the median of the whole series of 15 images was calculated for each method.

The results of the statistical tests performed to compare each method proposed with the standard method are shown in Table 3 in which the methods are ordered according to the median of the mean overall error. Considering the threshold of significance as a p value <0.05, this analysis demonstrated that the results obtained with all the algorithms except Hertz showed statistically significant differences with respect to those obtained with the relative threshold of 42%, with Yanni, Tsai, Otsu, Ridler, Ramesh, Sahoo, Kapur and Yen being statistically better than the standard method. As an example of the good functioning of the methods described, Fig. 4 presents the segmented volume using the algorithm of Yanni in the PET images of the 15 surgical pieces obtained in the Mosaic high resolution tomograph.

Segmentation of the images of the Biograph Duo PET

Fig. 2B shows a slice of each of the PET images of the 15 surgical pieces acquired ex vivo in Biograph Duo PET/CT tomograph. In comparison with the images of the high resolution Mosaic tomograph, it is clearly shown that in this case the images are blurry, that is, they have a worse spatial resolution and do not present the noise associated with the low count statistics.

The images of the 15 surgical pieces acquired in the PET/CT Biograph Duo tomograph were segmented with the standard method (threshold of 42% of the maximum value of the image) and with the 12 automatic algorithms evaluated in this study. In this case the VE
Fig. 3. Comparison among the methods of the median of the measurements of error for the 15 images. The methods are ordered according to the categories defined in Table 2. (A) Data obtained on segmenting the images of the Mosaic PET tomograph. (B) Data obtained on segmenting the images of the Biograph Duo PET/CT tomograph.

Table 3

Methods ordered according to the median of the measurement of overall error (mean of VE, CE, 1-DSI) in the 15 images of the Mosaic tomograph and according to the median of the VE parameter in the 15 images of the Biograph Duo PET/CT tomo-
graph. p value of significance obtained on comparison with the standard method of 42%.

<table>
<thead>
<tr>
<th>Method</th>
<th>Median</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mosaic PET</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yanni</td>
<td>3.7</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Tsai</td>
<td>4.1</td>
<td>0.007*</td>
</tr>
<tr>
<td>Otsu</td>
<td>8.5</td>
<td>0.035*</td>
</tr>
<tr>
<td>Ramesh</td>
<td>8.5</td>
<td>0.007</td>
</tr>
<tr>
<td>Sahoo</td>
<td>11.5</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Kapur</td>
<td>11.9</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Yen</td>
<td>18.7</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>42%</td>
<td>41.8</td>
<td></td>
</tr>
<tr>
<td>Lloyd</td>
<td>52.8</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Huang</td>
<td>69.0</td>
<td>0.035*</td>
</tr>
<tr>
<td>Li</td>
<td>72.7</td>
<td>0.035*</td>
</tr>
<tr>
<td>Hertz</td>
<td>80.8</td>
<td>0.118</td>
</tr>
<tr>
<td><strong>Biograph Duo PET/CT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42%</td>
<td>18.8</td>
<td>–</td>
</tr>
<tr>
<td>Otsu</td>
<td>38.9</td>
<td>0.092</td>
</tr>
<tr>
<td>Ramesh</td>
<td>38.9</td>
<td>0.092</td>
</tr>
<tr>
<td>Ridler</td>
<td>39.4</td>
<td>0.092</td>
</tr>
<tr>
<td>Yanni</td>
<td>54.8</td>
<td>0.267</td>
</tr>
<tr>
<td>Lloyd</td>
<td>55.7</td>
<td>0.013*</td>
</tr>
<tr>
<td>Tsai</td>
<td>57.1</td>
<td>0.581</td>
</tr>
<tr>
<td>Kapur</td>
<td>64.1</td>
<td>0.006*</td>
</tr>
<tr>
<td>Hertz</td>
<td>65.1</td>
<td>0.180</td>
</tr>
<tr>
<td>Sahoo</td>
<td>85.9</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Yen</td>
<td>92.3</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Huang</td>
<td>100.0</td>
<td>0.022</td>
</tr>
<tr>
<td>Li</td>
<td>100.0</td>
<td>0.003</td>
</tr>
</tbody>
</table>

* Statistically significant differences.

was used as the only measurement of error, calculating its median for all the images (Fig. 3).

The results of the comparison of each automatic method with the standard method (42%) are shown in Table 3. The standard method was found to be the best method and the only one with a median of less than 20%. The sign statistical test showed that despite obtaining worse results the algorithms of Otsu, Ramesh, Ridler, Yanni and Tsai did not present statistically significant differences compared to the reference method, with the median of error for the first 3 methods being less than 40%. The remaining methods (Lloyd, Kapur, Hertz, Sahoo, Yen, Huang and Li) presented significantly worse functioning than the reference method. In general, all the methods showed a worse behavior compared to the results obtained with the Mosaic PET. Additionally, detailed visual analysis of all the images was made. The visual analysis was centered on the best algorithms according to the statistical analysis: Otsu, Ridler and Ramesh. Fig. 5 shows the images corresponding to 3 of the 15 surgical pieces corresponding to one image with good results (p01), one with intermediate results (p04) and one with deficient results (p20). All the images acquired in this study, together with the reference image in each case may be seen at the following website: https://www.cun.es/la-clinica/servicios-medicos/departamento/medicina-nuclear/pet-segmentation-tools-0.

**Discussion**

Methodology of evaluation of the segmentation techniques based on the high resolution PET tomograph in surgical specimens.

A unique contribution of the present study is the inclusion of data from PET of real human tumors. Attainment of data of patients for segmentation usually leads to many economic, logistic and even ethical questions15 and thus, most studies work with simulated lesions. These simulations may be generated directly by the convolution of fictitious tumors with the function of response of the PET system, posteriorly adding a certain level of noise to model its effect
on the images. A more purist method of artificial image generation consists in the simulation of objects and of the PET tomograph with the Monte Carlo method. In any case, not real images are obtained with these 2 methods (convolution with the response of the system or Monte Carlo).

Ideal validation should include real data of patients as in the present study. Once the PET images have been obtained the greatest difficulty lies in generating a reference image apt for verifying the segmentation results. There are 2 possible ways to do this: acquire images with another imaging modality such as computerized tomography or surgical extraction of the tumor and analyze its characteristics anatomopathologically. Both options present important limitations.

In the first option, on comparing the data with another imaging modality the results depend on the positioning of the patient and/or the goodness of registry among the images. On the other hand, there is no guarantee that the alternative modality provides absolutely reliable information regarding the precise delimitation of the tumor since all the modalities are subject to limitations. In addition, comparison with another imaging method would be based on the questionable hypothesis that the contours of the tumor are equal in the 2 methods. However, the interest in PET lies in that it is able to detect small changes in cellular metabolism which are not detected with any other medical imaging technique.

With regard to the surgically removed tumor specimens, there are also numerous, varied difficulties. This type of information has been managed in the literature with the registry between the slices of the macroscopic surgical specimen and the presurgical PET images. This is a complex procedure which consists in freezing the surgical specimen, cutting it into fine slices, obtaining a digital image of each slice and reconstructing a 3D volume from the images of the slices. This methodology was developed by Daisne et al. for pharyngolaryngeal tumors and by Stroom et al. for lung tumors and presents the limitation that the surgical specimen may become deformed or lose volume during freezing, the scanning and the registry. In addition, the contours are drawn manually on the surgical specimen and are therefore subject to important variability.

In the present study a completely novel technique has been proposed for obtaining the 3D reference volume. Instead of freezing the whole specimen, cutting it, digitalizing the slices and stacking up the images, a high resolution PET image is proposed with a threshold visually selected on this image. This methodology precisely obtains the volume of the tumor without being affected by possible changes due to the complex processing of the macroscopic piece proposed by Daisne et al. One limitation of the technique proposed lies in that regional nontumoral uptake may be produced in the image due to manipulation during surgery after injection of the radiotracer.

Fig. 4. Volumes segmented using the algorithm of Yanni in the images of the 15 surgical specimens acquired in the Mosaic PET tomograph. The segmented volume is shown as a binary image and the red outline indicates the reference volume selected visually by an expert in nuclear medicine.

Fig. 5. Example of the segmentation of 3 surgical specimens. The high resolution PET image, the PET image of the Biograph Duo with and without interpolation and the results of the segmentation with the best methods (Otsu, Ridler, Ramesh and the threshold of 42%) are shown. The red outline represents the reference volume obtained from the high resolution image.
Evaluation of the algorithms of automatic segmentation

As mentioned previously, manual delimitation of tumor volume on a PET image is more complex than delimitation on anatomical images mainly due to their worse spatial resolution. The methods described in the literature as an alternative to manual delimitation are mainly based on establishing a threshold according to the SUV value whether by using a fixed threshold (SUV 2.5) or adaptively selecting a threshold related to the maximum value of tumor uptake (42%). These are two limiting factors which can be adjusted in each image based on the calibration curves or iteratively. All these options present some drawbacks. The most currently used semiautomatic technique is the establishment of a threshold based on visual criteria. However, the selection of the optimal threshold is critical for correct delimitation. Thresholds based on a fixed absolute or relative SUV value constitute completely automatic techniques, but they only work in determined conditions. Although some adaptive methods calculate the threshold for each image based on the characteristics of the object to be segmented, these methods are supervised, that is, they are not automatic. On the other hand, automatic segmentation techniques have been proposed based on the calculation of gradients, the Bayesian theory, or the theory of probabilities. However, these methods are generally complex, require lengthy mathematical equations or some type of pre- or post-processing for correct functioning.

Nonsupervised and nonparametrized automatic techniques were evaluated in the present study. These techniques obtain the threshold from some characteristic of the image which may be automatically extracted. Since the threshold value method is currently most frequently used, the physician is able to perfectly understand the fundamentals of this type of segmentation so that this technique may be introduced in their clinical practice, understanding the subjacent mechanism. In addition, segmentation based on the automatic selection of an optimum threshold is the simplest method of all the automatic alternatives proposed in the literature. Indeed, these techniques are computationally very rapid and do not require any type of pre- or post-processing for functioning.

These algorithms of automatic segmentation have been applied in a group of tumor images acquired in vivo in 2 different PET tomographs using the tumor delimited by an expert on the high resolution PET image as the reference. The quality of segmentation has been compared with a standard segmentation method in PET (relative threshold of 42%). The results may be summarized as follows:

- In the images acquired in the high resolution PET tomograph the segmentation methods described by Yanni, Tsai, Otsu, Rid-lér and Ramesh presented medians in the measurement of the overall error of less than 10% while the methods of Sahoo, Kapur and Yen presented medians between 10 and 20%. All these algorithms were significantly better than the threshold of 42% (median: 41.8%).
- In the images acquired in the clinical PET/CT, none of the automatic segmentation methods evaluated surpassed the reference method, which showed a median of VE of less than 20%. The best methods with medians less than 40% and without statistically significant differences compared to the threshold of 42% were the algorithms of Otsu, Ramesh and Ridler.

Thus, very different results have been obtained in the two tomographs. The intrinsic characteristics of the images acquired in the Biograph Duo PET/CT with regard to sampling and spatial resolution make their segmentation particularly complex. These 2 phenomena are clearly demonstrated in Fig. 5. With respect to the sampling, the voxel size in this tomograph is of 5.31 mm × 5.31 mm × 3.38 mm thereby impeding very precise determination of the shape of the tumor. The other important factor is the limited spatial resolution of the PET/CT tomograph (9.3 mm) which makes the tumor produce an important effect of partial volume (PVE) so that the tumor appears to have a considerably larger than real size. Both the sampling and PVE affect all the images but the latter is especially severe in tumors of small dimension or those presenting central zones without uptake (p04 and p20 in Fig. 5). These 2 limiting factors contribute to the VE having relatively high values in comparison with those observed in the series of ex vivo images from the Mosaic PET. However, the 2 factors are due to the intrinsic characteristics of the images acquired in this tomograph and not to errors in the segmentation using the methods proposed in this study.

These results suggest that segmentation based on automatic selection of a threshold may be of limited applicability for clinical tomographs with similar characteristics of resolution to those of the PET/CT tomograph used in this study. However, it may be a very useful tool for segmenting tumors in PET images acquired with the new generation tomographs. These are characterized by an important improvement in spatial resolution thanks to the time-of-flight and modeling techniques of the function of point scatter together with an increase in matrix size in the reconstruction, being adequate for clinical practice given the improvement in the capacity of computation of the processing stations. These characteristics draw the benefits of the new generation PET equipment closer to those of the high resolution Mosaic tomograph for which the algorithms of Otsu, Ramesh and Ridler have demonstrated their efficacy in this study.

Conclusions

We have proposed a novel technique of validation of segmentation techniques for oncologic PET images, acquiring ex vivo PET studies of surgical pieces and obtaining the size of the tumor for evaluation of the results of the segmentation using a high resolution PET image. The images obtained with this methodology are available to the scientific community through Internet.

Automatic segmentation of PET images of the surgical pieces acquired in the high resolution tomograph have demonstrated that the algorithms of Yanni, Tsai, Otsu, Ridler and Ramesh are clearly better than the standard method (relative threshold of 42%). However, no algorithm surpassed the standard technique in images of the surgical pieces obtained in the clinical PET/CT.

Funding

The development of this investigation was partially supported by the Instituto de Salud Carlos III (project FIS PI061261) and by Siemens Health Care.

Conflict of interests

The authors declare no conflict of interest.

References


