Implementation of cluster analysis in 3D dosimetry for targeted radionuclide therapy

Antigoni Divoli, Susan Buckley, Manuel Bardiés, and Glenn D. Flux

Abstract– **Dosimetry for targeted radionuclide therapy has evolved in recent years and current research is focusing on the production of spatial dose distributions using 3D voxel-based approaches. In this work we propose the use of cluster analysis for 3D dosimetric applications as an automatic way to identify non-homogeneities by grouping voxels within a volume of interest (VOI) according to their functionality. We implemented *k-means* methodology with IDL and we applied it to 3D and to 4D simulated and patient data. The results showed that the proposed methodology could recover the non-homogenous regions, place automatically sub-VOIs over a VOI, and produce cumulated activity (CA) maps directly by fitting the centroids of the clusters or by a second step process using the clustered image as a map. An additional benefit of the proposed methodology could be its use to identify regions of excess noise originating especially from misregistration. An assessment of the full benefits and limitations of the proposed methodology remains to be done.**

I. Introduction

Dosimetry for targeted radionuclide therapy (TRT) has evolved in the last two decades and is becoming a routine procedure in many centres world-wide. Image-based methodologies have been used to calculate mean absorbed doses (AD) to organs or tumors. For the calculation the cumulated activity (CA) within a source volume is needed which represents the total number of decays occurring in the source after the administration of the radionuclide. Given the CA the AD is calculated either according to the MIRD scheme [1] by multiplication with a pre-calculated dose conversion factor, or by Monte-Carlo simulations where particles are followed as they interact with matter and they deposit energy in tissues. For the estimation of the CA a series of nuclear medicine images over several days is obtained, the time-activity curve for each source volume is defined and the integral of this curve to infinity is calculated to give the CA. Non-uniform uptake is often observed in volumes of interest (VOIs) and current research is focusing on the production of spatial dose distributions either by identifying sub-VOIs or by using 3D voxel-based approaches. For the latter, the CA of each voxel is estimated and given a resulting CA map a dose map is calculated by convolution with a dose point kernel or by Monte-Carlo techniques.

A typical value of a reconstructed image voxel is 4 mm but the spatial resolution of PET and SPECT modalities is well above this value, therefore, the accuracy of the counts of a single voxel is questionable, and in addition the statistical quality of the data decreases as the radioisotope decays. Furthermore, nuclear medicine images obtained in the duration of several days can be subject to residual misregistration even if alignment has been performed. The above limitations can result in inaccurate CA maps and hence dose maps if the direct voxel-by-voxel approach is used. It might be advantageous therefore to work with group of voxels, clusters, rather than single voxels. However, if these clusters have been defined by a user the selection is prone to user susceptibility.

The motivation of this work was to implement an automatic, user-independent approach to segment a VOI to sub-VOIs based on their functionality. *k-means* cluster analysis [2] was used for this purpose which is a clustering technique that starts with a number of random clusters and then iteratively moves items between clusters minimizing variability within each cluster and maximizing variability between clusters. The classification is based on the centroids (mean values) of the clusters. The implementation of the methodology and preliminary results on simulated and patient data are described below.

II. Methods and Results

A. Implementation of the methodology

We incorporated *k-means* clustering methodology in home-written software for image analysis and display in IDL® language utilising existing functions provided by the software. Given a two dimensional matrix where the columns represent the number of variables and the rows the number of measurements, the cluster centres are first calculated starting from a random hypotheses for their value. In our study the number of variables are the individual voxels, therefore, the original data either 3D representing a single slice with multi measurements, or 4D representing a volume with multi measurements, were first rearranged into a 2D matrix. Fig. 1 shows a schematic representation of the methodology; for the demonstration a given slice with multiple measurements at n time points is shown, but the approach is identical for a 3D volume with multiple measurements, i.e. for a 4D data set.

This work was supported by the National Translational Cancer Research Network grant C21768/A7021 funded jointly by the Department of Health, UK and the Institute National de Cancer, France, and managed by Cancer Research UK.

A. Divoli is with the Institute of Cancer Research, Royal Marsden NHS Foundation trust, SM2 5PT UK (telephone: +44(0)2086613713, e-mail: Antigoni.Divoli@icr.ac.uk).

S. Buchley was with the Institute of Cancer Research, Royal Marsden NHS Foundation trust, SM2 5PT UK. She is now with the Singleton NHS Trust, SA2 8QA, UK (telephone: +44(0)1792205666, e-mail: Susan.Buckley2@wales.nhs.uk).

M. Bardiés is with the INSERM U982, Nantes, F-44093 France, (telephone: +33(0)240412821, e-mail: manu@nantes.inserm.fr).

G. D. Flux is with the Institute of Cancer Research, Royal Marsden NHS Foundation trust, SM2 5PT UK (telephone: +44(0)2086613485, e-mail: Glenn.Flux@icr.ac.uk).

The red and green regions represent regions of different activity uptake as they decay oven the n measurements.

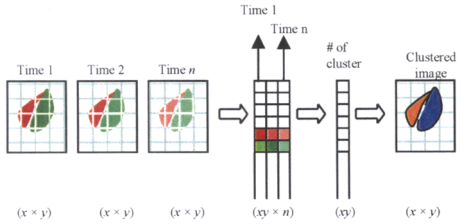

Fig. 1. Schematic representation of the implementation of the methodology for a given slice with n measurements over time.

Each transaxial slice of $(x \times y)$ dimensions is reformed into a 1D matrix of (xy) dimensions and the n slices are stuck together to form a 2D matrix of $(xy \times n)$ dimensions. Given the number of clusters *a priory* the *k-means* clustering algorithm is applied to the two dimensional matrix and the centroids of each cluster are calculated. Each row is classified into its corresponding cluster resulting into a 1D matrix of (xy) dimensions, which is then reformed into a 2D matrix of the original dimensions of each slice $(x \times y)$. The resulting clustered image can be used to map the original slices in order to recover time activity curves of the sub-VOIs or the centroids of each cluster can be used directly since if the centroids are plotted against the acquisition time points they are the desirable time activity curves. The latter assumes that the counts have been converted into activity by an appropriate calibration factor.

For a 3D volume with multiple measurements over time, i.e. for a data set of $(x \times y \times z \times n)$ dimensions the approach is identical with the reformed 2D matrix being of $(xyz \times n)$ dimensions, the resulting 1D matrix of (xyz) dimensions, and finally the resulting clustered image of $(x \times y \times z)$ dimensions.

B. *Application on simulated data*

The right kidney of the NCAT voxelised phantom [3] was used with 4 mm side cubic voxel. 6 decaying time points were simulated based on monoexponential decay for all but one slice in which an inhomogeneity was included which followed the decay of a different monoexponential. Fig. 2 shows plot of the simulated TACs for the 2 regions (indicated by solid lines), and also the centroids of the identified clusters (indicated by dotted lines). It is shown that the 2 exponentials are accurately recovered while two extra curves are shown which describe the edges that had been simulated so as to include edge effects, and a fifth curve lies on the x-axis describing the zero values surrounding the kidney. These results were obtained with 5 clusters *a priory*, with one pass to the whole volume, i.e. to a 4D data set.

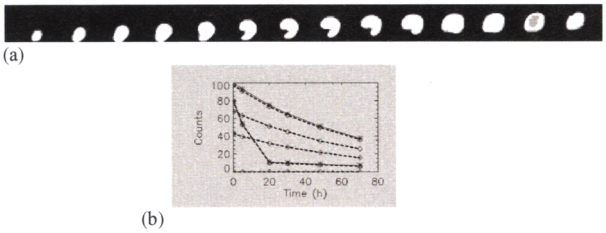

Fig. 2. (a) The kidney volume indicating the non homogeneity. (b) The centroids of the identified clusters in dotted lines recovering accurately the simulated curves shown by solid lines. Two extra curves are shown which describe the edges of the organ.

Artificial misregistration was induced in the slice containing the non-uniformity and the methodology was applied again to that slice only. Fig. 3 shows plot of the centoids of the defined clusters. It can be seen from the plot that the simulated curves were not recovered, and that the centroid curves would have no physical meaning as they would indicate excreted activity followed by uptake. These 'odd' time activity curves can appear as result of misregistration as it has been discussed by several groups [4, 5].

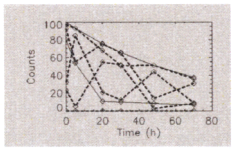

Fig. 3. The centroids of the identified clusters in dotted lines for the artificially misregistered slice indicate decay followed by sharp uptake which would have no physical meaning.

Poisson noise was added to each voxel by drawing random numbers from a Poisson distribution with mean value the original voxel value. Four different noise levels (N1-N4) were added to the data by multiplying the added noise with a where a=1-4. Fig. 4a shows the clustered images when the methodology was applied to the whole volume (4D data) for the noise N1 and N4. It is shown that the non-homogeneity is clearly recovered in both cases, but the clustered image of the volume of N4 level of noise is also noisy. For that case slice-by-slice analysis should be performed in addition or pre- or post-processing such as smoothing could be applied. Fig. 5 shows the resulted clustered images for the 4 different levels of noise studied for the non-uniform slice.

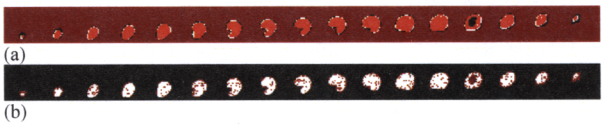

Fig. 4. The resulted clustered image when the *k-means* methodology was applied to the kidney volume subject to the Poisson noise a) noise level N1, and b) noise level N4.

Fig. 5. The resulted clustered images for the non-homogenous slice the 4 noise levels studied N1-N4 (from left to right). The non-homogeneity is detected in all cases, but at N4 noise level, the accuracy of the segmentation is poor.

C. *Application on patient data*

The methodology was applied to an abdominal tumor of a neuroblastoma patient treated with ^{131}I-mIBG (Meta-Iodo-Benzyl-Guanidine). 4 SPECT scans were available at 54 h, 125 h, 149 h, and 197 h after administration of 5000 MBq of ^{131}I-mIBG. The images were reconstructed and corrected based on a standardised clinical protocol [6], and, for the purposes of this work it was assumed that they described accurately the activity distribution within the volume. The tumour was isolated from the rest of the image, and a 30% of the maximum voxel count threshold was applied in order to reduce noise. The resulted volume of the tumour was 90 cc described by 14 slices. The methodology was applied to the whole volume and to each slice individually and there were either 1 or 2 clusters identified for each slice. Due to the image being smoothed the results were similar when the methodology was applied to the 4D data of the whole volume or to the 3D data of each slice. Fig.6 shows the transaxial slice of the maximum counts for the first time point and in the same figure the clustered image of the slice is shown.

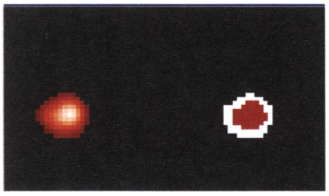

Fig. 6. Transaxial slice through the maximum count voxel for the first time point (left-hand side) and the clustered image (right-hand side).

Fig. 7 shows the resulting centroids when the methodology was applied to the whole tumor volume. The maximum counts over time are shown by solid lines and the centroids of the identified clusters by dotted lines. These were converted into time activity curves by using an appropriate calibration factor and they were then integrated to give the CA for the two identified sub-VOIs.

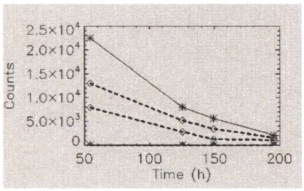

Fig. 7. The centroids of the two identified clusters for the tumour volume shown by dotted lines, and the maximume voxel counts shown by solid lines.

D. *Implementation in 3-D dosimetry*

For the patient tumour, AD maps were produced using the MCNPX 2.5.0 Monte-Carlo code [7] via a tool for personalized dosimetry called OEDIPE [8]. Given the CA, the energy deposited in each voxel was calculated using the full spectrum of ^{131}I and cumulated dose volume histograms (DVHs) were obtained from the resulted dose maps. Calculations based on 3 CA maps were performed: a) a uniform CA map, b) a CA map obtained by fitting each individual voxel's time activity curve, and c) a CA map obtained from the sub-VOIs defined by *k-means* clustering. The total of the cumulated activity was kept the same in all three cases to allow comparison of the distributions. Fig. 8 shows the 3 DVHs obtained. Assuming uniform cumulated activity indicated a relatively uniform dose deposition while the non uniform CA maps either as defined by clustering or by voxel-by-voxel fitting indicated non uniform dose distributions.

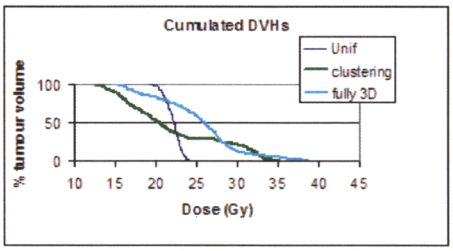

Fig. 8. Cumulated DVHs of dose maps obtained assuming uniform CA (Unif), CA calculated by fitting of the time activity curves of each voxel (fully 3D), and by using the clustered image (clustering).

III. Discussion

Cluster analysis has been used in many areas of science, such as astronomy, biology, psychiatry, and nuclear medicine; *k-means* algorithms in particular have been used mainly in dynamic PET cardiac and brain studies [9]. In this work we propose it for applications in 3D dosimetry for targeted radionuclide therapy. The methodology was implemented in IDL to perform to both 4D and 3D data sets representing a series of images of a volume or a slice respectively.

The initial results on simulated and patient data were encouraging; sub-VOIs were automatically defined even in the presence of noise. The automatic calculation of the CA maps is in principle possible simply by fitting the resulting centroids, but as the methodology is susceptible to noise a second step processing is advisable by using the clustered image as a map. In addition the proposed methodology could be used to identify areas of misregistration prior to applying 3D voxel-by-voxel dosimetry. A hypothesis would be that the methodology could be used to correct for residual misregistration and/or to assist the process of registration.

Dose maps and the respective cumulated DVHs were calculated for the patient tumour based on CA maps obtained by voxel-by-voxel fitting, and by clustering. Both showed non-homogenous AD deposition unlike it was the case when the CA activity was assumed uniform. There was not significant difference in the calculation time between the three cases studied, but this could change for a different radioisotope, for example for the more energetic beta particles of ^{90}Y.

Two inherent weaknesses of the *k-means* cluster analysis are: that the number of clusters even though an unknown has

to be given *a priory*, and that it starts with a random assumption on the values of the centroids. If the original assumption on the centoids is far from being true, the algorithm fails. Nevertheless the algorithm is very fast (seconds to minutes depending on the size of the original data) and these are not significant limitations since one can perform the analysis several times over the period of a few minutes if necessary in order to gain confidence. Furthermore, the exact number of clusters might not be as crucial for tumour dosimetry. For an organ might be known, as it is the case for the uptake of peptides in the kidney where two distinct regions have been observed. Nevertheless, other clustering algorithms could be investigated that might be more robust. The full benefits and limitations of the proposed methodology need to be assessed.

IV. Conclusions

Cluster analysis can be used in targeted radionuclide therapy for dosimetric calculations as an automatic method to identify inhomogeneities, place sub-VOIs over them, and produce CA maps. Furthermore it can be used as a tool for identifying residual misregistration and regions of excess noise. Further investigation is needed to explore the full benefits of the proposed methodology.

Acknowledgment

The authors would like to thank Antonello Spinelli and Sara Domenichelli (Nuclear Medicine, Azienda Ospedaliera di Bologna Policlinico S.Orsola-Malpighi, Bologna, Italy) for useful discussion at the initial stage of this project.

References

[1] R. Loevinger, T. F. Budinger, and E. E. Watson, "MIRD primer for absorbed dose calculations", *Revised Edition New York: Society of Nuclear Medicine.* 1991.

[2] B. S. Everitt, *Cluster Analysis*, New York: Halsted Press, 1993, pp.99-102.

[3] http://www.bme.unc.edu/~wsegars/index.html (8/5/08).

[4] S. J. Wilderman, A. M. Avram, J. Kritzman, R. Ackerman, and Y. Deqaraja, "Dosimetry in ^{131}I internal emitter therapy using voxel dependent integrated time-activities derived from multiple, registered SPECT and CT images", *Conference record (CD-ROM) of the IEEE 2006 Nuclear Science Symposium and MIC (San Diego, CA).*

[5] P. Papavasileiou, A. Divoli, K. Hatziioannou, and G. D. Flux, "The importance of the accuracy of image registration of SPECT images for 3D targeted radionuclide therapy dosimetry", *Phys., Med., Biol.*, vol. 52, N539-N548, 2007.

[6] S. E. Buckley, F. H. Saran, N. M. Gaze, et. Al., "Dosimetry for fractionated I-131-mIBG therapies in patients with primary resistant high-risk neuroblastoma: preliminary results", *Cancer Bio. Radiopharma.*, vol. 22, no. 1, pp, 105-112, 2007.

[7] J. S. Hendricks. MCNPX, version 2.5.e. *LA-UR-04-0569.* 2004.

[8] S. Chiavassa, I. Aubineau-Laniece, A. Bitar, et al. "Validation of a personalized dosimetric evaluation tool (Oedipe) for targeted radiotherapy based on the Monte Carlo MCNPX code", *Phys Med Biol.* vol. 51, pp. 601-616, 2006.

[9] K. P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Segmentation of dynamic PET images using cluster analysis", *IEEE Trans Nucl Sci*, vol. 49, no. 1, pp. 200-207, 2002.