

# Lung texture in serial thoracic CT scans: Registration-based methods to compare anatomically matched regions<sup>a)</sup>

Alexandra R. Cunliffe, Samuel G. Armato III, Xianhan M. Fei, and Rachel E. Tuohy  
*Department of Radiology, The University of Chicago, Chicago, Illinois 60637*

Hania A. Al-Hallaq<sup>b)</sup>  
*Department of Radiation and Cellular Oncology, The University of Chicago, Chicago, Illinois 60637*

(Received 1 April 2013; revised 29 April 2013; accepted for publication 1 May 2013; published 23 May 2013)

**Purpose:** The aim of this study was to compare three demons registration-based methods to identify spatially matched regions in serial computed tomography (CT) scans for use in texture analysis.

**Methods:** Two thoracic CT scans containing no lung abnormalities and acquired during serial examinations separated by at least one week were retrospectively collected from 27 patients. Over 1000 regions of interest (ROIs) were randomly placed in the lungs of each baseline scan. Anatomically matched ROIs in the corresponding follow-up scan were placed by mapping the baseline scan ROI center pixel to (1) the original follow-up scan, (2) the follow-up scan resampled to match the baseline scan voxel size, and (3) the follow-up scan aligned to the baseline scan through affine registration. Mappings used the vector field obtained through demons deformable registration of each follow-up scan variant to the baseline scan. 140 texture features distributed among five feature classes were calculated in all ROIs. Feature value differences between paired ROIs were evaluated using Bland-Altman 95% limits of agreement. For each feature, (1) the mean feature value change and (2) the difference between the upper and lower limits of agreement were normalized to the mean feature value to obtain, respectively, the normalized bias and normalized range of agreement (nRoA). Non-parametric tests were used to evaluate differences in normalized bias and nRoA across the three methods.

**Results:** Because patient CT scans contained no pathology, minimal changes in feature values were expected (i.e., low nRoA and normalized bias). Seventy-five features with very large feature value variability ( $nRoA \geq 100\%$ ) were excluded from further analysis. Across the remaining 65 features, significant differences in normalized bias were observed among the three methods. The lowest normalized bias (median: 0.06%) was achieved when feature values were calculated on original follow-up scans. The affine registration method achieved the lowest nRoA, though nRoA was not significantly increased using original follow-up scans. Features with low nRoA values also had low normalized bias, though the converse was not necessarily true. Using nRoA as a metric, a set of 20 features having both low nRoA and normalized bias were identified.

**Conclusions:** Three methods to facilitate texture analysis of serial CT scans using demons registration for ROI placement were evaluated. The bias in feature value change between matched ROIs was minimized when feature values were calculated on original baseline and follow-up scans. A set of features that had both low bias and variability (nRoA) in feature value change using this method were identified. This texture analysis approach could facilitate future measurement of pathologic changes between CT scans without necessitating calculation of feature values on deformed scans.

© 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4805110>]

Key words: texture analysis, deformable image registration, lung, CT

## I. INTRODUCTION

Texture analysis of thoracic computed tomography (CT) images can allow for identification of various disease patterns and/or aid in the diagnosis of lung disease. Several groups have used CT image texture analysis to identify various patterns of diffuse lung disease, including emphysema, ground-glass opacification, and honeycombing.<sup>1-3</sup> Yao *et al.*<sup>4</sup> showed that a set of first-order, gray-level co-occurrence matrix (GLCM), and gray-level run-length matrix texture features could distinguish H1N1 influenza from fibrosis, nonin-

fluenza infections, and normal lung tissue. Ganeshan *et al.*<sup>5</sup> found that lung cancer tumor metabolism and stage correlated with a set of first-order texture features calculated using CT scans that had been filtered to emphasize certain morphology. These researchers have demonstrated that, using texture analysis, CT scans can provide valuable information about a patient's disease status at a single point in time.

For patients with progressive lung disease, CT scans at multiple time points may be acquired to assess disease progression or treatment response. Coregistration of these serial CT scans facilitates assessment of disease changes in spatially

matched locations over time. A number of algorithms based on control point matching (splines deformable registration) (Refs. 6 and 7) or optical flow (demons deformable registration) (Refs. 8 and 9) have been applied for thoracic CT scan registration. These deformable registration techniques correct for differences both in patient positioning and respiratory phase between serial scans. Changes between serial scans that are not due to positioning differences can thus be identified.

Texture analysis of coregistered serial CT scans would allow for measurement of texture changes between anatomically matched regions over time, providing objective measurement of disease progression or treatment response. Several groups have performed quantitative analysis using coregistered CT scans. Palma *et al.*<sup>10</sup> measured pixel value changes between thoracic CT scans acquired before and after lung radiation therapy that had been registered using B-splines deformable registration; no higher-order features were investigated. Arzhaeva *et al.*<sup>11</sup> used B-splines registration to align CT scan pairs acquired from patients with interstitial lung disease, then quantified changes between scans using a set of first-order, dissimilarity, and filter-based texture features. They did not determine, however, whether the registration process itself may have also altered CT scan texture, potentially confounding changes due to disease. In a previous publication, we evaluated the registration accuracy achieved using four image registration algorithms and examined their effects on lung CT image texture features.<sup>12</sup> While demons deformable registration achieved the highest registration accuracy among the four methods, this registration method introduced nonzero bias in feature values due to the altered lung parenchymal texture in deformed CT scans. This study demonstrated that although texture analysis of coregistered CT scans appears a straightforward method for measuring temporal changes, this approach may not be reliable due to the registration-induced changes in the deformed scans. Alternatively, methods that facilitate texture comparisons using CT scans that have not been deformed should be investigated.

In the current work, several approaches to measure changes in image texture between serial CT scans are presented and evaluated. These methods use the registration accuracy achieved by demons deformable registration to identify matched regions in CT scan pairs. Texture features are then calculated directly on CT scans that have not been deformed, thus eliminating the possibility that deformation itself had altered texture. The bias and variability in feature value change introduced by these methods are calculated and compared. The goal is to identify a method that minimizes bias across features in the absence of disease change and to identify a set of texture features with low variability using this approach.

## II. METHODS

### II.A. Patient database

The patient database has been described previously.<sup>12</sup> Healthy thoracic CT scan pairs acquired between one week

and two years apart (mean: 126 days) for 27 patients were retrospectively collected under Institutional Review Board (IRB) approval. For each patient, CT scan pairs were acquired either with ( $n = 24$ ) or without ( $n = 3$ ) intravenous contrast injection. All scans were determined to have no lung abnormalities, defined by the absence of acute disease or nodules exceeding 4 mm, by an experienced radiologist. Scans were acquired after patients were instructed to inspire and hold their breath using multislice Philips Brilliance CT scanners (Brilliance 16, Brilliance 16P, Brilliance 64, or Brilliance iCT256) and reconstructed at submillimeter pixel spacing and 1 mm slice thickness using identical high-resolution lung reconstruction and smoothing kernels. The peak kilovoltage was 120 kVp for all scans, and the exposure ranged from 144 to 351 mAs. The median difference in exposure between scan pairs was 20 mAs (range: 0–156 mAs).

### II.B. Region of interest (ROI) identification and mapping

The method used to identify ROIs in a baseline scan CT and map them to their corresponding locations in a follow-up scan is shown in Fig. 1. Following automated lung segmentation,<sup>13</sup> randomly placed  $32 \times 32$ -pixel ROIs were automatically identified in each patient's baseline scan (Step 1). Only ROIs lying entirely within the lung borders while not overlapping previously selected ROIs were retained. A maximum of ten ROIs were identified in each axial CT section, resulting in 1435–2413 (mean: 1898) ROIs per patient.

Identification of ROIs in follow-up scans relied on application of fully automated demons deformable registration using the open-source software package Plastimatch (version 1.5.12-beta).<sup>8</sup> First, the follow-up scan was deformed to match the baseline scan, outputting a displacement vector field (Step 2) that mapped each pixel in the baseline scan to a corresponding location in the follow-up scan. The center of a baseline scan ROI was mapped to the original (i.e., not deformed) follow-up scan using the vector field, with the remainder of the ROI placed in the same axial section (Step 3). Follow-up scan ROIs with the same size in pixels as the baseline scan ROIs (i.e.,  $32 \times 32$  pixels) and with the same physical size in millimeters (i.e.,  $n \times n$  pixels) as the baseline scan ROIs were both considered, otherwise referred to as “follow-up<sub>32×32</sub>” and “follow-up<sub>n×n</sub>” ROIs, respectively. This distinction results from the fact that baseline and follow-up scans generally do not possess the same pixel size.

In addition, follow-up CT scans were transformed using two distinct techniques that could potentially improve ROI alignment between baseline and follow-up scans. For the first alternative approach, the patient's follow-up CT scan was resampled to match the voxel size of the baseline scan, with bilinear interpolation used to assign the corresponding gray-level values.  $32 \times 32$ -pixel ROIs in the baseline scans and resampled follow-up scans (“resampled<sub>32×32</sub>” ROIs) were therefore the same physical size, removing the possibility that differences in the pixel size between CT scans in a pair could affect texture feature value comparisons. For the second approach, the follow-up CT scan was globally registered to

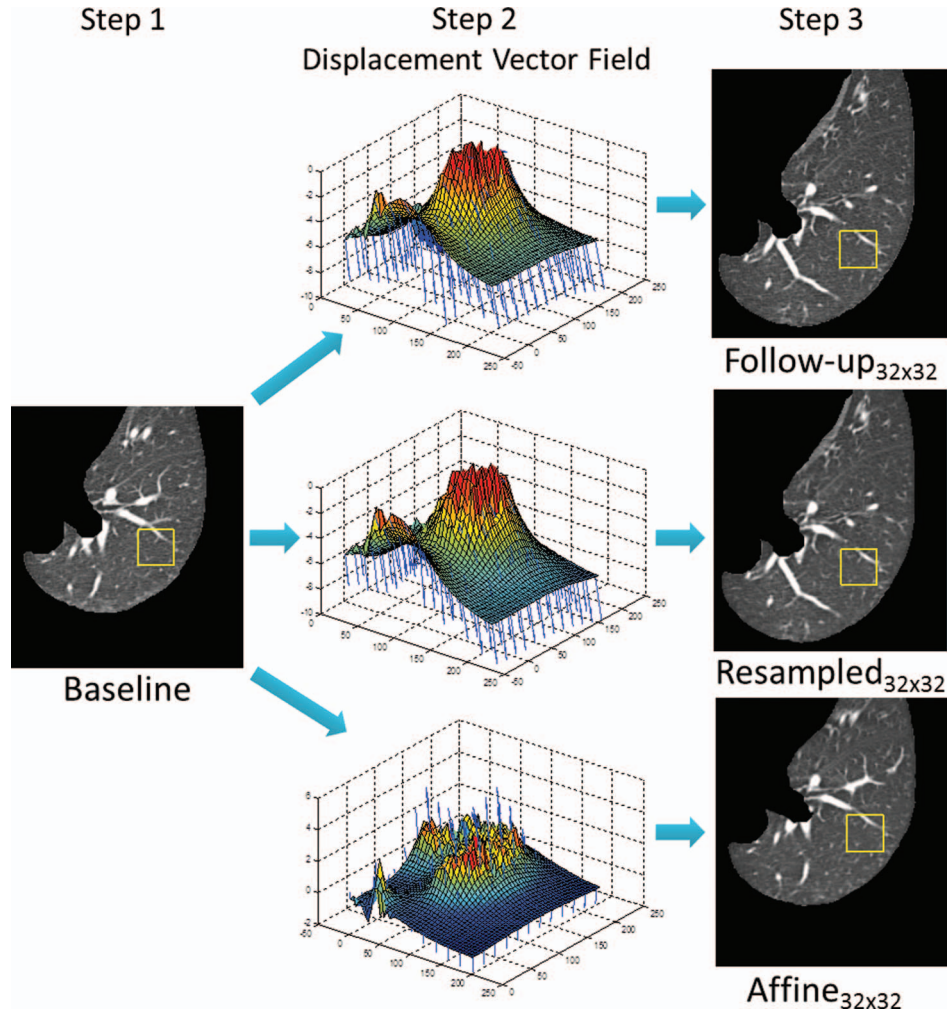


FIG. 1. Steps for ROI (yellow box) mapping from a baseline scan to a follow-up scan and two follow-up scan transformations using the displacement vector fields displayed in column 2.

the baseline scan using affine registration. While previously presented methods ensured only that the center of a baseline scan ROI was well-matched in the follow-up scan, this alternative approach increases the likelihood that an entire inplane  $32 \times 32$ -pixel ROI in the globally registered follow-up scan (“affine<sub>32x32</sub>” ROI) contains similar anatomy to the ROI in the baseline scan. The transformed follow-up scans were then registered to the baseline scan using demons registration to achieve a mapping between ROIs in the baseline and transformed follow-up scans as previously described (see Fig. 1).

### II.C. CT scan texture analysis

A set of 140 texture features distributed among first-order, fractal, Fourier, GLCM, and Laws’ filter classes were calculated. These features are presented in more detail in Sec. II.D. Feature values in matched ROIs in baseline and follow-up scans were compared using Bland-Altman limits of agreement for the case of multiple measurements per patient.<sup>14</sup> The difference between the upper and lower 95% limits of agreement was calculated and normalized to the

mean feature value across all baseline scan ROIs to obtain the normalized range of agreement (nRoA). The bias was also calculated as the mean difference between baseline and follow-up scan ROI feature values and normalized to the mean ROI feature value in the baseline scan to obtain the normalized bias (nBias)

$$\text{nRoA} = \frac{(95\% \text{ limit upper bound} - 95\% \text{ limit lower bound})}{|\text{Mean feature value}|} \times 100\%, \quad (1)$$

$$\text{nBias} = \frac{\text{Bias}}{|\text{Mean feature value}|} \times 100\%, \quad (2)$$

$$\text{Mean feature value} = \frac{1}{\text{nROIs}} \sum_{\text{all ROIs}} \text{Feature value}_{\text{baseline}}, \quad (3)$$

where nROIs is the total number of baseline scan ROIs across all patients. Because patient CT scans contained no pathology, minimal changes in feature values were expected between scans in a pair, with the expectation of narrow limits of agreement (low nRoA) and low bias.

To separate bias introduced during the mapping technique from bias introduced due to intrinsic CT scan differences, analysis was repeated in the reverse order as detailed in a previous publication.<sup>12</sup> ROIs were first identified in follow-up CT scans, then ROI center pixels were mapped to their corresponding locations in all of the variants of the baseline CT scans. The normalized bias was again calculated and averaged with the bias achieved using the forward registration method. For bias introduced by differences between scans in a pair, the magnitude of bias was expected to be similar using both forward and reverse methods, though the direction of bias (i.e., the sign) would be reversed. Bias introduced by scan resampling or registration, however, was expected to have the same sign and similar magnitude irrespective of the registration direction. Thus, the average normalized bias ( $n\text{Bias}_{\text{avg}}$ ) represented the bias introduced during the registration/mapping method rather than by differences between the CT scans themselves.

Values of  $n\text{RoA}$  and  $n\text{Bias}_{\text{avg}}$  were calculated for all features and compared among the approaches. Friedman rank sum tests were used to determine if significant ( $p < 0.05$ ) differences in  $n\text{RoA}$  and  $n\text{Bias}_{\text{avg}}$  existed among the three methods.  $n\text{RoA}$  and  $n\text{Bias}_{\text{avg}}$  values were then compared between each pair of approaches using Wilcoxon signed rank tests (paired, two-tailed). To maintain model significance at  $\alpha = 0.05$ , significance levels for individual tests were modified according to the Bonferroni method. Results were also compared with our previous study in which texture feature differences had been calculated between baseline scans and demons-deformed follow-up scans directly, after renormalizing values of  $n\text{RoA}$  and  $n\text{Bias}_{\text{avg}}$  to the average baseline scan feature values.<sup>12</sup>  $n\text{RoA}$  and  $n\text{Bias}_{\text{avg}}$  values for GLCM features reported in our previous study were adjusted for consistency with the methods reported by Haralick *et al.*<sup>15</sup>

## II.D. Texture features

Texture features were derived from five feature classes.

### II.D.1. First-order histogram features

First-order histogram features<sup>16–18</sup> characterize the gray-level histogram of an image or region. The 19 first-order features calculated were: mean, median, maximum, minimum, mean absolute deviation, range, interquartile range (IQR), standard deviation, skewness, kurtosis, energy, entropy, binned entropy (calculated after sorting data into 256 histogram bins), 5%, 30%, 60%, and 95% histogram quantiles, and balance of the inner 40% and inner 90% of the gray-level histogram.

### II.D.2. Fractal features

These features characterize region self-similarity at different scales and indicate image detail. Fractal dimension was calculated using three methods: the blanket method,<sup>19</sup> the Brownian motion method,<sup>20</sup> and the box-counting method<sup>18,21,22</sup> (including total, coarse, and fine aspects of the box-counting dimension).

### II.D.3. Fourier features

Fourier-based features characterize the spatial frequency components of a region. Fourier features calculated include the first moment of the power spectrum,<sup>23</sup> the root-mean-squared variation,<sup>23</sup> and the energy of the Fourier-transformed region and of several subspaces representing different frequency components. These subspaces were the high- and low-frequency rings when the region was subdivided into two sections; the low, moderately low, moderately high, and high frequency rings when the region was subdivided into four sections; and the eight sectors formed when the region was divided into 45° sectors.<sup>16</sup>

### II.D.4. Laws' filter features

These features emphasize five aspects of region microstructure: spot, wave, ripple, edge, and level surfaces.<sup>24</sup> Regions were first convolved with a particular filter, then a set of features (mean, energy, binned entropy, maximum, minimum, and standard deviation) were calculated on rotationally invariant filtered regions.

### II.D.5. Gray-level co-occurrence matrix features

GLCM features<sup>15–17</sup> quantify the spatial relationships of gray-level values in a region. For each region, a GLCM was first constructed to count all gray-level pairs separated by 1 pixel and at angle  $\theta$ . Fourteen features were calculated from the GLCM: correlation, inertia, absolute value, inverse difference, energy, entropy, contrast, sum of squares variance, sum average, sum variance, sum entropy, difference average, difference variance, and difference entropy. Four directions were examined ( $\theta = \{0^\circ, 45^\circ, 90^\circ, \text{and } 135^\circ\}$ ), and each feature was calculated by taking the average over the four directions.

## III. RESULTS

For 75 of 140 features,  $n\text{RoA}$  values were greater than 100% using all approaches, indicating that the variability exceeded the mean feature value across baseline scan ROIs [see Eq. (1)]. These findings were similar to our previous study, in which 52 features had  $n\text{RoA}$  values that exceeded 100%.<sup>12</sup> For the remaining 65 features with  $n\text{RoA} < 100\%$ , follow-up<sub>32×32</sub> ROI and follow-up<sub>n×n</sub> ROI texture feature values were compared to determine if it was necessary to change ROI size to accommodate differences in pixel dimension in the follow-up scan. Boxplots (Fig. 2) display the  $n\text{Bias}_{\text{avg}}$  and  $n\text{RoA}$  for these two approaches.  $n\text{Bias}_{\text{avg}}$  appeared slightly higher when ROI size was altered (median: 0.14% versus 0.06%), though this difference was not significant. Furthermore, several outlier values of  $n\text{Bias}_{\text{avg}}$  existed using the follow-up<sub>n×n</sub> ROIs. There was no significant difference in  $n\text{RoA}$  between the two approaches (median: 67.32% versus 67.44%). Due to the similarities between the two approaches, only the measurements obtained with 32 × 32-pixel ROIs (follow-up<sub>32×32</sub>) were considered for the remainder of the analysis.

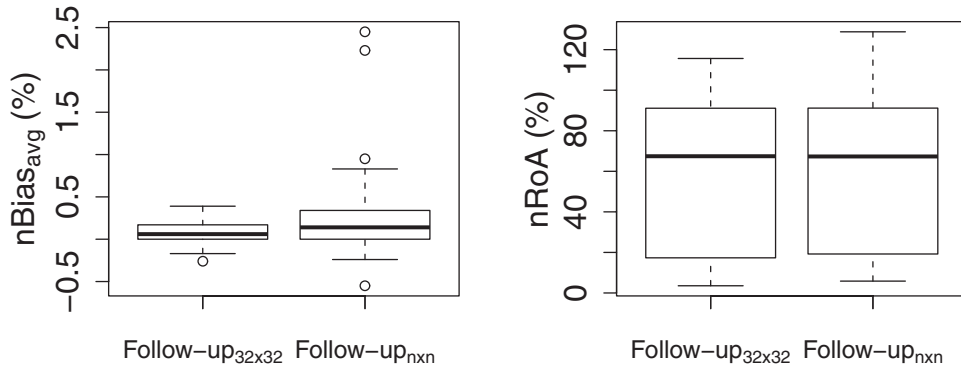


FIG. 2. Boxplot compare  $nBias_{avg}$  (left) and  $nRoA$  (right) values when follow-up scan ROI size was fixed at  $32 \times 32$  pixels versus when size was adjusted to match the physical ROI size in the baseline scan.

For the 65 features with  $nRoA < 100\%$ , Friedman rank sum tests indicated that significant differences in  $nBias_{avg}$  existed among the follow-up $_{32 \times 32}$ , resampled $_{32 \times 32}$ , and affine $_{32 \times 32}$  ROI approaches ( $p = 3 \times 10^{-9}$ ). Wilcoxon signed rank tests showed that the differences in  $nBias_{avg}$  between each pairwise comparison of the three methods were significant ( $p < \alpha/3 = 0.017$  for all three tests). The lowest median value of  $nBias_{avg}$  across features occurred for follow-up $_{32 \times 32}$  ROIs, while affine $_{32 \times 32}$  ROIs yielded the highest median value of  $nBias_{avg}$  (Table I). Significant difference in  $nRoA$  among the three methods were also observed ( $p = 1 \times 10^{-9}$ ). Using Wilcoxon signed ranks tests for pairwise comparisons, no significant differences in  $nRoA$  were observed between the follow-up $_{32 \times 32}$  and the affine $_{32 \times 32}$  ROIs ( $p = 0.30$ ); however, these two mapping methods achieved significantly lower  $nRoA$  than the resampled $_{32 \times 32}$  ROIs ( $p < 0.017$ ), indicating lower variability in feature value change. Figure 3 depicts histograms of  $nBias_{avg}$  and  $nRoA$  values across the 65 features for each ROI mapping technique. For features calculated from resampled $_{32 \times 32}$  and affine $_{32 \times 32}$  ROIs, values of  $nBias_{avg}$  were more widely distributed about zero than for follow-up $_{32 \times 32}$  ROIs, indicating that feature values were consistently altered from baseline using these methods.

The relationship between  $nRoA$  and  $nBias_{avg}$  across features was investigated. For all three mapping methods, the variability in  $nBias_{avg}$  increased with increasing  $nRoA$ , as indicated in Fig. 4. Although the features with the lowest  $nRoA$  values also had low  $nBias_{avg}$  values, many of the features with low  $nBias_{avg}$  had large values of  $nRoA$ .

Table II compares  $nRoA$  and  $nBias_{avg}$  values obtained using the three presented mapping methods with the values obtained previously when feature values were calculated on

demons-deformed follow-up scans directly (demons $_{32 \times 32}$ ).<sup>12</sup> The 20 features displayed in the table are those that yielded  $nRoA \leq 20\%$  for at least one of the three currently presented approaches.

Beyond  $nRoA = 20\%$ , the number of features per  $nRoA$  increment decreased, indicated by the changing slope in Fig. 5. For the 20 features presented,  $nRoA$  was significantly smaller using the demons $_{32 \times 32}$  method than the follow-up $_{32 \times 32}$  method ( $p = 4 \times 10^{-4}$ ), while  $nBias_{avg}$  was significantly smaller for follow-up $_{32 \times 32}$  ROIs ( $p = 0.008$ ).

#### IV. DISCUSSION

This study investigated the effects on texture feature values of mapping ROIs between serial CT scans. Compared to resampled $_{32 \times 32}$  and affine $_{32 \times 32}$  ROIs, follow-up $_{32 \times 32}$  ROIs were found to yield the lowest values of  $nBias_{avg}$  without significantly increasing  $nRoA$ . This mapping method may therefore be the most appropriate technique for texture analysis of serial CT scans, as minimal bias in feature values was introduced. With this method, low values of  $nRoA$  also had low values of normalized bias (Fig. 4), though low normalized bias did not imply low values of  $nRoA$ . As the variability in feature value change increased ( $nRoA$  increased), it is possible that the accuracy with which bias could be measured decreased, resulting in increasing variability in  $nBias_{avg}$  with increasing  $nRoA$ . Thus, the use of  $nRoA$  is more selective of features than  $nBias_{avg}$ . Table II displays features that had low  $nRoA$  and  $nBias_{avg}$  when feature values were calculated on follow-up scans and compares the results with a previous study.<sup>12</sup> Although mean  $nRoA$  was lower using the previously reported methods that directly deformed scans,  $nBias_{avg}$  was an order of magnitude larger, indicating that feature values were consistently altered using these methods. In the previous study, feature values were also calculated on scans that were deformed through B-splines registration. For 19 of the 20 features, values of  $nRoA$  and  $nBias_{avg}$  calculated using follow-up $_{32 \times 32}$  ROIs remained less than or equal to the values calculated previously using B-splines-deformed follow-up ROIs, indicating that localizations of ROI center pixels using demons registration remained more accurate than B-splines. The previous study demonstrated that deformable

TABLE I. Median normalized bias, median  $nRoA$ , and the respective IQR for the three methods using the 65 features with  $nRoA < 100\%$  for at least one method.

Mapping	Median $nBias_{avg}$ [IQR] (%)	Median $nRoA$ [IQR] (%)
Follow-up $_{32 \times 32}$	0.1 [0.0, 0.2]	67.4 [17.3, 91.1]
Resampled $_{32 \times 32}$	-5.2 [-12.7, -0.8]	72.3 [18.6, 97.1]
Affine $_{32 \times 32}$	-5.9 [-14.2, -0.9]	64.6 [16.2, 86.6]

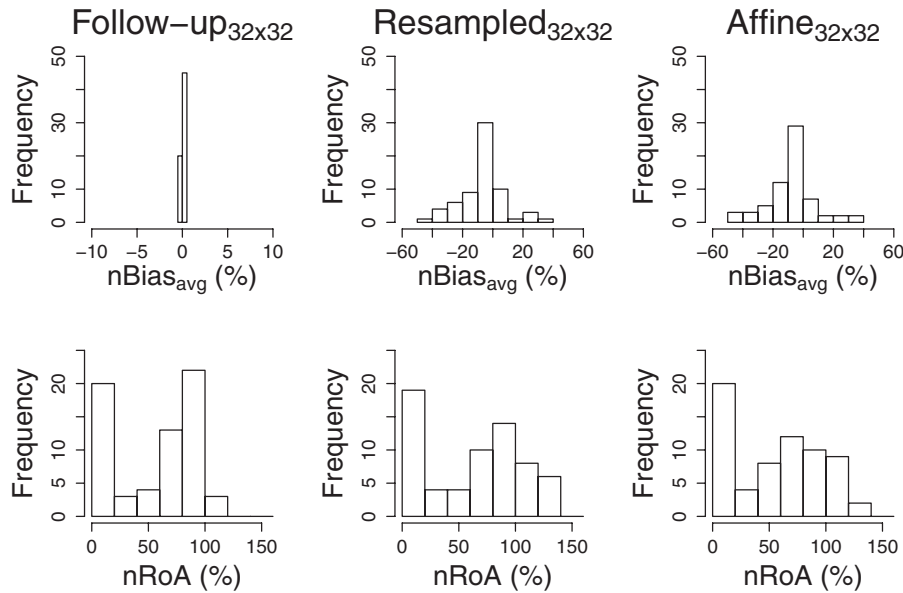


FIG. 3. Histograms of  $nBias_{avg}$  (top) and  $nRoA$  (bottom) values obtained for 65 features using the three mapping methods. For visualization purposes, the range of  $nBias_{avg}$  values displayed on the x-axis for the follow-up scan mapping (top left) is smaller than for the two transformed scan mappings (top middle and top right).

registration of CT scans introduced bias in texture feature values, potentially restricting the utility of this approach in future texture analysis applications. The techniques presented in this study, however, provide an avenue for texture analysis of serial scans that utilizes demons for coregistration of ROIs but not for scan deformation, thereby reducing the effects of image distortion.

The 20 features displayed in Table II had the lowest values of  $nRoA$  ( $nRoA < 20\%$ ) using follow-up $_{32 \times 32}$  ROIs. Above  $nRoA = 20\%$ , the number of features per  $nRoA$  increment declined so that approximately one feature was added per 5% increase in  $nRoA$  (Fig. 5). Because increasing the  $nRoA$  threshold (and thus increasing feature value variability) would include few additional features,  $nRoA = 20\%$  seemed an appropriate cutoff for this study. The choice of features to select for future texture analysis applications, however, will ultimately depend on the disease change itself. If, for example, the ap-

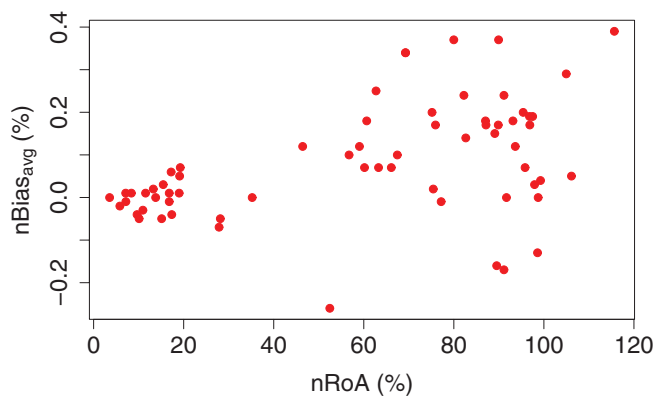


FIG. 4. Scatterplot of  $nBias_{avg}$  versus  $nRoA$  when features were calculated directly on follow-up scans. Each of the 65 features with  $nRoA < 100\%$  for at least one method is included. Although not shown, plots generated using the two alternative methods (resampled $_{32 \times 32}$  and affine $_{32 \times 32}$ ) were similar.

pearance of a particular disease pattern results only in small changes in feature values from baseline, features with low variability (i.e., small  $nRoA$ ) relative to the disease change should be selected. In general, the  $nRoA$  of features used for the detection task should be smaller than the change introduced by disease to achieve sufficiently high sensitivity. Thus, higher- $nRoA$  features could be included in analysis, particularly if they measure independent aspects of image texture. Steps should be taken, however, to minimize  $nBias_{avg}$ , as nonzero bias indicates that the image (and thus the disease pattern itself) is being distorted due to CT scan manipulation. As demonstrated in our previous work, deformable registration had similar effects on  $nBias_{avg}$  as the transformations presented here (resampling and affine registration).<sup>12</sup>

It is possible that some of the features identified may be insensitive to changes in lung parenchymal texture, resulting in similar feature values (thus, low  $nRoA$  and  $nBias_{avg}$ ) irrespective of actual differences between CT scans in a pair. In our previous work,<sup>12</sup> we observed that two of the features presented in Table II (minimum and 5% quantile) remained similar between CT scans in a pair even when registration accuracy was low, suggesting that these features may be ill-equipped to detect true changes between scans. In future studies, features that are insensitive to lung parenchymal changes will be identified using a database of patients with lung disease.

Only small changes in  $nRoA$  and  $nBias_{avg}$  were introduced when follow-up scan ROI size was adjusted. The reason for the generally small differences using the two techniques may be due to the similarities between matched ROIs in CT scan pairs. Differences in pixel size between scans in a pair did not exceed 0.18 mm, resulting in small differences in the extent of anatomy included in  $32 \times 32$ -pixel ROIs in the baseline and follow-up scans. ROI size adjustment was therefore not necessary to correct for differences in pixel size

TABLE II. Values of nRoA and nBias<sub>avg</sub> for the 20 features with nRoA ≤ 20% for at least one of the three methods.

	Follow-up <sub>32×32</sub>		Resampled <sub>32×32</sub>		Affine <sub>32×32</sub>		Demons <sub>32×32</sub>	
	nRoA (%)	nBias <sub>avg</sub> (%)	nRoA (%)	nBias <sub>avg</sub> (%)	nRoA (%)	nBias <sub>avg</sub> (%)	nRoA (%)	nBias <sub>avg</sub> (%)
First-order features								
Minimum	5.82	-0.02	7.77	1.35	8.09	1.57	9.79	3.39
Mean	13.79	0.00	13.56	0.03	12.94	-0.02	7.58	0.24
Median	10.95	-0.03	10.83	0.13	10.74	0.14	7.30	0.19
Binned entropy	16.79	0.01	16.96	-1.20	16.15	-1.43	14.42	-4.03
Unbinned entropy	11.53	0.01	12.00	-1.66	11.52	-2.02	10.13	-4.66
5% quantile	10.11	-0.05	11.32	2.93	10.86	3.26	10.84	4.83
30% quantile	9.67	-0.04	9.65	1.15	9.56	1.28	7.46	1.85
70% quantile	16.77	-0.01	16.45	-0.82	15.88	-0.92	9.19	-1.45
Fractal features								
Box counting dim.	15.13	-0.05	15.65	-3.92	13.53	-4.29	11.51	-5.91
Fine dimension	17.33	-0.04	21.03	-6.58	19.16	-7.08	16.27	-8.46
Brownian dimension	7.16	-0.01	6.85	-0.83	5.70	-0.91	4.48	-1.28
Laws' filter features								
E5L5 entropy	17.21	0.06	16.79	-1.46	14.05	-1.76	10.68	-2.54
R5L5 entropy	19.00	0.01	19.82	-5.23	17.66	-5.75	14.17	-4.01
S5L5 entropy	19.22	0.07	18.57	-2.67	15.65	-3.12	11.52	-3.67
W5L5 entropy	19.11	0.05	18.87	-4.04	16.31	-4.55	12.45	-4.36
GLCM features								
Difference entropy	13.26	0.02	15.20	-4.93	13.54	-5.53	12.84	-9.48
Entropy	3.55	0.00	2.88	0.02	2.88	-0.03	3.04	-0.50
Sum of squares var.	15.48	0.03	15.75	-1.77	14.59	-2.12	12.94	-3.95
Sum average	7.12	0.01	7.32	-0.78	6.91	-0.94	6.36	-1.86
Sum entropy	8.38	0.01	8.78	-1.68	8.34	-1.94	7.23	-3.63
Mean across all features	12.87	0.00	13.30	-1.60	12.20	-1.81	10.01	-2.46

between scans. Instead, changing the ROI size may have affected the values of features, resulting in slightly larger values of nBias<sub>avg</sub>. As an alternative to changing ROI size, follow-up scans were resampled to match voxel sizes in the correspond-

ing baseline scans prior to texture analysis in 32 × 32-pixel ROIs (resampled<sub>32×32</sub>). Resampling the scans, however, introduced consistent changes in feature values, indicated by an increased nBias<sub>avg</sub>. We therefore concluded that changing ROI size to match a baseline scan ROI introduced considerably smaller bias than transforming the CT scan itself. In applications where pixel size differences between scan pairs are larger than those observed here, the effects of ROI size on feature values will need to be reevaluated.

While all CT scans in this study scans were acquired using Philips Brilliance CT scanners, CT scan pairs were not necessarily acquired using the same scanner model (n = 12 with the same model for both scans versus n = 15 with different models). To determine the effects of scanner differences on nRoA and nBias<sub>avg</sub>, the results for patients with scans acquired using different scanners were compared with those of the patients who had both scans acquired using the same scanner model. Across the 20 features displayed in Table II, the mean difference in nRoA between the two groups did not exceed 2% using the three methods (resampled<sub>32×32</sub>, affine<sub>32×32</sub>, and follow-up<sub>32×32</sub>), while the mean difference in nBias<sub>avg</sub> remained below 0.1%. Wilcoxon signed rank tests (paired, one-sided) showed that, across the 20 features, nRoA and nBias<sub>avg</sub> were not significantly increased for any of the methods when scans were acquired using different scanner models. One limitation of this study is that it only considers CT scans acquired with Philips scanners and reconstructed using a single

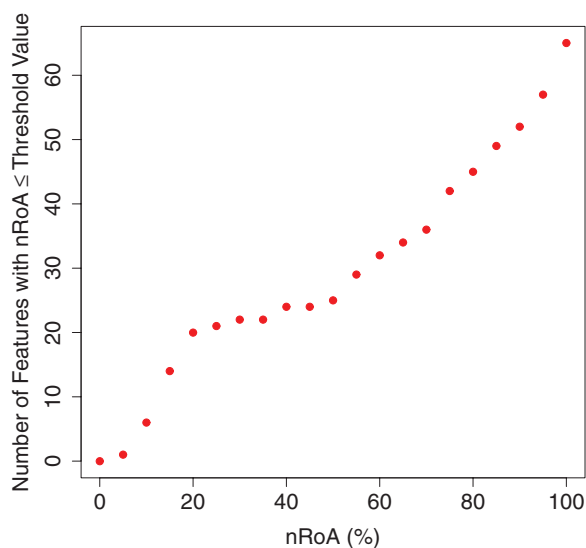


FIG. 5. The cumulative number of features with at most a given nRoA value for at least one of the methods. Plots generated using each of the methods individually (not shown) were similar. nRoA values ranged from 0% to 100% in 5% increments.

high-resolution lung reconstruction algorithm. The utility of the features identified here remains to be tested using other CT scanner models and reconstruction techniques. Furthermore, the results of this study should be validated using a database of patients with progressive disease. Differences between CT scan pairs resulting from disease changes may complicate the registration process, leading to inferior anatomic matching between baseline and follow-up scan ROIs. Future studies will evaluate the differences in anatomic alignment achieved when registering diseased patient CT scans and examine the effects on  $nRoA$  and  $nBias_{avg}$ .

To our knowledge, this is the first investigation that quantifies the effects of lung CT scan transformations on a variety of texture features. A study by Palma *et al.*<sup>25</sup> concluded that the registration algorithm used (B-splines) did not introduce notable changes in mean pixel value between CT scans. These results remain consistent with our study, in which minimal changes in mean pixel value due to CT scan transformation were observed with all of the approaches studied, resulting in low values of  $nRoA$  and  $nBias_{avg}$  (Table II). For the majority of the features investigated, however, nonzero bias was introduced when scans underwent transformation. Further studies that use these higher-order features to detect disease changes should consider the effects of scan transformation on the values of these features.

## V. CONCLUSIONS

Three methods for mapping ROIs between anatomically matched regions of serial CT scans were evaluated based on their effects on texture feature values. For each approach, the anatomic alignment accuracy achieved through demons registration was used to map ROI centers from baseline to follow-up scans. Feature values were calculated, however, on scans that had not been deformed, thus eliminating the possibility for CT scan deformation to alter image texture. One method (follow-up<sub>32×32</sub>) achieved significantly lower  $nBias_{avg}$  than the other approaches, indicating that feature values were minimally perturbed.  $nBias_{avg}$  for this method was also smaller than in a previous study where features were calculated directly on demons-deformed scans. Compared with the method that achieved lowest variability (affine<sub>32×32</sub>),  $nRoA$  values for follow-up<sub>32×32</sub> ROIs were not significantly higher, indicating low variability in feature value differences between scans with this approach. Twenty features with both low  $nRoA$  and  $nBias_{avg}$  were identified for use with the follow-up<sub>32×32</sub> method. In future studies of patients with lung disease, this method may facilitate quantitative measurement of differences between serial CT scans due to disease progression or response to treatment, allowing for improved detection and characterization of pathologic change.

## ACKNOWLEDGMENTS

This work was supported, in part, by the Coleman Endowment through the University of Chicago Comprehensive Cancer Center, NSF REU Award No. 1062909, and NIH Grant Nos. S10 RR021039, P30 CA14599, and T32 EB002103-23.

S.G.A. receives royalties and licensing fees related to computer-aided diagnosis technology through the University of Chicago and research funding from Riverain Technologies through the University of Chicago. The authors would like to thank Kristen Wroblewski, the University of Chicago Department of Health Studies, for her guidance in statistical analysis.

- <sup>a)</sup>Presented in part at the 2013 Annual SPIE Conference in Medical Imaging.
- <sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: hal-hallaq@radonc.uchicago.edu; Fax: 773-832-3309 and 773-834-7299.
- <sup>1</sup>Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.-T. Zhang, K. Suzuki, and D. Doi, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Med. Phys.* **30**, 2440–2454 (2003).
- <sup>2</sup>R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *Am. J. Respir. Crit. Care Med.* **160**, 648–654 (1999).
- <sup>3</sup>S. Delorme, M. A. Keller-Reichenbecher, I. Zuna, W. Schlegel, and G. Van Kaick, "Usual interstitial pneumonia: Quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis," *Invest. Radiol.* **32**, 566–574 (1997).
- <sup>4</sup>J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," *Acad. Radiol.* **18**, 306–314 (2011).
- <sup>5</sup>B. Ganeshan, S. Abaleke, R. C. D. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imaging* **10**, 137–143 (2010).
- <sup>6</sup>M. Staring, S. Klein, J. H. C. Reiber, W. J. Niessen, and B. C. Stoel, "Pulmonary image registration with elastix using a standard intensity-based algorithm," *Medical Image Analysis for the Clinic: A Grand Challenge, in Proceedings of the Workshop of MICCAI*, Beijing, China, 2010.
- <sup>7</sup>Z. Wu, E. Rietzel, V. Boldea, D. Sarrut, and G. C. Sharp, "Evaluation of deformable registration of patient lung 4D CT with subanatomical region segmentations," *Med. Phys.* **35**, 775–781 (2008).
- <sup>8</sup>G. C. Sharp, N. Kandasamy, H. Singh, and M. Folkert, "GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration," *Phys. Med. Biol.* **52**, 5771–5783 (2007).
- <sup>9</sup>S. S. Samant, J. Xia, P. Muyan-Ozcelik, and J. D. Owens, "High performance computing for deformable image registration: Towards a new paradigm in adaptive radiotherapy," *Med. Phys.* **35**, 3546–3553 (2008).
- <sup>10</sup>D. A. Palma, J. van Sornsen de Koste, W. F. A. R. Verbakel, A. Vincent, and S. Senan, "Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients," *Int. J. Radiat. Oncol., Biol., Phys.* **81**, 974–978 (2011).
- <sup>11</sup>Y. Arzhaeva, M. Prokop, K. Murphy, E. M. van Rikxoort, P. A. de Jong, H. A. Gietema, M. A. Viergever, and B. van Ginneken, "Automated estimation of progression of interstitial lung disease in CT images," *Med. Phys.* **37**, 63–73 (2010).
- <sup>12</sup>A. R. Cunliffe, H. A. Al-Hallaq, Z. E. Labby, C. A. Pelizzari, C. S. Straus, W. F. Sensakovic, M. Ludwig, and S. G. Armato III, "Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration," *Med. Phys.* **39**, 4679–4690 (2012).
- <sup>13</sup>A. R. Cunliffe, H. A. Al-Hallaq, X. M. Fei, R. E. Tuohy, and S. G. Armato III, "Comparison of demons deformable registration-based methods for texture analysis of serial thoracic CT scans," *Proc. SPIE* **8670**, 86700D-1–86700D-6 (2013).
- <sup>14</sup>J. M. Bland and D. G. Altman, "Agreement between methods of measurement with multiple observations per individual," *J. Biopharm. Stat.* **17**, 571–582 (2007).
- <sup>15</sup>R. M. Haralick, S. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**, 610–621 (1973).
- <sup>16</sup>W. K. Pratt, "Image feature extraction" *Digital Image Processing*, 3rd ed. (Wiley, New York, 2001), pp. 509–550.
- <sup>17</sup>T. Wagner, "Texture analysis" in *Handbook of Computer Vision and Applications*, edited by B. Jahne, H. Hanssecker, and P. Geissler (Academic, San Diego, 1999), Vol. 2, pp. 275–308.
- <sup>18</sup>H. Li, M. L. Giger, Z. Huo, O. I. Olopade, L. Lan, B. L. Weber, and I. Bonta, "Computerized analysis of mammographic parenchymal patterns for



- assessing breast cancer risk: Effect of ROI size and location,” *Med. Phys.* **31**, 549–555 (2004).
- <sup>19</sup>S. Peleg, J. Naor, R. Hartley, and D. Avnir, “Multiple resolution texture analysis and classification,” *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 518–523 (1984).
- <sup>20</sup>C.-C. Chen, J. S. DaPonte, and M. D. Fox, “Fractal feature analysis and classification in medical imaging,” *IEEE Trans. Med. Imaging* **8**, 133–142 (1989).
- <sup>21</sup>J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jone, and M. J. Yaffe, “Automated analysis of mammographic densities,” *Phys. Med. Biol.* **41**, 909–923 (1996).
- <sup>22</sup>R. Creutzburg and E. Ivanov, “Fast algorithm for computing fractal dimensions of image segments,” in *Recent Issues in Pattern Analysis and Recognition*, Lecture Notes in Computer Science Vol. 399, edited by V. Cantoni, R. Creutzburg, S. Levialdi, and G. Wolf (Springer, Berlin, 1989), pp. 42–51.
- <sup>23</sup>S. Katsuragawa, K. Doi, and H. MacMahon, “Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung disease in digital chest radiographs,” *Med. Phys.* **15**(3), 311–319 (1988).
- <sup>24</sup>K. I. Laws, “Textured image segmentation,” USCIP Technical Report No. 940 (University of Southern California, Los Angeles, CA, 1980).
- <sup>25</sup>D. A. Palma, J. R. Van Sornsed de Koste, W. F. A. R. Verbakel, and S. Senan, “A new approach to quantifying lung damage after stereotactic body radiation therapy,” *Acta Oncol.* **50**, 509–517 (2011).