



<http://www.diva-portal.org>

This is the published version of a paper published in *Physics in Medicine and Biology*.

Citation for the original published paper (version of record):

Garpebring, A., Brynolfsson, P., Kuess, P., Georg, D., Helbich, T H. et al. (2018)
Density Estimation of Grey-Level Co-Occurrence Matrices for Image Texture Analysis
Physics in Medicine and Biology, 63(19): 9-15
<https://doi.org/10.1088/1361-6560/aad8ec>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-152488>

OPEN ACCESS



PAPER

Density estimation of grey-level co-occurrence matrices for image texture analysis

RECEIVED
17 April 2018REVISED
24 July 2018ACCEPTED FOR PUBLICATION
8 August 2018PUBLISHED
2 October 2018

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Anders Garpebring¹, Patrik Brynolfsson¹, Peter Kuess^{2,4}, Dietmar Georg^{2,4}, Thomas H Helbich^{3,4}, Tufve Nyholm¹ and Tommy Löfstedt^{1,5}¹ Department of Radiation Sciences, Umeå University, Umeå, Sweden² Department of Radiotherapy, Medical University of Vienna, Vienna, Austria³ Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria⁴ Christian Doppler Laboratory for Medical Radiation Research for Radiation Oncology, Vienna, Austria⁵ Author to whom any correspondence should be addressed.E-mail: tommy.lofstedt@umu.se**Keywords:** Haralick features, invariant features, GLCM, density estimation, texture analysis, image analysisSupplementary material for this article is available [online](#)**Abstract**

The Haralick texture features are common in the image analysis literature, partly because of their simplicity and because their values can be interpreted. It was recently observed that the Haralick texture features are very sensitive to the size of the GLCM that was used to compute them, which led to a new formulation that is invariant to the GLCM size. However, these new features still depend on the sample size used to compute the GLCM, i.e. the size of the input image region-of-interest (ROI).

The purpose of this work was to investigate the performance of density estimation methods for approximating the GLCM and subsequently the corresponding invariant features.

Three density estimation methods were evaluated, namely a piece-wise constant distribution, the Parzen-windows method, and the Gaussian mixture model. The methods were evaluated on 29 different image textures and 20 invariant Haralick texture features as well as a wide range of different ROI sizes.

The results indicate that there are two types of features: those that have a clear minimum error for a particular GLCM size for each ROI size, and those whose error decreases monotonically with increased GLCM size. For the first type of features, the Gaussian mixture model gave the smallest errors, and in particular for small ROI sizes (less than about 20×20).

In conclusion, the Gaussian mixture model is the preferred method for the first type of features (in particular for small ROIs). For the second type of features, simply using a large GLCM size is preferred.

1. Introduction

In their seminal paper from 1973, Haralick *et al* (1973) introduced 14 texture feature descriptors computed from grey-level co-occurrence matrices (GLCMs, containing the joint frequency of neighbouring pixel grey-levels). These were used in the classification of photomicrographs, aerial, and satellite image data. The Haralick texture features were computed as functions of a GLCM. Such features are simple to implement, easily computed, and result in a set of interpretable and easy-to-understand texture descriptors. Each Haralick feature captures a different aspect of the GLCM, and the features taken together are assumed to capture the characteristics of a particular texture in an image. Thus, the texture features are useful when e.g. classifying either whole images, or regions in images, based on texture characteristics.

The Haralick texture features have been applied in many different fields. For instance, they have been used in many aspects of remote sensing applications, such as forest inventory (Tokola and Kilpelainen 1999), land-use and forest-type classification (Ulaby *et al* 1986) and tree species identification (Kukunda *et al* 2018); in automatic detection of pollen (Marcos *et al* 2015), finger print analysis (Agarwal *et al* 2016), classification of hydrophobicity

in polymers insulators (Thomazini *et al* 2012), classification of lamellar graphite in cast iron (Roberts *et al* 2005), and have even been compared favourably to deep learning methods (Basu *et al* 2016).

Further, they have been used in medical image analysis, in the analysis of ultrasound and MRI images of the liver (Lerski *et al* 1979, Mayerhoefer *et al* 2010), the heart (Skorton *et al* 1983); in the analysis of x-ray mammographic images (Chan *et al* 1995, Soltanian-Zadeh *et al* 2004, Mustra *et al* 2012, Garma and Hassan 2014); and in the analysis of MRI images in the study of breast cancer (Chen *et al* 2007, Nie *et al* 2008, Ahmed *et al* 2013, Nagarajan *et al* 2013a, 2013b, Gensheimer *et al* 2015, Prasanna *et al* 2016, Wu *et al* 2016) and of brain cancer (Georgiadis *et al* 2009, Assefa *et al* 2010, Eliat *et al* 2012).

There are of course other features and methods that can be used to describe image texture and to classify texture, such as e.g. fractal dimensions (Lopes *et al* 2011), wavelets (Livens *et al* 1997), local binary patterns (Ojala *et al* 1994) and Gabor filters (Zacharaki *et al* 2009), and not least convolutional neural networks (Le Cun *et al* 1998). However, the Haralick texture features, computed from GLCMs, have been very successful and are still widely used, in large part due to their simplicity and intuitive interpretations.

When creating a GLCM from a region in an image, the number of grey-levels in the image is often reduced in a process called quantisation. It turns out that the Haralick texture feature values depend strongly on the image quantisation (Brynnolfsson *et al* 2017), which means that features cannot be compared directly between images when different numbers of grey-levels have been used, even when they are computed on images with the same underlying texture.

The number of grey-levels determines the size of the GLCM, and this in turn influences the feature values computed from the GLCM. This relation was investigated by Löfstedt *et al* (2017), and an approach was proposed to construct similar features that are asymptotically invariant to the number of grey-levels. Although this is a very attractive property, the result may still depend heavily on the size of the GLCM if the GLCM is constructed from an insufficient number of samples (i.e. small ROIs).

The purpose of this work was therefore to evaluate whether the GLCM can be better estimated as a density (the alternative interpretation of the GLCM that was employed by Löfstedt *et al* (2017)) to construct GLCMs for different sample sizes and to see how these estimates affect the (asymptotically invariant) Haralick features. A particular focus was on how the methods performed when the sample sizes were small. This is an important domain, since the invariant texture features have their largest errors for small sample sizes (Löfstedt *et al* 2017); hence, reduced errors for small sample sizes may improve e.g. radiomic or other clinical applications, where small ROIs are common (e.g. small tumours).

Three density estimation methods were evaluated in terms of the errors of the resulting feature values compared to ground truth values. The methods were evaluated on constructed texture images (the Kylberg texture dataset v.1.0 (Kylberg 2011)) and a real prostate cancer data set (Kuess *et al* 2017).

2. Theory

In this section, we will give a brief description of the GLCM, the original Haralick texture features, and the invariant Haralick texture features proposed by Löfstedt *et al* (2017). Finally, we will describe the theory and the details regarding the density estimation methods we investigate, with which we can avoid having to select the number of grey-levels to use.

An overview of how to produce GLCMs from image data can be found in e.g. Brynnolfsson *et al* (2017) and Löfstedt *et al* (2017).

2.1. The Haralick texture features

The Haralick texture features are functions of a normalised GLCM (from now on denoted simply as GLCM), $\mathbf{P} = \mathbf{X}/N \in \mathbb{R}^{n \times n}$, where \mathbf{X} is the un-normalised GLCM and $N = \sum_i \sum_j X_{ij}$ is the *sample size*. The Haralick texture features can all be written on the general form (Löfstedt *et al* 2017)

$$f = \sum_{i=a}^A \sum_{j=b}^B \phi(i, j, \mathbf{g}(\mathbf{P})) \psi(P_{ij}), \quad (1)$$

where ψ is a function of the elements of the GLCM, P_{ij} , and ϕ is a function of the indices and a vector-valued function, \mathbf{g} , of the whole GLCM. For instance, the function \mathbf{g} could return the marginal means, μ_x and μ_y , and marginal standard deviations, σ_x and σ_y , of the GLCM, as in the feature *Correlation*, f_{corr} , (see Löfstedt *et al* (2017)). In that case we have

$$a = b = 1, \quad A = B = n, \quad \mathbf{g}(\mathbf{P}) = (\mu_x, \mu_y, \sigma_x, \sigma_y),$$

$$\phi(i, j, \mathbf{g}(\mathbf{P})) = \left(\frac{i - \mu_x}{\sigma_x} \right) \left(\frac{j - \mu_y}{\sigma_y} \right) \quad \text{and} \quad \psi(P_{ij}) = P_{ij},$$

Table 1. The identified two types of features. Those in the left column have a clear minimum relative RMSE, while those in the right column benefit when using the largest possible GLCM size.

Clear optimum	Decreases monotonically
Angular second moment	Autocorrelation
Difference entropy	Cluster shade
Entropy	Cluster prominence
Information measure of correlation 1	Contrast
Information measure of correlation 2	Correlation
Maximal correlation coefficient	Difference average
Sum entropy	Difference variance
Maximum probability	Inverse difference
	Inverse difference moment
	Sum average
	Sum of squares: variance
	Sum variance

and hence

$$f_{\text{corr}} = \sum_{i=a}^A \sum_{j=b}^B \phi(i, j, \mathbf{g}(\mathbf{P})) \psi(P_{ij}) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{i - \mu_x}{\sigma_x} \right) \left(\frac{j - \mu_y}{\sigma_y} \right) P_{ij}.$$

The Haralick texture features used in this work were the same as those used in Löfstedt *et al* (2017). We thus evaluated 20 features in total. The features evaluated are also listed in table 1.

2.2. The invariant texture features

The asymptotically invariant texture features, proposed in Löfstedt *et al* (2017), interprets the GLCM as a discrete approximation of a probability density function (i.e. as piece-wise constant densities), instead of as a mass function. In this interpretation, the features are expressed as integrals over functions of this distribution by

$$\int_0^1 \int_0^1 \phi(i, j, \mathbf{g}(p^*)) \psi(p^*(i, j)) \, djdi, \tag{2}$$

where p^* is the true underlying probability density function. The normalised GLCM must therefore integrate to 1, i.e.

$$\int_0^1 \int_0^1 p^*(i, j) \, djdi = 1.$$

Note that the bounds of the integrations are arbitrary, but also that they must be independent of n , the size of the GLCM. Ideally they should be constant. In this study, they were set to 0 and 1 for simplicity.

The integral in equation (2) can be approximated by a Riemann sum as

$$\int_0^1 \int_0^1 \phi(i, j, \mathbf{g}(p^*)) \psi(p^*(i, j)) \, djdi \approx \sum_{i=1}^n \sum_{j=1}^n \phi\left(\frac{i}{n}, \frac{j}{n}, \mathbf{g}(\tilde{\mathbf{P}})\right) \psi(\tilde{P}_{ij}) \Delta_j \Delta_i,$$

where

$$\Delta_j = \Delta_i = \frac{1}{n},$$

and

$$\tilde{\mathbf{P}} = \frac{\mathbf{P}}{\Delta_i \Delta_j}. \tag{3}$$

In short, the indices, i and j , are normalised to the half-open interval $(0, 1]$, the summands are multiplied by a differential, and the GLCM is normalised such that its Riemann sum is 1.

The invariant texture features are presented together with the corresponding original features in table 2 in Löfstedt *et al* (2017).

While the approximation of the integral depends on the discretisation of the distribution, (i.e. the number of grey-levels), the approximation is equal to the integral in the limit as $n \rightarrow \infty$, assuming that $\phi(i, j, \mathbf{g}(p^*)) \psi(p^*(i, j))$ is Riemann integrable. The original Haralick features do not have this property that they

give the same approximated values for different discretisations, and in fact, many of the original Haralick features either approach zero or infinity when $n \rightarrow \infty$, as was illustrated in Löfstedt *et al* (2017).

3. Methods

Three density estimation methods were compared to the traditional approach, where the number of grey-levels is set to an arbitrary fixed value, in order to see which had the smallest expected error when used to estimate the invariant Haralick texture features for a range of different sample sizes (or equivalently, image ROI sizes). The methods differ in the step where the GLCM is estimated. GLCM estimation is, in the new interpretation, essentially the problem of estimating a probability density function from data. Therefore, two parametric density estimation methods, a method that automatically selects an appropriate GLCM size (equivalent to a piece-wise constant density), and the traditional approach (with fixed numbers of grey-levels) were compared to the true values. The evaluated methods were:

- Fixed number of grey-levels, which corresponds to the traditional approach of *a priori* selecting a fixed number of grey-levels.
- Variable number of grey-levels, where the numbers of grey-levels are estimated using cross-validation. This is more flexible than selecting a fixed value and thus have the potential to give better results.
- Parzen windows, a non-parametric method that estimates a density using kernel smoothing.
- Gaussian mixture model, which is a parametric model that estimates a density as a sum of weighted Gaussians.

Finally, the methods were compared to an approach that we refer to as the *Optimal method*. In this case, the number of grey-levels were found by comparing the computed feature values to feature values computed from a GLCM generated from a very large number of samples; the number of grey-levels that minimised the error relative to the estimated true feature value was selected by this method. This method requires knowledge about the true feature values, but since those are rarely available in practical applications, an approximation computed from a very large number of samples was assumed to be sufficient when used here as a benchmark.

3.1. Evaluation metric

A metric of the methods' performance was required in order to evaluate them. For this purpose we used the relative root mean-squared error (relative RMSE) of the estimated texture features. Let f_i , for $i = 1, \dots, 20$, be estimates of the 20 different invariant Haralick texture features investigated in this work. Further, let f_i^k be a sample of the random variable $f_i \sim p(f_i)$, then the relative RMSE of that feature is defined as

$$\begin{aligned} \text{RelRMSE}(f_i) &:= \frac{\sqrt{\mathbb{E}[(f_i - f_i^*)^2]}}{|\mathbb{E}[f_i]|} \\ &\approx \frac{\sqrt{\frac{1}{K} \sum_{k=1}^K (f_i^k - f_i^*)^2}}{|f_i^*|}, \end{aligned} \quad (4)$$

where K is the number of samples and $f_i^* = \mathbb{E}[f_i]$ is the true value of the parameter we want to estimate.

This metric was selected since the scale of different features, f_i , can vary substantially, both between features and depending on the texture in the analysed images. The magnitudes of the RMSE are thus less important than their relative sizes.

3.2. Data

Two data sets were used in the evaluation of the methods. These were the *Kylberg texture dataset v.1.0* (Kylberg 2011) and a prostate cancer data set (Kuess *et al* 2017). The first data set was selected because of its size and diversity, with many different textures and a large number of images for each texture. The prostate data was selected to represent a typical medical image application.

3.2.1. The Kylberg texture dataset

This image data set contains 28 texture classes with 160 images in each class, capturing the textures of rice, sand, a floor, etc. Figure 2 contains five example images and corresponding GLCM shapes, produced from the Kylberg data.

3.2.2. The prostate cancer data set

This data set contained 25 patients diagnosed with prostate cancer. The patients were a sub-cohort from a multimodal imaging study (approved by the local institutional review board of the Medical University of Vienna) and was identical to the patient cohort in Kuess *et al* (2017).

The imaging was performed on a 3 T MAGNETOM TimTrio MRI scanner (Siemens Healthcare, Erlangen, Germany) and contained an anatomical high resolution T_2 -weighted turbo spin echo in three planes and a DWI (single-shot echo-planar) with inversion recovery fat suppression from which ADC maps were computed directly using the scanner software (four b -values were used, namely: 0, 100, 400, and 800 s mm^{-2}). Other images were also acquired, but they were not used in this study. The ADC images were subsequently resampled to the voxel size of the T_2 -weighted images (i.e. $0.625 \times 0.625 \times 3.6 \text{ mm}^3$), and the different regions of the prostate were manually delineated on the T_2 -weighted images by two radiation oncologists (Kuess *et al* 2017). The texture features were subsequently computed from the resampled ADC images.

3.3. Data processing

The images from the Kylberg data set were resized to 288×288 , but where otherwise not altered.

The central gland from the prostate images was extracted using the oncologists' delineations. Image slices were extracted as transversal planes. The ROIs containing the central glands of all patients were used to generate the GLCMs.

After these steps, sample points were generated from the underlying GLCM distribution of each texture class or of the prostate data by considering a pixel and one of its neighbours in an image as a sample point from that distribution. Neighbours were defined as spatially adjacent pixels in eight directions, defined by the displacement vectors

$$\delta \in \{(-1, 1), (0, 1), (1, 1), (0, -1), (0, 1), (-1, -1), (-1, 0), (-1, 1)\}.$$

3.3.1. The 'true' feature values

The true feature values, f_i^* , for $i = 1, \dots, 20$, are generally not accessible, since the data does not cover the entire population. This is also the case in this work. Therefore, we instead estimated the true invariant feature values using very close approximations to the true GLCMs. We will denote these features as the true features below, but acknowledge that these are still only approximations (not being computed from their whole populations), albeit very good ones, of their true population GLCMs.

In order to have ground truths to compare to, we used a large GLCM size (256 grey-levels), such that the invariant features are close to their asymptotic values (all features appeared to have converged for 256 grey-levels in Löfstedt *et al* (2017)), and we used many sampled points. This lead to well-defined GLCMs from which the approximations of the true feature values were computed.

The approximated *true* GLCMs were generated by randomly selecting 60% of the sample points from each texture class in the Kylberg texture data set and similarly from the prostate cancer data. For the Kylberg texture data set, the approximations of the *true* GLCMs were generated from 63.4 million samples each, and for the prostate cancer data, the approximation of the *true* GLCM was generated from 2.0 million samples.

These GLCMs are very well-defined and are assumed to accurately capture the characteristics of the different textures classes. Hence, the computed approximations of the true feature values are assumed to be very good approximations.

3.4. Comparison

The 40% of sample points not used when approximating the true feature values were used in the comparison of the different methods. When drawing samples in the comparison described below, subsets of samples were selected uniformly with replacement.

The invariant Haralick texture feature values from each method were compared to the true feature values using the relative RMSE, see equation (4), for sample sizes ranging from $N = 200$ to $N = 100\,000$ in 21 equidistant steps on a logarithmic scale. The number of samples, N , corresponds to a square ROI with side length approximately $\sqrt{N}/8$; hence, the sample sizes correspond roughly to ROI sizes ranging from 5×5 to 110×110 pixels. Note that a square ROI is only used as a reference for presentation purposes, and in fact, the ROI can have any shape (as in the prostate data example). The number of sampled invariant Haralick feature values, f_i^k , used in equation (4) were $K = 100$ for both the Kylberg data for the prostate data.

The three methods for GLCM estimation introduced above and the *Optimal method* are described in more detail in the following subsections.

3.4.1. Fixed number of grey-levels

The usual way of computing Haralick texture features is to decide on a fixed number of grey-levels and use this value irrespective of the ROI size. In the literature, the number of grey-levels varies substantially (Brynnolfsson *et al* 2017), and we therefore investigated a wide range of levels, namely $n \in \{8, 16, 32, 64, 128\}$. We denote this method the *Fixed method*.

3.4.2. Variable number of grey-levels

We recall that the GLCM, $\tilde{\mathbf{P}}$, in equation (3) is a discrete approximation to an underlying continuous function, with n discrete steps in each dimension. We here interpret this as an approximation of the *true* underlying probability density function by a piece-wise constant function, normalised such that it is a probability density. The *true* underlying probability density function is denoted p^* , and the piece-wise constant approximation is denoted by $p_{\text{var}}(\mathbf{x}|n, \boldsymbol{\theta})$, where $\mathbf{x} = (i, j) \in [0, 1]^2$, n is the number of grey-levels (i.e. the number of piece-wise patches), and $\boldsymbol{\theta}$ contains all other parameters that define $\tilde{\mathbf{P}}$.

We consider the sampled points as coordinates in the GLCM, and then pose the log-likelihood function over a set of sampled points as

$$\begin{aligned} \log \mathcal{L}(\mathcal{X}|n) &= \sum_{\mathbf{x} \in \mathcal{X}} \log(p^*(\mathbf{x})) \\ &\approx \sum_{\mathbf{x} \in \mathcal{X}} \log(p_{\text{var}}(\mathbf{x}|n, \boldsymbol{\theta})) \\ &\approx \sum_{\mathbf{x}=(ij) \in \mathcal{X}} \log\left(\tilde{P}_{\lceil ni \rceil, \lceil nj \rceil} + \frac{1}{N \cdot n^2}\right) \\ &\equiv \log \mathcal{L}_{\text{var}}(\mathcal{X}|n), \end{aligned} \quad (5)$$

where the second term in the second approximation is a Laplace smoothing term (corresponding roughly to a Dirichlet prior), that was added to avoid the logarithm of zero. The Laplace smoothing term here distributes a mass of $1/N$ among all elements of the GLCM.

Then, we performed seven-fold cross-validation by splitting the set of points in two for each fold: one training set containing $6/7$ of the points and one validation set containing the other $1/7$ remaining points. We generated a GLCM with n grey-levels from the training set and evaluated the log-likelihood over the validation set. The number of grey-levels, n , that corresponded to the largest log-likelihood was selected for each fold. The process was repeated over the seven folds of sampled points and the computed mean number of grey-levels (rounded to the nearest integer) was used.

Hence, the cross-validated maximum likelihood estimate of the number of grey-levels was

$$n^* = \left\lfloor \frac{1}{B} \sum_{b=1}^B \arg \max_{n \in \mathbb{N}^+} \log \mathcal{L}_{\text{var}}(\mathcal{X}_b|n) \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the round to nearest integer function, $B = 7$, and \mathcal{X}_b is the b th fold of the sampled points, \mathcal{X} . This value of n^* was used to generate the final GLCM by this method, that we denote the *Variable method*.

3.4.3. Parzen windows

The third method we investigate is the Parzen-window method (also called kernel density estimation or the Parzen–Rosenblatt window) to estimate the true underlying probability density function, p^* (Parzen 1962, Duda *et al* 2000).

By this method, the approximated probability density function is defined as

$$p_{\text{pw}}(\mathbf{x}|h, \kappa) = \frac{1}{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} \frac{1}{h^2} \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right),$$

where κ is a window function, or kernel, h is the window width, or bandwidth, and $|\mathcal{X}|$ denotes the number of elements (its cardinality) in the set of samples, $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_{|\mathcal{X}|}\}$.

We used a Gaussian kernel, i.e. one such that

$$\frac{1}{h^2} \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) = \frac{1}{2\pi\sigma^2} e^{-\frac{(\mathbf{x} - \mathbf{x}_i)^T(\mathbf{x} - \mathbf{x}_i)}{2\sigma^2}},$$

where thus $h = \sigma$, the standard deviation of the Gaussian kernel.

As with the previous method, we consider the sampled points as coordinates in the GLCM, and pose the log-likelihood function over the set of sampled points as

$$\log \mathcal{L}_{\text{pw}}(\mathcal{X}|\sigma) = \sum_{\mathbf{x} \in \mathcal{X}} \log(p_{\text{pw}}(\mathbf{x}|\sigma)).$$

Similar to the previous method, we used cross-validation to compute the maximum log-likelihood value of the window size using the left-out samples, and retained the value of the window size,

$$\sigma^* = \frac{1}{B} \sum_{b=1}^B \arg \max_{\sigma > 0} \log \mathcal{L}_{pw}(\mathcal{X}_b | \sigma),$$

where each fold maximised the log-likelihood value (Awate 2006). This cross-validated maximum likelihood value of σ^* was then used to generate the final GLCM by this method, that we denote the *PW method*.

3.4.4. Gaussian mixture models

The fourth method we investigated was the Gaussian mixture model, that also approximates the underlying probability density function, p^* , from a set of samples (Duda et al 2000, Hastie et al 2009). The model has the form

$$p_{gmm}(\mathbf{x} | M, \boldsymbol{\theta}) = \sum_{m=1}^M \alpha_m \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m),$$

where M is the number of Gaussian components and $\boldsymbol{\theta}$ contains all other model parameters. The α_m are mixing coefficients (such that $\sum_m \alpha_m = 1$), the \mathcal{N} is the bivariate normal probability density function, the $\boldsymbol{\mu}_m$ are the means of the components, and the $\boldsymbol{\Sigma}_m$ are the covariance matrices for the components, $m = 1, \dots, M$.

In this case, the log-likelihood function over a set of samples, \mathcal{X} , is defined as

$$\log \mathcal{L}_{gmm}(\mathcal{X} | M) = \sum_{\mathbf{x} \in \mathcal{X}} \log (p_{gmm}(\mathbf{x} | M, \boldsymbol{\theta})),$$

where the model parameters, $\boldsymbol{\theta}$, i.e. the mixing coefficients, means, and covariance matrices, were estimated from the data using the `gmdistribution` class in Matlab 2016b (The MathWorks, Inc., Natick, MA, USA). This class initialises the parameters using the K-means++ algorithm and uses the expectation-maximisation algorithm to fit the parameters. To avoid singular covariance matrices, we added a small regularisation term that corresponds to adding a small value (0.001) to the diagonal of the covariance matrices.

Like in the previous methods, we again used seven-fold cross-validation to estimate the unknown statistic, in this case the number of components. The number of components corresponding to the maximum log-likelihood value of the left-out samples was used from each fold, and the mean number of components (rounded to the nearest integer),

$$M^* = \left\lfloor \frac{1}{B} \sum_{b=1}^B \arg \max_{M \in \mathbb{N}^+} \log \mathcal{L}_{gmm}(\mathcal{X}_b | M) \right\rfloor,$$

was then used to generate the estimated GLCM. We denote this method the *GMM method*.

3.4.5. Optimal number of grey-levels

We generated GLCMs of sizes $n = 4, 5, \dots, 256$, for each sample size, $N = 200, \dots, 100\,000$, in the same 21 equidistant logarithmic steps as above. Then an *optimal* GLCM size was selected for each number of samples and for each feature as the GLCM size that minimised the relative RMSE, as illustrated in figure 1. This approach will be denoted the *Optimal method*.

3.5. Statistical test of results

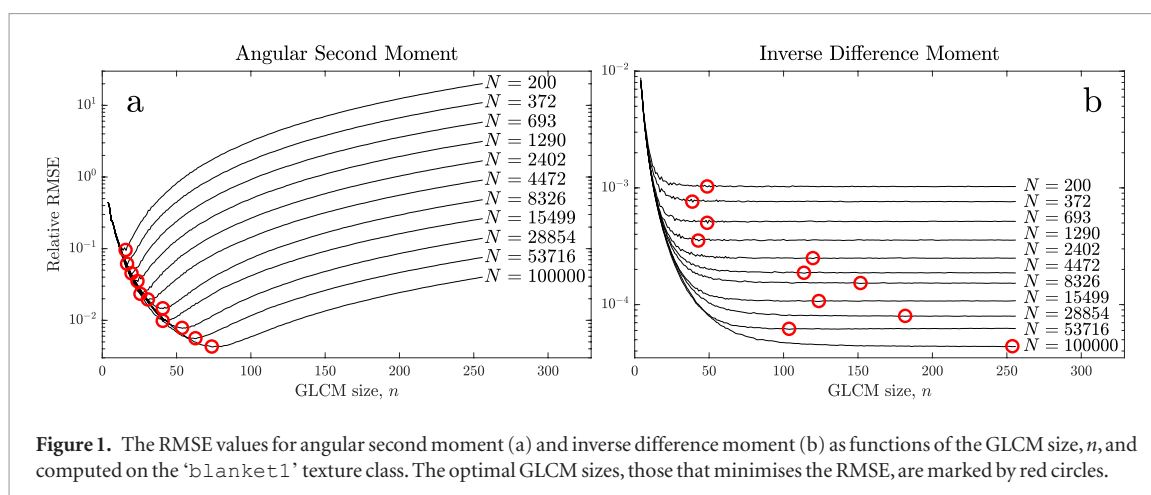
In order to determine whether any of the proposed methods outperform any other methods, we performed the Friedman test followed by a Nemenyi post-hoc test (Demšar 2006). The minimum relative RMSE over the fixed methods and the relative RMSE values of the three proposed methods were compared over all textures (from the Kylberg data set and the prostate data set) and all features for each ROI size.

The Friedman test was used in this case to test the hypothesis that all relative RMSE values are equal against the alternative that they are not. When the Friedman test determines that the observed data is significantly different from what can be expected by chance under the null hypothesis, the Nemenyi post-hoc test can be used to compare each method to each other method. We performed this test for each sample size (i.e. ROI size).

The Nemenyi post-hoc test was computed on the 5% level and was Bonferroni corrected for 21 tests (one Nemenyi test for each ROI size).

3.6. Computational time of the results

The wall-clock time of the computations were measured for each method when computed on Intel Xeon E5 2.60 GHz processors using a single core. We will present the results for each ROI size as medians over the required computational time for all textures and all features.



4. Results

Figure 1 illustrates the general behaviour of the relative RMSE as functions of the GLCM size, n , for two features. The other features also displayed the behaviour that is clearly seen in figure 1, namely that they either have a clear minimum, or that the relative RMSE decreases monotonically as a function of n . Similar plots for all features and all textures are presented in the supplementary material (stacks.iop.org/PMB/63/195017/mmedia). Table 1 lists the two apparent types of features in different columns.

Figure 2 illustrates the general behaviour of the estimated GLCMs for five Kylberg textures and the central gland from a prostate. The GLCMs from each method were computed from 10 000 samples, illustrated as a high-resolution GLCM in the third column. The GLCMs in the last three columns should be compared to the corresponding approximation of the true GLCMs in the second column.

With sufficient data the computed GLCMs will be smooth, as can be seen in the second column. However, when only small ROIs are available the smoothness cannot be reproduced. If very small bins are used (i.e. large n , where points are placed at or close to their coordinates), the resulting GLCMs will be noisy, as seen in the third column in figure 2. If larger bins are used (i.e. smaller n), as is often the case in the *Variable*, *Fixed*, and *Optimal methods*, the result is a blocky GLCM, as can be seen for the *Variable method* in the fourth column of figure 2. The inherent smoothing in the *PW* and *GMM methods* lead to GLCMs that reproduce the smoothness of the reference GLCM well, despite the small number of samples.

4.1. Analysis of the relative RMSE

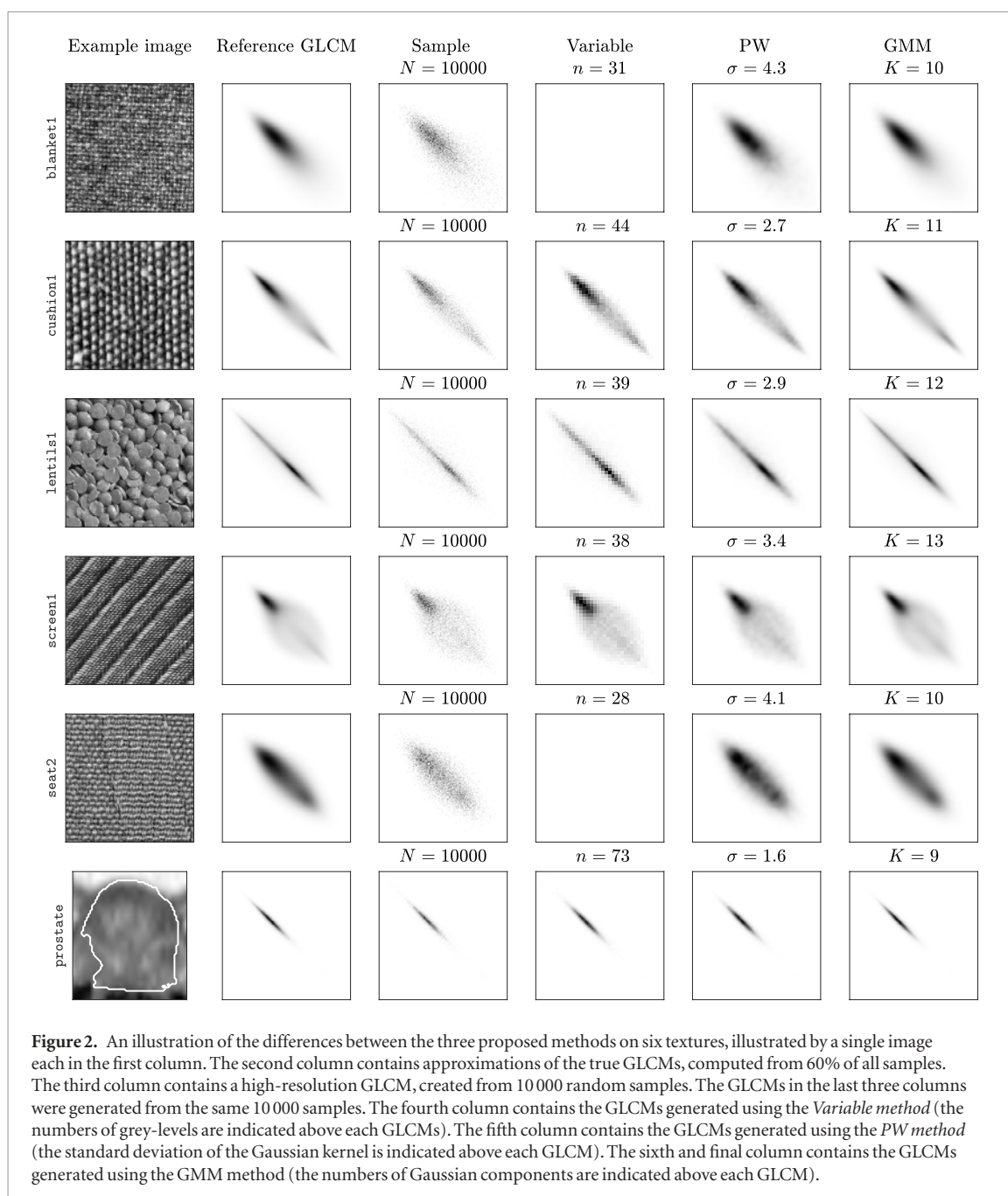
Figures 3 and 4 illustrate the relative RMSE of five representative invariant Haralick texture feature values for the 'blanket1' texture in the Kylberg data set and the prostate data, respectively. For plots of relative RMSE for all features and all textures, see the supplementary material. The behaviour of the different methods is not consistent in the sense that certain methods always outperform other methods. However, there are some general trends. As expected, the *Optimal method* give the smallest error in most cases. For features in the second column of table 1, larger GLCM sizes consistently give smaller errors for all ROI sizes (this is clearly seen for the *Fixed method* in the figures). Among the *Variable method*, the *PW method*, and the *GMM method*, the *GMM method* usually produces the smallest relative RMSE, but this is not always the case for all features, textures, and ROI sizes as can be seen in e.g. figure 3.

By ranking which method produces which relative RMSE for all methods (smallest error ranked first), an overview of the performance across all textures can be obtained, despite the fact that the relative RMSE values can have different scales for different textures. The results for five representative features are illustrated in figure 5.

4.1.1. Testing all methods on all data

The Friedman test was significant ($p \ll 10^{-16}$) for all ROI sizes. The ranks computed for the Friedman and Nemenyi tests are presented in figure 5. The figure contains the mean ranks over all textures (all Kylberg textures and the prostate data) for each ROI size, plotted with one standard deviation confidence bands computed point-wise.

The results from the Nemenyi post-hoc tests are presented in table 2. The first column contains the methods for which test results are presented on the same row; the tests were for significantly lower relative RMSE values. The top-most row contains the ROI sizes, with columns on distances proportional to the actual ROI sizes. A dash (—) indicates a non-significant Nemenyi test and a circle (○) indicates a significant test, i.e. a significantly lower



relative RMSE value (or, more accurately, significantly lower rank than what can be expected by chance) for the method on the left of the less than sign.

4.2. Timing results

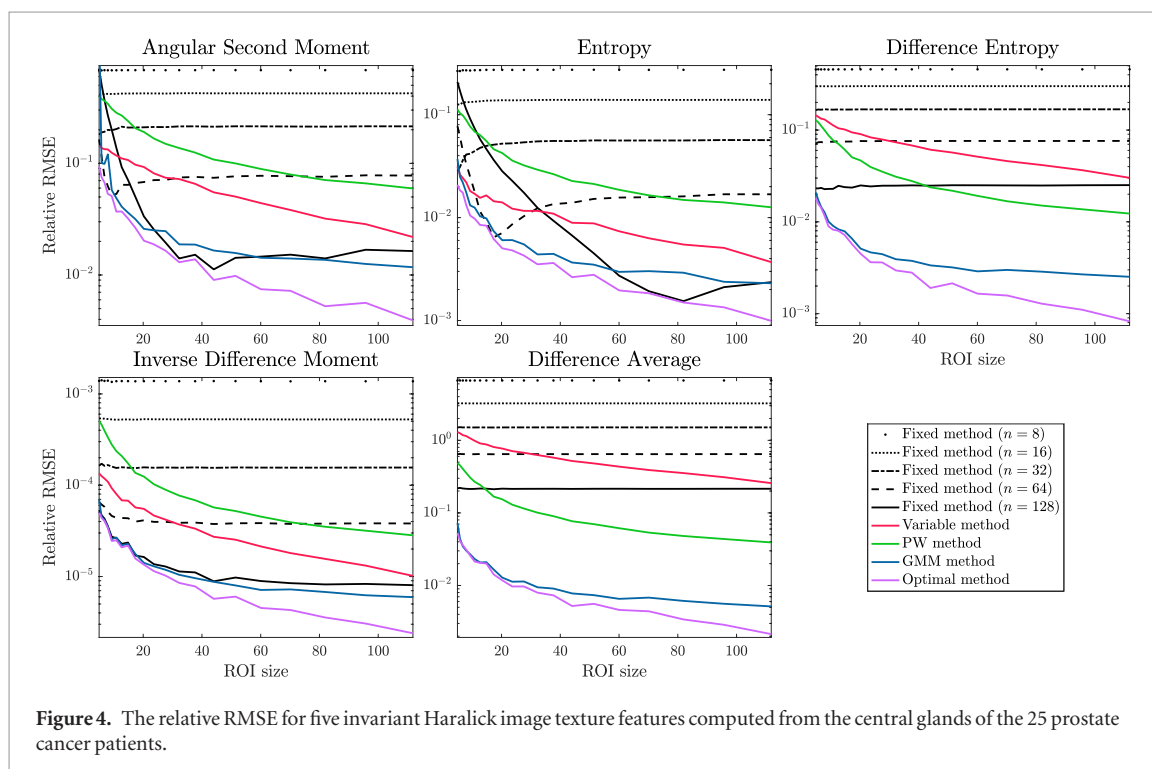
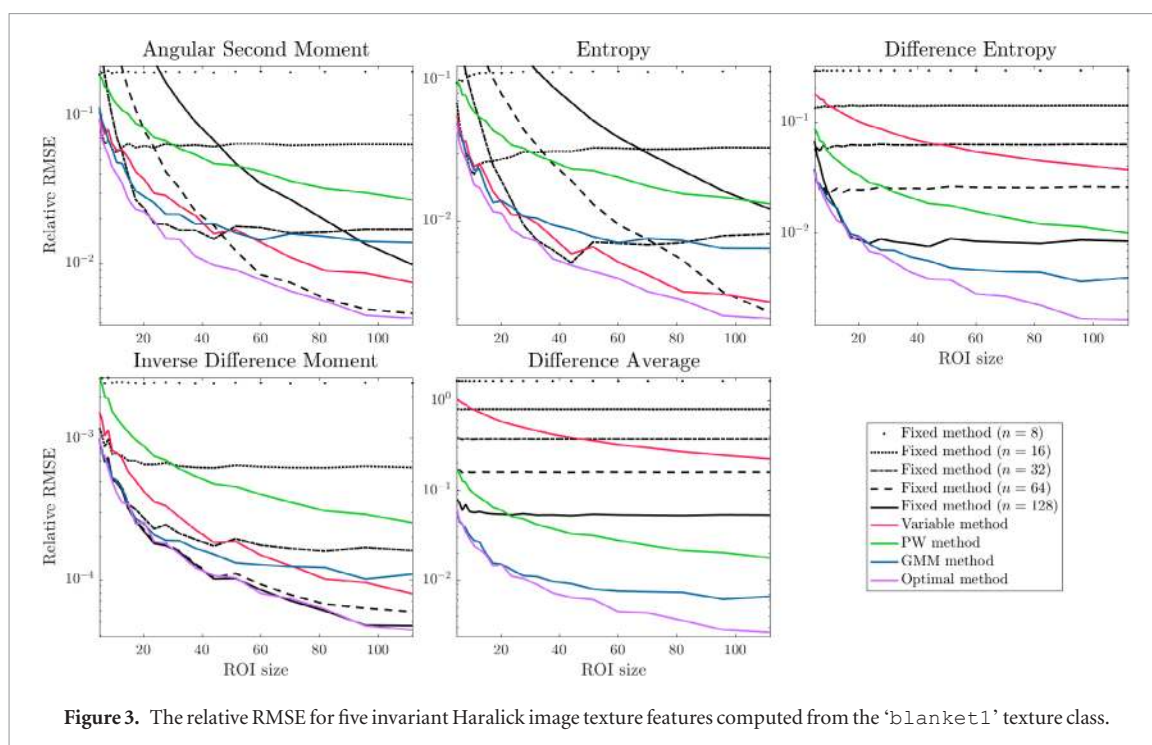
The median times for each method over all textures and all features are presented in figure 6. It is clear that the *GMM method* is substantially slower than the other methods.

5. Discussion

The purpose of this work was to evaluate methods to estimate GLCMs for different sample sizes and to see how these estimates affect the (asymptotically) invariant Haralick texture features, and in particular how the methods performed when the sample sizes were small.

The data suggests that the different methods perform quite different for different textures and features. However, the following general trends are supported by the statistical test we have performed:

- The *PW method* gives significantly smaller relative RMSE errors compared to the *Variable method* for ROI sizes ≤ 13 pixels.



- The *GMM method* gives significantly smaller errors compared to the minimum of the fixed methods for ROI sizes ≤ 20 pixels.
- The *GMM method* gives significantly smaller errors compared to the *Variable method* for all evaluated ROI sizes.
- The *GMM method* gives significantly smaller errors compared to the *PW method* for all evaluated ROI sizes.
- None of the proposed methods gives significantly smaller errors than the *Optimal method* for any ROI size.
- The *Optimal method* does not always give significantly smaller errors than the *GMM method* for ROI sizes smaller than 13, which suggests that the *GMM method* may perform only marginally worse than the *Optimal method* in this regime.

Further, we can say that the *Variable method* requires at least ten times longer computational time compared to the *Fixed methods*, the *PW method* requires about ten times longer computational time compared to

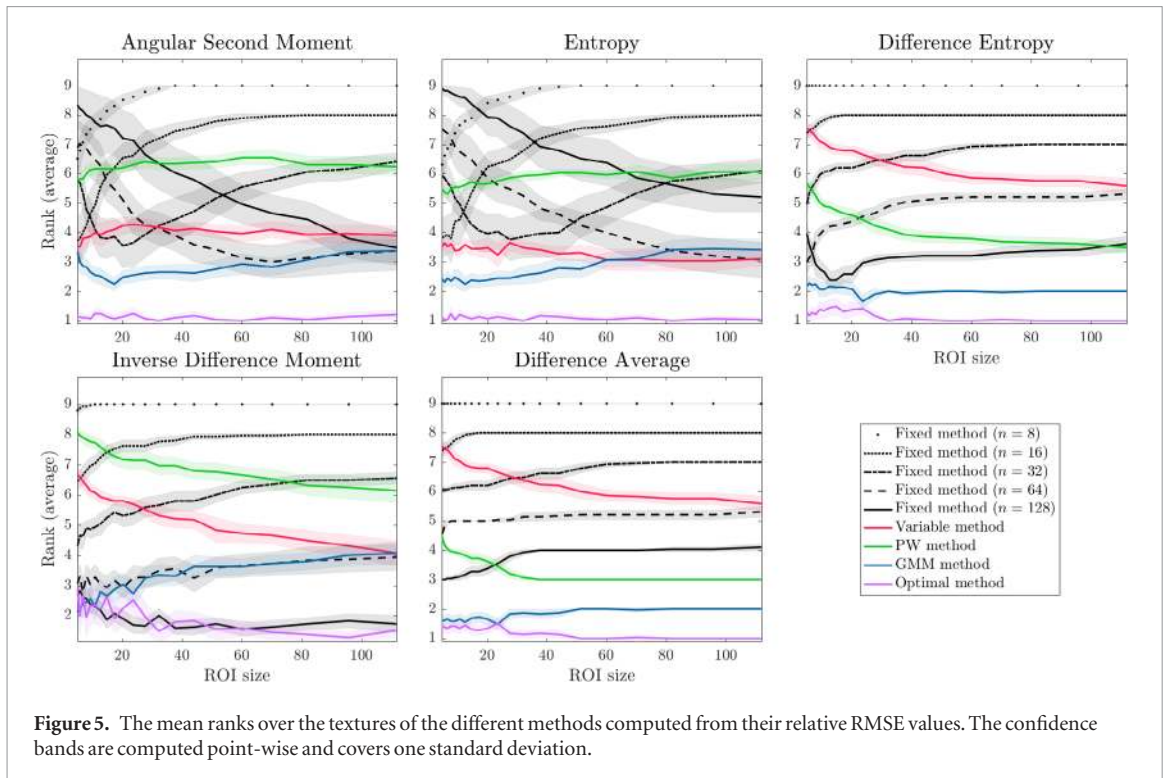


Figure 5. The mean ranks over the textures of the different methods computed from their relative RMSE values. The confidence bands are computed point-wise and covers one standard deviation.

Table 2. The results from the null-hypothesis significance test. The left-most column contains the names of the two methods for which results are given on that row. The *Fixed method* is denoted *Fix*, the *Variable method* is denoted *Var*, the *PW method* is denoted *PW*, the *GMM method* is denoted *GMM*, and the *Optimal method* is denoted *Opt*. The rows contains the results when testing whether the first method has a significantly lower relative RMSE than the second method. The Nemenyi test was performed for each ROI size (top-most row, proportional distances). A dash (–) means not significant. A circle (°) means that the first method had a significantly lower error than the second method.

Method \ ROI size	56	78	111	135	157	20	24	28	32	38	44	51	60	70	82	96	112
Var < Fix	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Var < PW	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Var < GMM	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Var < Opt	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PW < Fix	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PW < Var	°	°	°	°	°	–	–	–	–	–	–	–	–	–	–	–	–
PW < GMM	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PW < Opt	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
GMM < Fix	°	°	°	°	°	°	–	–	–	–	–	–	–	–	–	–	–
GMM < Var	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
GMM < PW	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
GMM < Opt	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Opt < Fix	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
Opt < Var	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
Opt < PW	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
Opt < GMM	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

the *Variable method*, and the *GMM method* requires about ten times longer computational time compared to the *PW method*. There is thus a trade-off between accuracy and computational time that may have to be taken into account when selecting a method to use in practice.

The results of this work indicate that the investigated features can be subdivided in two types. The first type have a clear optimal number of grey-levels for a given sample size. For the second type, the relative RMSE values decreased monotonically (when ignoring the noise) with the number of grey-levels for all sample sizes. This observation may be used for selecting GLCM sizes that reduce errors in the estimated feature values, which leads to more accurate results in e.g. radiomic applications.

When looking at tables 1 and 2 in Löfstedt et al (2017), we note that the features appearing to belong to the second type are all linear in $\tilde{\mathbf{P}}$, the probability density (or the corresponding sum and difference distributions; also, assuming the means and standard deviations are constant). This means that the second type of features are computing expectations over ϕ with respect to p^* , i.e. $\mathbb{E}_{p^*}[\phi | \mathbf{g}] \approx \mathbb{E}_{\tilde{\mathbf{P}}}[\phi | \mathbf{g}]$, and we can therefore expect them to converge to an asymptotic value as the accuracy in the estimation of \mathbf{P} increases. In fact, by the central limit theorem, and under mild assumptions, the convergence speed should be at least $\mathcal{O}(1/\sqrt{N})$. We have not analysed this rigorously, however, and instead leave this as an hypothesis for future work.

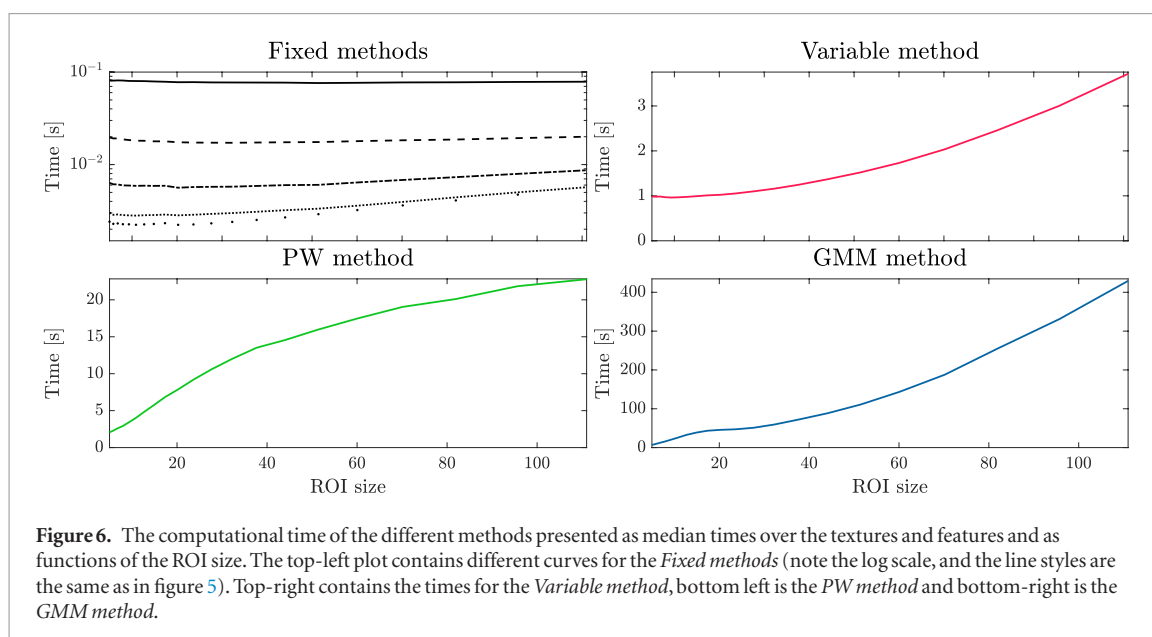


Figure 6. The computational time of the different methods presented as median times over the textures and features and as functions of the ROI size. The top-left plot contains different curves for the *Fixed methods* (note the log scale, and the line styles are the same as in figure 5). Top-right contains the times for the *Variable method*, bottom left is the *PW method* and bottom-right is the *GMM method*.

In practical applications, and in particular in medical imaging applications, small ROI sizes are common. An example is e.g. small tumours in MRI images, or when feature maps are computed as local texture from small ROIs centred at every pixel in the image. Because of this, the ability to compute features from very small ROIs is highly valuable, and could e.g. result in more robust diagnoses, better predictions in radiomic applications, more accurate predictions of clinical outcomes, and smaller errors in local texture features. The results in this work gives an indication regarding how to approach this. For the second type of features (those where the error decreases with increased number of grey-levels), the largest possible numbers of grey-levels yields both the highest accuracy and the lowest computational time (when compared to the three proposed methods). For the first type of features (those with an optimal number of grey-levels for each ROI size), the *GMM method* appears to be preferred for small ROI sizes, and the threshold appears to be at around 20×20 pixels. For ROI sizes larger than 20×20 pixels the answer is not as clear. The *GMM method* is significantly better, on average, than the other methods evaluated in our test, and would therefore likely suffice for most applications. The *GMM method* is, however, not always the best one for all features and all textures (see the supplementary plots). In fact, it might be in this case that further subdivision is required, such that different features are treated differently. The case for the first type of features with ROI sizes larger than 20×20 pixels could therefore be a subject for future studies.

We would like to stress, though, that (for the first type of features) there is no single fixed value for the GLCM size that performs better than the *GMM method* on average. I.e. if evaluated for different ROI sizes, there is only a single ROI size that a fixed value is optimal for (or, at best, a small range for which it performs well), while the *GMM method* will perform at worst moderately well for *all* ROI sizes.

The practical utility of the proposed methods in e.g. classification or regression models was outside of the scope of this paper. However, the results presented here implies that the accuracy in such studies when using e.g. the *GMM method* would be better on average compared to when using a fixed GLCM size—in particular if the features used were computed on images of different sizes. This is certainly a highly interesting subject of future work.

The methods investigated here are limited in scope, for instance with the use of Gaussian kernels in the *PW method* and a mixture of Gaussians in the *GMM method*. Using cross-validation for the estimation of hyper-parameters (e.g. kernel size and number of Gaussian components) was also a choice we made where other methods could have been of interest. Further investigation into other methods and variations in the model parameters we tested was out of the scope of this work but would be interesting for future studies.

A trend recently in medical imaging is to use volumetric images when computing the GLCM, i.e. to also include voxels across neighbouring transversal planes when computing the GLCM (Tesař et al 2008). The proposed methods would work in this setting as well, but care must be taken (like today) in case the voxels are anisotropic.

Another recent trend in longitudinal studies is to use ‘delta-features’ (see e.g. Aerts et al (2016) and Chang (2018)), where e.g. post-treatment features are normalised with respect to pre-treatment features. The proposed methodology may be particularly beneficial in this case since the ROI sizes are likely to change before and after treatment. E.g. the size of a tumour may be reduced after treatment leading to larger errors in the estimated feature values after treatment than before. Hence, using the proposed methodology, the errors in the pre-treat-

ment and post-treatment features would likely be smaller, leading to smaller errors in the delta-features as well. This would be an interesting topic for future work.

Another possible continuation of the present work includes improving the speed of the *GMM method*. Currently, the estimation of the number of components involve a grid search over a range of plausible values. It is likely that this can be made more efficient.

The *Optimal method* was used in this work as a kind of *oracle*, giving an ‘objective’ ground truth to which we compared the other methods. Another possible future work would be to investigate when this method could be used directly to estimate the best GLCM size. For instance, whenever several ROIs of the same texture are available it is possible that there is sufficient data to compute a *good enough* (in some sense) estimate of the optimal GLCM size using the *Optimal method*.

6. Conclusion

This study suggests that there might be two main types of (invariant) Haralick texture features and that they should be treated differently. This is, as far as the authors are aware, the first time this observation has been made about the Haralick texture features. The first type of features (those in the left column of table 1) have a clear optimal number of grey-levels for a given sample size. For the second type of features, the relative RMSE values appears to decrease monotonically with the number of grey-levels for all sample sizes. For the first type of features, the *GMM method* is preferred among the investigated methods—in particular for ROI sizes smaller than about 20×20 pixels. For the second type of features, selecting a large (as large as possible) GLCM size appears to be the best choice.

Acknowledgments

We are grateful for the financial support obtained from the Cancer Research Foundation in Northern Sweden, Karin and Krister Olsson, Umeå University, The Västerbotten regional county, and Vinnova, the Swedish innovation agency.

The financial support by the Austrian Federal Ministry of Science, Research and Economy, and the National Foundation for Research, Technology and Development is also gratefully acknowledged.

This research was conducted using the resources of the High Performance Computing Center North (HPC2N).

Anders Garpebring, Patrik Brynolfsson, Tufve Nyholm, and Tommy Löfstedt are co-owners of Nonpi Medical AB, the developer of MICE Toolkit—a software for medical image analysis.

We would also like to thank Dr Gustaf Kylberg for kindly answering our questions regarding the Kylberg Texture Dataset v.1.0.

ORCID iDs

Anders Garpebring  <https://orcid.org/0000-0002-0532-232X>

Patrik Brynolfsson  <https://orcid.org/0000-0002-0455-8904>

Peter Kuess  <https://orcid.org/0000-0003-2961-1692>

Dietmar Georg  <https://orcid.org/0000-0002-8327-3877>

Thomas H Helbich  <https://orcid.org/0000-0003-3169-778X>

Tufve Nyholm  <https://orcid.org/0000-0002-8971-9788>

Tommy Löfstedt  <https://orcid.org/0000-0001-7119-7646>

References

- Aerts H J W L, Grossmann P, Tan Y, Oxnard G R, Rizvi N, Schwartz L H and Zhao B 2016 Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC *Sci. Rep.* **6** 33860
- Agarwal A, Singh R and Vatsa M 2016 Fingerprint sensor classification via mélange of handcrafted features *23rd Int. Conf. on Pattern Recognition* (International Association of Pattern Recognition, IEEE Computer Society) pp 3001–6
- Ahmed A, Gibbs P, Pickles M and Turnbull L 2013 Texture analysis in assessment and prediction of chemotherapy response in breast cancer *J. Magn. Reson. Imaging* **38** 89–101
- Assefa D, Keller H, Menard C, Laperriere N, Ferrari R J and Yeung I 2010 Robust texture features for response monitoring of glioblastoma multiforme on T1-weighted and T2-FLAIR MR images: a preliminary investigation in terms of identification and segmentation *Med. Phys.* **37** 1722–36
- Awate S P 2006 Adaptive, nonparametric markov models and information-theoretic methods for image restoration and segmentation *PhD Thesis* The University of Utah
- Basu S, Karki M, DiBiano R, Mukhopadhyay S, Ganguly S, Nemani R and Gayaka S 2016 A theoretical analysis of deep neural networks for texture classification *Int. Joint Conf. on Neural Networks* pp 992–9

- Brynnolfsson P, Nilsson D, Torheim T, Asklund T, Thellenberg-Karlsson C, Trygg J, Nyholm T and Garpebring A 2017 Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters *Sci. Rep.* **7** 4041
- Chan H P, Wei D, Helvie M A, Sahiner B, Adler D D, Goodsitt M M and Petrick N 1995 Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space *Phys. Med. Biol.* **40** 857–76
- Chang Y 2018 An investigation of machine learning methods for delta-radiomic feature analysis *Master's Thesis* Duke Kunshan University and Duke University
- Chen W, Giger M L, Li H, Bick U and Newstead G M 2007 Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images *Magn. Reson. Med.* **58** 562–71
- Demšar J 2006 Statistical comparisons of classifiers over multiple data sets *J. Mach. Learn. Res.* **7** 1–30
- Duda R O, Hart P E and Stork D G 2000 *Pattern Classification* 2nd edn (New York: Wiley)
- Eliat P A, Oulivé D, Saikali S, Carsin B, Saint-Jalmes H and de Certaines J D 2012 Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? *Neurol. Res. Int.* **2012** 1951–76
- Garma F B and Hassan M A 2014 Classification of breast tissue as normal or abnormal based on texture analysis of digital mammogram *J. Med. Imaging Health Inform.* **4** 647–53
- Gensheimer M F, Hawkins D S, Ermoian R P and Trister A D 2015 Assessing the scale of tumor heterogeneity by complete hierarchical segmentation of MRI *Phys. Med. Biol.* **60** 977–93
- Georgiadis P, Cavouras D, Kalatzis I, Glotsos D, Athanasiadis E, Kostopoulos S, Sifaki K, Malamas M, Nikiforidis G and Solomou E 2009 Enhancing the discrimination accuracy between metastases, gliomas and meningiomas on brain MRI by volumetric textural features and ensemble pattern recognition methods *Magn. Reson. Imaging* **27** 120–30
- Haralick R M, Shanmugam K and Dinstein I 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **3** 610–21
- Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference and Prediction* 2nd edn (New York: Springer)
- Kuess P et al 2017 Association between pathology and texture features of multi parametric MRI of the prostate *Phys. Med. Biol.* **62** 7833–54
- Kukunda C B, Duque-Lazo J, González-Ferreiro E, Thaden H and Kleinn C 2018 Ensemble classification of individual Pinus crowns from multispectral satellite imagery and airborne LiDAR *Int. J. Appl. Earth Obs. Geoinf.* **65** 12–23
- Kylberg 2011 *The Kylberg Texture Dataset v. 1.0 (External Report (Blue Series) vol 35)* (Uppsala: Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University) (www.cb.uu.se/~gustaf/texture/)
- Le Cun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
- Lerski R, Barnett E, Morley P, Mills P, Watkinson G and MacSween R 1979 Computer analysis of ultrasonic signals in diffuse liver disease *Ultrasound Med. Biol.* **5** 341–3
- Livens S, Scheunders P, van de Wouwer G and Van Dyck D 1997 Wavelets for texture analysis, an overview *6th Int. Conf. on Image Processing and its Applications IET Conf. Proc.* pp 581–5
- Löfstedt T, Brynnolfsson P, Asklund T, Nyholm T and Garpebring A 2017 Gray-level invariant Haralick texture features. Applications of statistical methods in quantitative magnetic resonance imaging *PhD Thesis* Patrik Brynnolfsson and Umeå University
- Lopes R, Ayache A, Makni N, Puech P, Villers A, Mordon S and Betrouni N 2011 Prostate cancer characterization on MR images using fractal features *Med. Phys.* **38** 83–95
- Marcos J V et al 2015 Automated pollen identification using microscopic imaging and texture analysis *Micron* **68** 36–46
- Mayerhoefer M E, Schima W, Trattnig S, Pinker K, Berger-Kulemann V and Ba-Ssalamah A 2010 Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: a feasibility study in cysts and hemangiomas *J. Magn. Reson. Imaging* **32** 352–9
- Mustra M, Grgic M and Delac K 2012 Breast density classification using multiple feature selection *Automatika* **53** 362–72
- Nagarajan M B, Huber M B, Schlossbauer T, Leinsinger G, Krol A and Wismueller A 2013a Classification of small lesions in breast MRI: evaluating the role of dynamically extracted texture features through feature selection *J. Med. Biol. Eng.* **33** 59–68
- Nagarajan M B, Huber M B, Schlossbauer T, Leinsinger G, Krol A and Wismueller A 2013b Classification of small lesions in dynamic breast MRI: eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement *Mach. Vis. Appl.* **24** 1371–81
- Nie K, Chen J H, Yu H J, Chu Y, Nalcioğlu O and Su M Y 2008 Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI *Acad. Radiol.* **15** 1513–25
- Ojala T, Pietikainen M and Harwood D 1994 Performance evaluation of texture measures with classification based on Kullback discrimination of distributions *Proc. of the 12th IAPR Int. Conf. on Pattern Recognition (Conference A: Computer Vision & Image Processing)* vol 1 (Los Alamitos, CA: IEEE Computer Society Press) pp 582–5
- Parzen E 1962 On estimation of a probability density function and mode *Ann. Math. Stat.* **33** 1065–76
- Prasanna P, Tiwari P and Madabhushi A 2016 Co-occurrence of local anisotropic gradient orientations (CoLLAGe): a new radiomics descriptor *Sci. Rep.* **6** 37241
- Roberts K, Weikum G and Mucklich F 2005 Examinations on the automatic classification of lamellar graphite using the support vector machine *Pract. Metallography* **42** 396–410
- Skorton D J, Collins S M, Woskoff S D, Bean J A and Melton H E 1983 Range- and azimuth-dependent variability of image texture in two-dimensional echocardiograms *Circulation* **68** 834–40
- Soltanian-Zadeh H, Rafiee-Rad F and Pourabdollah-Nejad S D 2004 Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms *Pattern Recognit.* **37** 1973–86
- Tesař L, Shimizu A, Smutek D, Kobatake H and Nawano S 2008 Medical image analysis of 3D CT images based on extension of Haralick texture features *Comput. Med. Imaging Graph.* **32** 513–20
- Thomazini D, Gelfuso M V and Altafim R A C 2012 Classification of polymers insulators hydrophobicity based on digital image processing *Mater. Res.* **15** 365–71
- Tokola T and Kilpeläinen P 1999 The forest stand margin area in the interpretation of growing stock using landsat TM imagery *Can. J. Forest Res.* **29** 303–9
- Ulaby F T, Kouyate F, Brisco B and Williams T H L 1986 Textural information in SAR images *IEEE Trans. Geosci. Remote Sens.* **24** 235–45
- Wu J, Gong G, Cui Y and Li R 2016 Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy *J. Magn. Reson. Imaging* **44** 1107–15
- Zacharaki E I, Wang S, Chawla S, Yoo D S, Wolf R, Melhem E R and Davatzikos C 2009 Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme *Magn. Reson. Med.* **62** 1609–18