Quantification of Tumor Changes during Neoadjuvant Chemotherapy with Longitudinal Breast DCE-MRI Registration

Jia Wu[1], Yangming Ou[2], Susan P. Weinstein[1], Emily F. Conant[1], Ning Yu[3], Vahid Hoshmand[3], Brad Keller[1], Ahmed B. Ashraf[1], Mark Rosen[1], Angela DeMichele[1], Christos Davatzikos[1], Despina Kontos[1,*]

[1]Department of Radiology, University of Pennsylvania
[2]Massachusetts General Hospital, Harvard University
[3]School of Engineering and Applied Sciences, University of Pennsylvania
*Contact Author - em: Despina.Kontos@uphs.upenn.edu // ph: 215-615-0827 // fx: 215-614-0266

ABSTRACT

Imaging plays a central role in the evaluation of breast tumor response to neoadjuvant chemotherapy. Image-based assessment of tumor change via deformable registration is a powerful, quantitative method potentially to explore novel information of tumor heterogeneity, structure, function, and treatment response. In this study, we continued a previous pilot study to further validate the feasibility of an open source deformable registration algorithm DRAMMS developed within our group as a means to analyze spatio-temporal tumor changes for a set of 14 patients with DCE-MR imaging. Two experienced breast imaging radiologists marked landmarks according to their anatomical meaning on image sets acquired before and during chemotherapy. Yet, chemotherapy remarkably changed the anatomical structure of both tumor and normal breast tissue, leading to significant discrepancies between both raters for landmarks in certain areas. Therefore, we proposed a novel method to grade the manually denoted landmarks into different challenge levels based on the inter-rater agreement, where a high level indicates significant discrepancies and considerable amounts of anatomical structure changes, which would indeed impose giant problem for the following registration algorithm. It is interesting to observe that DRAMMS performed in a similar manner as the human raters: landmark errors increased as inter-rater differences rose. Among all selected six deformable registration algorithms, DRAMMS achieves the highest overall accuracy, which is around 5.5 mm, while the average difference between human raters is 3 mm. Moreover, DRAMMS performed consistently well within both tumor and normal tissue regions. Lastly, we comprehensively tuned the fundamental parameters of DRAMMS to better understand DRAMMS to guide similar works in the future. Overall, we further validated that DRAMMS is a powerful registration tool to accurately quantify tumor changes and potentially predict early tumor response to chemotherapy. Therefore, future studies that aim at examining if DRAMMS can generate valuable biomarkers for tumor response prediction during chemotherapy become feasible.

Keywords: Deformable Registration, Breast Tumor, DCE-MRI, Tumor Response, Neoadjuvant Chemotherapy

1. INTRODUCTION

Breast cancer is the second most frequent cancer and the second leading cause of death for women in the United States[1]. A patient receives neoadjuvant chemotherapy for breast cancer to shrink a tumor for two main reasons[2]: 1) allowing surgically removal of the tumor initially inoperable at pre-chemotherapy stage [3, 4], and 2) allowing breast-conserving surgery rather than mastectomy [5, 6]. In order to optimize clinical treatment outcomes, it is critical to effectively monitor tumor response during neoadjuvant chemotherapy as well as early predict long-term pathologic response from a series of longitudinal breast magnetic resonance images[7, 8]. Image registration algorithms [9, 10] that aim to build voxel-wise anatomical correspondences, are able to automatically and efficiently quantify both tumor and normal breast tissue local changes over time, and therefore provide a potentially more attractive pathway (known as parametric response map[11-13]) to uncover tumor response[12, 14-16], in comparison to current global measurements such as RECIST [17] and World Health Organization (WHO) [18] and more [19]. However, since there are considerable amounts of anatomical changes brought in by chemotherapy, it becomes challenge to identify these correspondences in imaging spaces. Therefore, it is necessary to determine the most suitable registration algorithm among the available algorithms specifically for breast DCE-MR imaging.

Previously, our group has performed a pioneering study to compare our in-house algorithm DRAMMS[20] with five other deformable registration algorithms for longitudinal breast image registration, in general with DRAMMS giving the highest registration accuracy [21]. In this study, we intend to further evaluate and validate this open source registration

package DRAMMS from our group. The main contributions of this paper are two-fold: 1) we have proposed new ways to assess and more deeply understand the performance of DRAMMS within the imaging subregions (note the whole breast region are labeled with various registration challenge levels based on inter-rater disagreement and 2) comprehensively tune the parameters of DRAMMS to optimize the configuration for longitudinal breast DCE-MR imaging registration application, which will guide future large population studies.

2. METHODS

2.1 Study population and image dataset

Longitudinal breast DCE-MR images for 14 women with biopsy-proven T2-3 breast tumors were retrospectively collected from the ACRIN 6657 I-SPY trial (see details in [7]). All women in this trial underwent standard neoadjuvant chemotherapy, which at the time of the study consisted of four cycles of Adriamycin/Cytoxan, followed by four cycles of Taxotere. At the end of the 3-4 months of chemotherapy, the women were evaluated by pathological analysis and classified as either pathologic complete responder (pCR) or pathologic partial responder (pPR). In this study, seven women were classified as pCR and seven as pPR.

MR imaging parameters were: FOV 18-20 cm, image size 256×256×64,voxel size 0.70×0.70×2.0mm^3, TR=27.0ms, TE=4.76ms, flip angle=45^o. Pre-gadodiamide imaging was performed followed by immediate post-gadolinium images (at two minutes) and delayed post-gadodiamide images (at seven minutes). The baseline image was the image obtained right before the first chemotherapy, and final images were acquired after therapy was completed. Note, in this study not all four image sets were investigated, and only the image sets obtained before chemotherapy (1st imaging) and after the second treatment chemotherapy cycle (3rd imaging) were studied.

2.2 Deformable registration algorithms

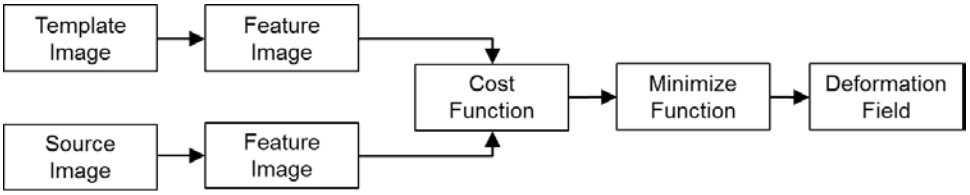


Fig. 1. A general descriptive pipeline of typical deformable registration algorithm

Two other open source software packages (five algorithms) which were previously applied to register longitudinal breast MR images were selected for longitudinal breast MRI registration, including Demons [22] and FFD [23] for comparison purposes. In general, each deformable registration algorithm aims to find the optimal deformation field that warps the source image to the template image through minimization of a specified cost function (measure the spatial difference) as shown in Fig. 1. Note in this study, the template image is always the image acquired before chemotherapy, and the source image is always the image obtained after chemotherapy. Each deformable registration algorithm is typically defined by three essential components: 1) voxel characterization, 2) similarity metric and 3) transformation model. The details for the selected six algorithms of three packages are listed in Table 1, and the detailed parameter setting for the selected algorithms can be referred in [21]. Note, in order to carry out a fair comparison, the parameters for all the studied software were optimally adjusted according to the given imaging set (see more details in [21]).

Table 1. Three principal components (voxel characterization, similarity metric and transformation model) of three deformable registration packages, including Demons, FFD, and DRAMMS.

	Voxel Characterization	***Similarity Metric***	***Transformation Model***
Demons	Intensity	Intensity Difference	Optical Flow[24]/Diffeomorphism
FFD	Intensity	1) Normalized Mutual Information 2) Correlation Coefficient 3) Sum of Squared Differences	Cubic B-spline
DRAMMS	Gabor Texture[25]	Texture Difference	Cubic B-spline

2.3 Evaluation criteria

2.3.1 Human rater

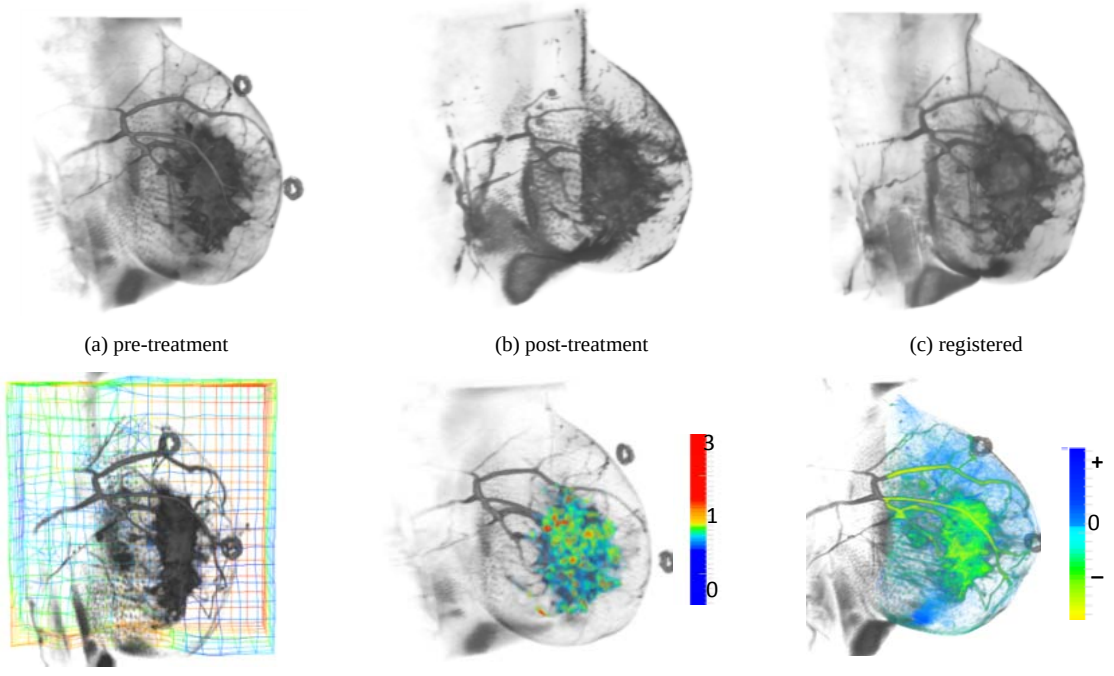
To quantitatively measure registration output, two experienced breast imaging radiologists separately denoted landmarks according to their anatomical meaning for both the template and source images while blinded to the patient response outcome. Both radiologists initially agreed on a collective set of 2380 anatomical landmarks in the template image set, which were scattered in the whole breast space (includes nipples, breast boundaries, chest walls, internal milk ducts, vessels, glandular structure and tumor). Then they independently identified corresponding landmarks in the source image for each patient.

2.3.2 Registration challenge from landmark grading

Since neoadjuvant chemotherapy significantly affects both tumor and normal breast tissue structures, it is very difficult to identify the corresponding landmarks based on the template image between two raters. Moreover, corresponding landmarks from the template image may not even exist in the source image since chemotherapy causes noticeable shrinkage within a particular tumor region. Before directly utilizing these landmarks as a basis for future evaluation, we proposed a novel idea to grade each landmark at different registration challenge areas, based on the value of landmark variation between the two raters. The details are as following: if there is no variation between two raters for the given landmark, then this landmark is marked as level 0; if the variation $\in \big((k-1)*d, k*d\big]$, $k = 1,2,3$ ..., where d is the longest cube diagonal of the image voxel, this landmark is marked as level k. A higher level landmark indicates larger variation between the two raters and great anatomical structure change. Thus the level of the landmark correlates to the challenge for each registration algorithm. In general, the high level landmarks are distributed across the whole region of breast, which indicates the chemotherapy significantly alternates both tumor and normal breast tissue regions. Thus, we are curious to evaluate the performance of the selected registration algorithm DRAMMS at various challenge levels. For further calculation purposes, these high level landmark regions should be excluded to perform an accurate comparison.

3. RESULTS

3.1 Registration output from DRAMMS



(a) pre-treatment (b) post-treatment (c) registered

(d) displacement (e) Jacobian (f) residual

Fig. 2. Jacobian determinants and intensity residual colormap from registering (c) a post-treatment (b) to a pre-treatment scan (a) for a selected breast cancer patient, using DRAMMS, showing (e) tumor shrinkage as response to neoadjuvant chemotherapy, but also (f) residual intensity decrease.

Fig. 2 displays the registration results of one selected patient image set from DRAMMS with the template image as the image acquired before chemotherapy (pre-treatment), the source image as the image obtained after chemotherapy (post-treatment), and the registered image being the deformed source image produced from DRAMMS (with the preselected optimal parameters g = 0.2). Anatomical variations between pre-treatment and post-treatment breast structures within the two images are visually noticeable, due to the chemotherapy. After registration, the aligned post-treatment image (Fig. 2c) is deformed in a way more similar to the pre-treatment image in comparison with post-treatment image, which qualitatively indicates the effectiveness of the deformable registration algorithm DRAMMS. The displacement field is generated from the deformable registration algorithm DRAMMS at each image voxel as shown by the mesh deformation in Fig. 2d. The Jacobian determinants were also computed (Fig. 2e), revealing volumetric voxel changes with value=1 for volume preservation, <1 for shrinkage and >1 for volume expansion. Also, the residual image were calculated by subtracting the registered post-treatment image from the pre-treatment image, which revealed the voxel-wised image intensity variation superimposed on the pre-treatment 3D breast structure as shown in Fig. 2f. Moreover, the registration-derived information (e.g., deformation field and residual image) could be applied to analyze spatial heterogeneity of interested multi-parameter maps to augment tumor response prediction to chemotherapy response (known as parametric response mapping technique).

3.2 Landmark error of DRAMMS ordered by challenge level

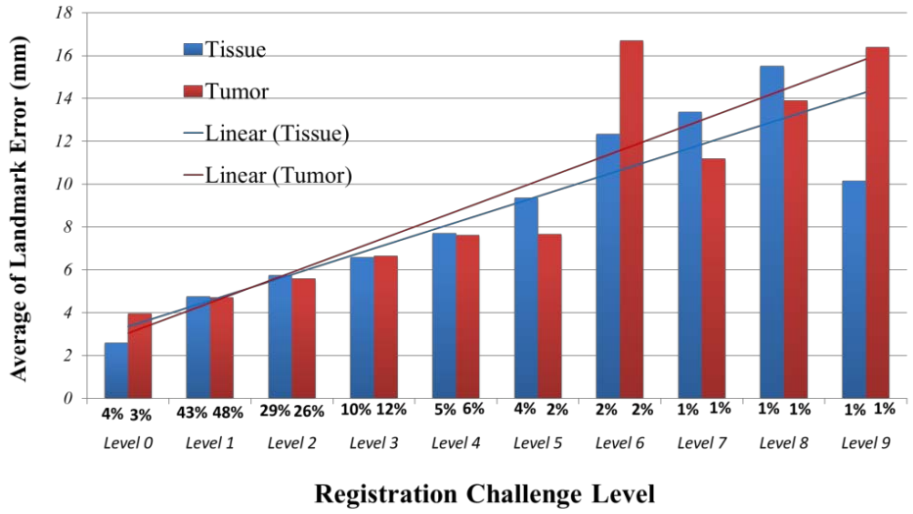


Fig. 3. Plot of relation between average of landmark errors within tumor and normal tissue region calculated from DRAMMS and registration challenge levels. Two straight lines represent the linear regression of average landmark error for tumor or tissue region.

Fig. 3 shows the landmark errors from registration algorithm DRAMMS ordered by landmark grading levels (i.e., registration challenge) with most landmarks falling within low level (relatively little challenge) and more than 95% landmarks within level 5. Both landmark errors from normal breast tissue and tumor region show a similar high linear correlation with registration challenge levels. For high level landmarks (level 5 and greater), the variation between two raters is significantly high, indicating that these landmarks may not be suitable to serve as ground truth for evaluating registration accuracy. Moreover, it is easy to find out the performance of DRAMMS is equivalently sound within both the normal breast tissue and tumor regions for low-grade regions (may have relatively small anatomical variation during chemotherapy), and for high-grade regions, both human raters and DRAMMS share similar uncertainty patterns for identifying structure correspondences within both imaging spaces. To sum up, DRAMMS performs well in the sub-regions where the inter-rater agreement is high, while DRAMMS fails where the inter-rater disagreement is strong.

3.3 Algorithm comparison

Fig. 4 shows the average and standard deviation for the landmark errors of tumor and/or normal breast regions from all six deformable registration algorithms (exclude landmark error higher than level 5), along with the mean and standard deviation of inter-rater differences. Overall, DRAMMS shows the highest overall accuracy among all six algorithms and performed equivalently well within both tumor and normal breast tissue regions, which is lower than 5mm

misalignment based on the ground truth generated by human raters. Since imaging space is discrete, and the voxel size is $0.70\times0.70\times2.0mm^3$. Therefore, compared with human rater that is around 1~2 voxel size misalignment, DRAMMS generate an extremely accurate alignment result (around 2~3 voxel size misalignment).

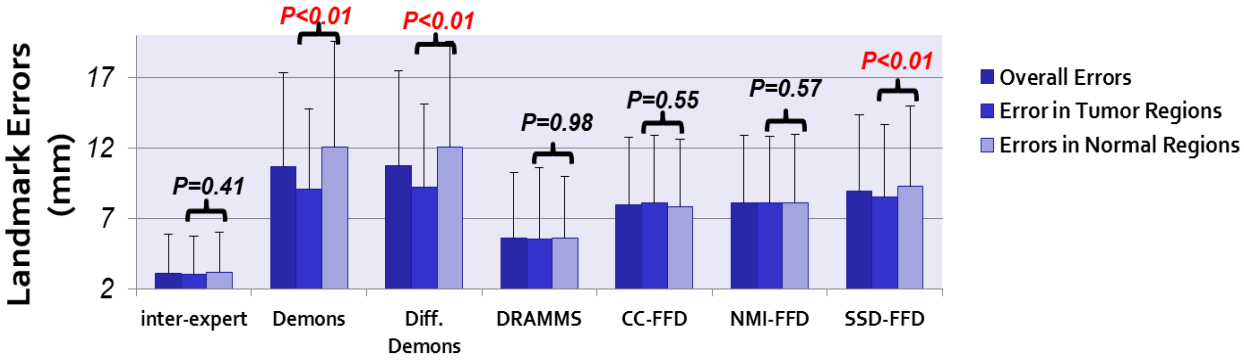


Fig. 4. Landmark errors of different deformable registration algorithm at different regions (tumor and/or normal tissue). The p-value from Student's t-test indicates if there are substantial differences between tumor and normal breast tissue area.

3.4 Parameter adjustment

Similar to other registration algorithm, there are many parameters inside DRAMMS, which could significantly affect its performance. For instance, Fig. 5 shows the tendency of average landmark error for tumor or breast tissue region to change with respect to an increasing g-value (i.e., the regularization term adjusts the smoothness of the deformation field). The average landmark errors for both regions initially decreased slightly, and then started to increase quickly, and lastly reached maximal boundary. However, there are many other metrics (e.g., determinant of Jacobian > 0) that should be considered to evaluate the deformation field besides landmark error in order to obtain a realistic and meaningful result. Finally, g-value is set to be 0.2 to optimize the received deformation field.

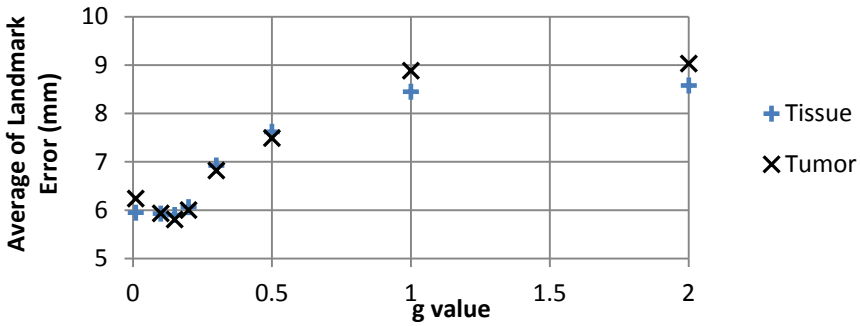


Fig. 5. Average of landmark errors variation with respect to the g value setting in DRAMMS in both tumor and tissue regions.

4. CONCLUSION

In this study, we further validated our in-house deformable registration algorithm DRAMMS for longitudinal breast DCE-MR images. DRAMMS showed a high linear correlation with registration challenge level (i.e., human rater performance), and achieved the highest accuracy in comparison to other competitive registration algorithms. A larger dataset from the ACRIN 6657 I-SPY trial is currently being acquired and processed to further validate these findings, and the features generated from DRAMMS have potential to serve as novel biomarkers to assess early tumor response to neoadjuvant chemotherapy as well as long term pathologic response by evaluating tissue deformation.

5. ACKNOWLEDGMENTS

This work was supported by a pilot grant from the University of Pennsylvania Center for Biomedical Image Computing and Analysis (CBICA). We would also like to thank Dr. Nola Hylton from UCSF for the permission to use the I-SPY data and useful discussions on this research work.

6. REFERENCE

[1] American Cancer Society., [Breast cancer facts & figures] American Cancer Society, Atlanta, Ga.

[2] W. F. Symmans, F. Peintinger, C. Hatzis *et al.*, "Measurement of residual breast cancer burden to predict survival after Neoadjuvant chemotherapy," Journal of Clinical Oncology, 25(28), 4414-4422 (2007).

[3] S. Yeluri, P. Jose, A. Saha *et al.*, "Actual 5-year Survival after Neoadjuvant Chemotherapy and Resection for Esophageal Adenocarcinoma," Annals of Surgical Oncology, 21, S150-S151 (2014).

[4] C. S. Fisher, C. X. Ma, W. E. Gillanders *et al.*, "Neoadjuvant Chemotherapy Is Associated with Improved Survival Compared with Adjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer Only after Complete Pathologic Response," Annals of Surgical Oncology, 19(1), 253-258 (2012).

[5] E. S. R. Currin, L. K. Dunnwald, R. K. Doot *et al.*, "Quantitative measures of FDG PET after neoadjuvant chemotherapy to predict breast cancer patient survival," Journal of Clinical Oncology, 30(15), (2012).

[6] A. R. Davies, J. A. Gossage, J. Zylstra *et al.*, "Tumor Stage After Neoadjuvant Chemotherapy Determines Survival After Surgery for Adenocarcinoma of the Esophagus and Esophagogastric Junction," Journal of Clinical Oncology, 32(27), 2983-+ (2014).

[7] N. M. Hylton, J. D. Blume, W. K. Bernreuter *et al.*, "Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL," Radiology, 263(3), 663-72 (2012).

[8] L. Martincich, F. Montemurro, G. De Rosa *et al.*, "Monitoring response to primary chemotherapy in breast cancer using dynamic contrast-enhanced magnetic resonance imaging," Breast Cancer Research and Treatment, 83(1), 67-76 (2004).

[9] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable Medical Image Registration: A Survey," Ieee Transactions on Medical Imaging, 32(7), 1153-1190 (2013).

[10] F. P. M. Oliveira, and J. M. R. S. Tavares, "Medical image registration: a review," Computer Methods in Biomechanics and Biomedical Engineering, 17(2), 73-93 (2014).

[11] C. J. Galban, T. L. Chenevert, C. R. Meyer *et al.*, "The parametric response map is an imaging biomarker for early cancer treatment outcome," Nature Medicine, 15(5), 572-576 (2009).

[12] J. L. Boes, B. A. Hoff, N. Hylton *et al.*, "Image Registration for Quantitative Parametric Response Mapping of Cancer Treatment Response," Translational Oncology, 7(1), 101-110 (2014).

[13] C. Tsien, C. J. Galban, T. L. Chenevert *et al.*, "Parametric Response Map As an Imaging Biomarker to Distinguish Progression From Pseudoprogression in High-Grade Glioma," Journal of Clinical Oncology, 28(13), 2293-2299 (2010).

[14] N. Cho, S. A. Im, I. A. Park *et al.*, "Breast Cancer: Early Prediction of Response to Neoadjuvant Chemotherapy Using Parametric Response Maps for MR Imaging," Radiology, 272(2), 385-396 (2014).

[15] X. Li, H. Kang, L. R. Arlinghaus *et al.*, "Analyzing Spatial Heterogeneity in DCE- and DW-MRI Parametric Maps to Optimize Prediction of Pathologic Response to Neoadjuvant Chemotherapy in Breast Cancer," Translational Oncology, 7(1), 14-22 (2014).

[16] H. C. Thoeny, and B. D. Ross, "Predicting and Monitoring Cancer Treatment Response with Diffusion-Weighted MRI," Journal of Magnetic Resonance Imaging, 32(1), 2-16 (2010).

[17] E. A. Eisenhauer, P. Therasse, J. Bogaerts *et al.*, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," European Journal of Cancer, 45(2), 228-247 (2009).

[18] C. C. Jaffe, "Measures of response: RECIST, WHO, and new alternatives," Journal of Clinical Oncology, 24(20), 3245-3251 (2006).

[19] J. G. Perez-Larraya, M. Lahutte, G. Petrirena *et al.*, "Response assessment in recurrent glioblastoma treated with irinotecan-bevacizumab: comparative analysis of the Macdonald, RECIST, RANO, and RECIST plus F criteria," Neuro-Oncology, 14(5), 667-673 (2012).

[20] Y. M. Ou, A. Sotiras, N. Paragios *et al.*, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," Medical Image Analysis, 15(4), 622-639 (2011).

[21] Y Ou, SP Weinstein, EF Conant *et al.*, "Deformable Registration for Quantifying Longitudinal Tumor Changes During Neoadjuvant Chemotherapy," Magnetic Resonance in Medicine, (2014).
[22] T. Vercauteren, X. Pennec, A. Perchant *et al.*, "Diffeomorphic demons: Efficient non-parametric image registration," Neuroimage, 45(1), S61-S72 (2009).
[23] D. Rueckert, L. I. Sonoda, C. Hayes *et al.*, "Nonrigid registration using free-form deformations: Application to breast MR images," Ieee Transactions on Medical Imaging, 18(8), 712-721 (1999).
[24] B. K. P. Horn, and B. G. Schunck, "Determining Optical-Flow," Artificial Intelligence, 17(1-3), 185-203 (1981).
[25] C. J. Liu, and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," Ieee Transactions on Image Processing, 11(4), 467-476 (2002).