Nodule Characteristics to Consider:

- Location
- Volume
- Shape
- Density
- Texture

Due to time constraints I am only considering location, volume, and shape of the nodules.

I have come up with two different equations that assign weights to the different characteristics. I think these would have some success in determining whether scans are similar.

General Equation:

Similarity Measure = $aV_e + bL_e + cS_e$

where a, b, and c are weights assigned to each characteristic and V_e is volume error, L_e is location error and S_e is shape error

To calculate V_e:

$$Ve = \frac{\sum \frac{Vc - Vt}{Vc}}{N}$$

where V_c is the volume of the current nodule in the query scan, V_t is the volume of the corresponding nodule in the current test scan , and N is the number of nodules in the scan (query scan or current test scan) with the most nodules.

To calculate L_e :

$$Le = \frac{\sum \frac{Lc - Lt}{Lc}}{N}$$

where L_c is the location of the current nodule, L_t is the location of the corresponding nodule in the current test scan, and N is the number of nodules in the scan (query scan or current test scan) with the most nodules.

To calculate S_e:

First calculate S (the sphere occupancy) for each nodule:

$$S = \frac{3V}{4\pi r^3}$$

where V is the volume and r is the radius.

Then Calculate S_e:

$$Se = \frac{\sum \frac{Sc - Lt}{Lc}}{N}$$

where S_c is the sphere occupancy for the current nodule, S_t is the sphere occupancy of the corresponding nodule in the current test scan, and N is the number of nodules in the scan (query scan or current test scan) with the most nodules.

General Equation Continued…

- Each element in the equation (V_e, S_e, and L_e) should yield a value between 0 and 1.
- Given the weights that I define in the following equations the similarity measure is between 0 and 10. A measure of 0 should result from comparing a scan to itself while a measure of 10 should occur when 2 very different scans are compared.
- N is chosen as the number of nodules in the scan (query scan or current test scan) with the most nodules so that a penalty is added when the number of nodules in the scans differ. I was assuming that an error of 1 is added for V_e, S_e, and L_e when no corresponding nodule exists in the test scan or when there is an extra nodule in the test scan.

Equation 1:

$$\text{Similarity Measure} = 3Ve + 5Le + 2Se$$

In this equation the location of each nodule is the most important characteristic. I think that this equation will produce scans that are visually more similar to the query scan than Equation 2, but I have not tested either equation yet. Because the location is assigned the highest weight, not as much emphasis will be put on size and location so at first glance I the scans produced by the query should appear to be similar.

Equation 2:

$$\text{Similarity Measure} = 5Ve + 2Le + 3Se$$

In this equation the most important characteristic is the volume of each nodule. I think that this equation should produce scans that are less visually similar than those in equation 1 but more similar in terms of patient diagnosis, however, this hypothesis still needs to be tested. Because nodule volume and shape are assigned higher weights, scans returned by the query may not look as much like the query scan but the nodules in the scans hopefully will be similar. If two scans have similar nodules it is likely that the same disease is present in both scans.

Other Things to Consider:

- Texture and density of the nodules
- More complex shape measures (i.e. measuring smoothness of the nodule boundary)
- More complex location measures (i.e. distances from landmarks to nodules)
- These simple equations only consider nodule characteristics. Measure the shape and/or volume of the lung containing the nodules
- Effects of differences in number of nodules in the query scan and scans produced by query