DEFINITION OF TEXTURE FEATURES

Input volume: Volume of interest $V(x,y,z)$ with isotropic voxel size. The necessity for isotropically resampling V to a given voxel dimension prior to texture analysis results from the fact that: 1) *Global* features are computed from histograms counting the number of gray-levels in 3D space; and 2) all higher-order texture measurements explicitly or implicitly involve a distance parameter in the matrix computation (GLCM, GLRLM, GLSZM and NGTDM). For this parameter to be meaningful in 3D space and in order for the orientation dependence of the tumour to be minimized, isotropic resolution is required.

Global texture features (first-order gray-level statistics).
Let P define the first-order histogram of a volume $V(x,y,z)$ with isotropic voxel size. $P(i)$ represents the number of voxels with gray-level i, and N_g represents the number of gray-level bins set for P. The i^{th} entry of the normalized histogram is then defined as:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)}.$$

The *Global* texture features are then defined as:

- Variance:

$$\sigma^2 = \sum_{i=1}^{N_g} (i-\mu)^2 \, p(i)$$

- Skewness:

$$s = \sigma^{-3} \sum_{i=1}^{N_g} (i-\mu)^3 \, p(i)$$

- Kurtosis

$$k = \sigma^{-4} \sum_{i=1}^{N_g} \left[(i-\mu)^4 \, p(i) \right] - 3$$

Gray-Level Co-occurence Matrix (GLCM) texture features.
Let P define the GLCM of a quantized volume $V(x,y,z)$ with isotropic voxel size. $P(i,j)$ represents the number of times voxels of gray-level i were neighbours with voxels of gray-level j in V, and N_g represents the pre-defined number of quantized gray-levels set in V. Only one GLCM of size $N_g \times N_g$ is computed per volume V by simultaneously adding up the frequency of co-occurences of all voxels with their 26-connected neighbours in 3D space, with all voxels (*including*

the peripheral region) considered once as a center voxel (according to [1], thus always using $d = 1$). To account for discretization length differences, neighbours at a distance of $\sqrt{3}$ voxels around a center voxel increment the GLCM by a value of $\sqrt{3}$, neighbours at a distance of $\sqrt{2}$ voxels around a center voxel increment the GLCM by a value of $\sqrt{2}$, and neighbours at a distance of 1 voxel around a center voxel increment the GLCM by a value of 1. The entry (i, j) of the of the normalized GLCM is then defined as:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j)}.$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} p(i,j), \qquad \mu_j = \sum_{j=1}^{N_g} j \sum_{i=1}^{N_g} p(i,j),$$

$$\sigma_i = \sum_{i=1}^{N_g} (i-\mu_i)^2 \sum_{j=1}^{N_g} p(i,j), \qquad \sigma_j = \sum_{j=1}^{N_g} (j-\mu_j)^2 \sum_{i=1}^{N_g} p(i,j).$$

The GLCM texture features are then defined as:

- Energy [1]:

$$energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i,j)]^2$$

- Contrast [1]:

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 \, p(i,j)$$

- Correlation (adapted from [1]):

$$correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i-\mu_i)\,(j-\mu_j)\,p(i,j)}{\sigma_i\,\sigma_j}$$

- Homogeneity (adapted from [1]):

$$homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1+|i-j|}$$

- Variance (adapted from [1]):

$$variance = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left[(i-\mu_i)^2\, p(i,j) + (j-\mu_j)^2\, p(i,j)\right]$$

- Sum Average (adapted from [1]):

$$sum\ average = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i\, p(i,j) + j\, p(i,j)]$$

- Entropy [1]:

$$entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2\big(p(i,j)\big)$$

- Dissimilarity [2]:

$$dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j|\, p(i,j)$$

Gray-Level Run-Length Matrix (GLRLM) texture features.
Let P define the GLRLM of a quantized volume $V(x,y,z)$ with isotropic voxel size. $P(i,j)$ represents the number of runs of gray-level i and of length j in V, N_g represents the pre-defined number of quantized gray-levels set in V, and L_r represents the length of the longest run (of any gray-level) in V. Only one GLRLM of size $N_g \times L_r$ is computed per volume V by simultaneously adding up all possible longest run-lengths in the 13 directions of 3D space (one voxel can be part of multiple runs in different directions, but can be part of only one run in a given direction). A MATLAB toolbox created by Xunkai Wei [8] computes GLRLMs from 2D images, and it can be used to facilitate the computation of GLRLMs from 3D volumes. To account for discretization length differences, runs constructed from voxels separated by a distance of $\sqrt{3}$ increment the GLRLM by a value of $\sqrt{3}$, runs constructed from voxels separated by a distance of $\sqrt{2}$ increment the GLRLM by a value of $\sqrt{2}$, and runs constructed from voxels separated by a distance of 1 increment the GLRLM by a value of 1. The entry (i,j) of the of the normalized GLRLM is then defined as:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} P(i,j)}.$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_r} p(i,j), \qquad \mu_j = \sum_{j=1}^{L_r} j \sum_{i=1}^{N_g} p(i,j).$$

The GLRLM texture features are then defined as:

- Short Run Emphasis (SRE) [3]:

$$SRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{p(i,j)}{j^2}$$

- Long Run Emphasis (LRE) [3]:

$$LRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} j^2 \, p(i,j)$$

- Gray-Level Nonuniformity (GLN) (adapted from [3]):

$$GLN = \sum_{i=1}^{N_g} \left(\sum_{j=1}^{L_r} p(i,j) \right)^2$$

- Run-Length Nonuniformity (RLN) (adapted from [3]):

$$RLN = \sum_{j=1}^{L_r} \left(\sum_{i=1}^{N_g} p(i,j) \right)^2$$

- Run Percentage (RP) (adapted from [3]):

$$RP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} p(i,j)}{\sum_{j=1}^{L_r} j \sum_{i=1}^{N_g} p(i,j)}$$

- Low Gray-Level Run Emphasis (LGRE) [5]:

$$LGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{p(i,j)}{i^2}$$

- High Gray-Level Run Emphasis (HGRE) [5]:

$$HGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} i^2 \, p(i,j)$$

- Short Run Low Gray-Level Emphasis (SRLGE) [6]:

$$SRLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{p(i,j)}{i^2 j^2}$$

- Short Run High Gray-Level Emphasis (SRHGE) [6]:

$$SRHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{i^2\, p(i,j)}{j^2}$$

- Long Run Low Gray-Level Emphasis (LRLGE) [6]:

$$LRLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{j^2\, p(i,j)}{i^2}$$

- Long Run High Gray-Level Emphasis (LRHGE) [6]:

$$LRHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} i^2 j^2\, p(i,j)$$

- Gray-Level Variance (GLV) (adapted from [7]):

$$GLV = \frac{1}{N_g \times L_r} \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \left(i\, p(i,j) - \mu_i\right)^2$$

- Run-Length Variance (RLV) (adapted from [7]):

$$RLV = \frac{1}{N_g \times L_r} \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \left(j\, p(i,j) - \mu_j\right)^2$$

Gray-Level Size Zone Matrix (GLSZM) texture features.
Let P define the GLSZM of a quantized volume $V(x, y, z)$ with isotropic voxel size. $P(i, j)$ represents the number of 3D zones of gray-levels i and of size j in V, N_g represents the pre-defined number of quantized gray-levels set in V, and L_z represents the size of the largest zone (of any gray-level) in V. One GLSZM of size $N_g \times L_z$ is computed per volume V by adding up all possible largest zone-sizes, with zones constructed from 26-connected neighbours of the same gray-level in 3D space (one voxel can be part of only one zone). The entry (i, j) of the normalized GLSZM is then defined as:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} P(i,j)}.$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_z} p(i,j), \qquad \mu_j = \sum_{j=1}^{L_z} j \sum_{i=1}^{N_g} p(i,j).$$

The GLSZM texture features are then defined as:

- Small Zone Emphasis (SZE) [3, 7]:

$$SZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i,j)}{j^2}$$

- Large Zone Emphasis (LZE) [3, 7]:

$$LZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} j^2 \, p(i,j)$$

- Gray-Level Nonuniformity (GLN) (adapted from [3, 7]):

$$GLN = \sum_{i=1}^{N_g} \left(\sum_{j=1}^{L_z} p(i,j) \right)^2$$

- Zone-Size Nonuniformity (ZSN) (adapted from [3, 7]):

$$ZSN = \sum_{j=1}^{L_z} \left(\sum_{i=1}^{N_g} p(i,j) \right)^2$$

- Zone Percentage (RP) (adapted from [3, 7]):

$$ZP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} p(i,j)}{\sum_{j=1}^{L_z} j \sum_{i=1}^{N_g} p(i,j)}$$

- Low Gray-Level Zone Emphasis (LGZE) [5, 7]:

$$LGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i,j)}{i^2}$$

- High Gray-Level Zone Emphasis (HGZE) [5, 7]:

$$HGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} i^2 \, p(i,j)$$

- Small Zone Low Gray-Level Emphasis (SZLGE) [6, 7]:

$$SZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i,j)}{i^2 j^2}$$

- Small Zone High Gray-Level Emphasis (SZHGE) [6, 7]:

$$SZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{i^2\, p(i,j)}{j^2}$$

- Large Zone Low Gray-Level Emphasis (LZLGE) [6, 7]:

$$LZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{j^2\, p(i,j)}{i^2}$$

- Large Zone High Gray-Level Emphasis (LZHGE) [6, 7]:

$$LZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} i^2 j^2\, p(i,j)$$

- Gray-Level Variance (GLV) (adapted from [7]):

$$GLV = \frac{1}{N_g \times L_z} \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \left(i\, p(i,j) - \mu_i\right)^2$$

- Zone-Size Variance (ZSV) (adapted from [7]):

$$ZSV = \frac{1}{N_g \times L_z} \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \left(j\, p(i,j) - \mu_j\right)^2$$

Neighbourhood Gray-Tone Difference Matrix (NGTDM) texture features.

Let $P(i)$ define the NGTDM of a quantized volume $V(x,y,z)$ with isotropic voxel size. $P(i)$ represents the summation of the gray-level differences between all voxels with gray-level i and the average gray-level of their 26-connected neighbours in 3D space. N_g represents the pre-defined number of quantized gray-levels set in V, and $(N_g)_{eff}$ is the effective number of gray-levels in V, with $(N_g)_{eff} < N_g$ (let the vector of gray-levels values in V be denoted as $\mathbf{g} = g(1), g(2), \ldots, g(N_g)$; some gray-levels excluding $g(1)$ and $g(N_g)$ may not appear in V due to different quantization schemes). One NGTDM of size $N_g \times 1$

is computed per volume V. To account for discretization length differences, all averages around a center voxel located at position (j,k,l) in V are performed such that the neighbours at a distance of $\sqrt{3}$ voxels are given a weight of $1/\sqrt{3}$, the neighbours at a distance of $\sqrt{2}$ voxels are given a weight of $1/\sqrt{2}$, and the neighbours at a distance of 1 voxel are given a weight of 1. The i^{th} entry of the NGTDM is then defined as:

$$P(i) = \begin{cases} \sum_{\text{all voxels} \in \{N_i\}} |i - \overline{A}_i| & \text{if } N_i > 0, \\ 0 & \text{if } N_i = 0. \end{cases}$$

where $\{N_i\}$ is the set of all voxels with gray-level i in V (*including* the peripheral region), N_i is the number of voxels with gray-level i in V, and $\overline{A}_i$ is the average gray-level of the 26-connected neighbours around a center voxel with gray-level i and located at position (j,k,l) in V such that:

$$\overline{A}_i = \overline{A}(j,k,l) = \frac{\sum_{m=-1}^{m=1}\sum_{n=-1}^{n=1}\sum_{o=-1}^{o=1} w_{m,n,o} \cdot V(j+m,k+n,l+o)}{\sum_{m=-1}^{m=1}\sum_{n=-1}^{n=1}\sum_{o=-1}^{o=1} w_{m,n,o}},$$

$$\text{where} \quad w_{m,n,o} = \begin{cases} 1 & \text{if } |j-m|+|k-n|+|l-o| = 1, \\ \frac{1}{\sqrt{2}} & \text{if } |j-m|+|k-n|+|l-o| = 2, \\ \frac{1}{\sqrt{3}} & \text{if } |j-m|+|k-n|+|l-o| = 3, \\ 0 & \text{if } V(j+m,k+n,l+o) \text{ is undefined.} \end{cases}$$

The following quantity is also defined:

$$n_i = \frac{N_i}{N}.$$

where N is the total number of voxels in V. The NGTDM texture features are then defined as:

- Coarseness [4]:

$$coarseness = \left[\epsilon + \sum_{i=1}^{N_g} n_i\, P(i)\right]^{-1}$$

  where ϵ is a small number to prevent *coarseness* becoming infinite.

- Contrast [4]:

$$contrast = \left[\frac{1}{(N_g)_{eff}\left[(N_g)_{eff} - 1\right]}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} n_i\, n_j\,(i-j)^2\right]\left[\frac{1}{N}\sum_{i=1}^{N_g} P(i)\right]$$

- Busyness [4]:

$$busyness = \frac{\sum_{i=1}^{N_g} n_i\, P(i)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i\, n_i - j\, n_j)}, \quad n_i \neq 0, n_j \neq 0$$

- Complexity [4]:

$$complexity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{|i-j| \left[n_i \, P(i) + n_j \, P(j) \right]}{N \, (n_i + n_j)}, \quad n_i \neq 0, n_j \neq 0$$

- Strength [4]:

$$strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (n_i + n_j) \, (i-j)^2}{\left[\epsilon + \sum_{i=1}^{N_g} P(i) \right]}, \quad n_i \neq 0, n_j \neq 0$$

  where ϵ is a small number to prevent *strength* becoming infinite.

Online resources

MATLAB® software code is freely shared under the GNU General Public License at: `https://github.com/mvallieres/radiomics`.

References

[1] Haralick, R.M., Shanmugam, K. and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, smc **3**(6), 610-621.

[2] Thibault, G. (2009). Indices de formes et de textures: de la 2D vers la 3D. Application au classement de noyaux de cellules. *PhD Thesis*, Université AIX-Marseille: p.172.

[3] Galloway, M.M. (1975). Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, **4**(2), 172-179.

[4] Amadasun, M. and King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**(5), 1264-1274.

[5] Chu, A. Sehgal, C.M. and Greenleaf, J.F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, **11**(6), 415-419.

[6] Dasarathy, B.V. and Holder, E.B. (1991). Image characterization based on joint gray level-run length distributions. *Pattern Recognition Letters*, **12**(8), 497-502.

[7] Thibault, G. et al. (2009). Texture indexes and gray level size zone matrix. Application to cell nuclei classification. In *Pattern Recognition and Information Processing*. Minsk, Belarus, 140-145.

[8] Wei, X. (2007). Gray Level Run Length Matrix Toolbox v1.0, computer software. Beijing Aeronautical Technology Research Center. Available from: http://www.mathworks.com/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox. [11 January 2013]