

AbstractID: 4863 Title: Evaluation of similarity measures for use in intensity-based registration in radiotherapy

Purpose: In radiotherapy, the implementation of automated patient positioning depends on optimal selection of a similarity measure. Visual inspection of a registration is insufficient for evaluating performance of a measure, since high precision registration is usually involved. A general methodology is presented for evaluation of a measure based on convergence, precision and systematic errors for 2D and 3D imaging.

Method and Materials: The following common metrics were studied: mean pixelwise product (PROD), normalized correlation coefficient (NCC), partitioned intensity uniformity (PIU), mutual information (MI), normalized mutual information (NMI), entropy of the difference image (ED), gradient correlation (GC), gradient difference (GD), and pattern intensity (PI). Imaging from an anthropomorphic phantom and clinical data were used. Portal images were acquired with a video camera EPID and a flat-panel imager. Kilovoltage digitally reconstructed radiographs (DRR) and megavoltage DRRs (MVDRR) were used for reference images. The following shifts were carried out: translations from -10 mm to +10 mm with 0.2 mm increments, and/or rotations from -10° to $+10^\circ$ with 0.2° increments. The effect down-sampling and image type were also investigated.

Results: The selection of a measure depended on image type and down-sampling schedule. PROD introduced the largest systematic error. NCC, PIU, MI, NMI, and GD typically had systematic errors less than 1 mm and 1° . Down-sampling has less effect on NCC, PIU, and GC. MI, NMI, GC, GD, and PI have sharp "peaks" around their global optima. Asymmetry in the objective function was observed in NCC, PIU, MI, NMI, and ED. Registrations using GC, GD, and PI were less sensitive to the variation of anatomical structures. Using MVDRRs instead of DRRs increased objective function smoothness.

Conclusion: A general methodology was presented for evaluation and selection of a measure. Selection was application dependent, indicating necessity of a validation study, as presented here, prior to clinical use.