AUTOMATIC ESTIMATION OF THE NOISE VARIANCE FROM THE HISTOGRAM OF A MAGNETIC RESONANCE IMAGE

1,2 D. Poot, 2 J. Sijbers, 3 A. J. den Dekker, 2 W. Pintjens

1 E-mail: Dirk.Poot@ua.ac.be, Telephone: +32 (0)3/820.24.44
2 Vision Lab, University of Antwerp, Belgium
3 Delft Center for Systems and Control, Delft University of Technology, The Netherlands

ABSTRACT

Estimation of the noise variance of a magnetic resonance (MR) image is important for various post-processing tasks. In the literature, various methods for noise variance estimation from MR images are available. In this paper, we focus on automatic histogram-based noise variance estimation techniques. Previously described methods are reviewed and a new method based on the maximum likelihood (ML) principle is presented. The noise variance estimation methods are compared in terms of the root mean-squared error (RMSE). The results show that the newly proposed method is superior in terms of the RMSE.

1. INTRODUCTION

The noise variance in magnetic resonance (MR) images has always been an important parameter to account for when processing and analyzing magnetic resonance imaging (MRI) data. Algorithms for noise reduction, segmentation, clustering, restoration, and registration highly depend on the noise variance. Also, many applications that employ statistical analysis techniques, such as functional MRI or voxel based morphometry, often base their conclusions on assumptions about the underlying noise characteristics. In the past, many techniques have been proposed to estimate the image noise variance. These can be subdivided into two classes:

multiple images The noise variance can be estimated by subtracting two acquisitions of the same object and calculating the standard deviation of the resulting pixel values [1]. These methods are relatively insensitive to structured noise such as ghosting, ringing, and DC artifacts, but perfect geometrical alignment and temporal stationarity of the imaging process is required.

single image The image noise variance can be estimated from a single magnitude image. A common approach is to estimate the noise variance from a large, manually selected, uniform signal region or non-signal (i.e., noise only) region [2, 3, 4, 5, 6]. Often, magnitude MR images contain a large number of background data points. Hence, the histogram of such images often shows a background mode that is clearly distinguishable from the signal contributions in the histogram. Therefore, automatic and robust noise variance estimation methods have been reported that exploit this background mode along with the knowledge that the noise-only contribution represents a Rayleigh distribution [9, 7, 8]. In this paper, these procedures are reviewed and a new method is presented.

2. METHODS

2.1. Noise properties of MR data

It is well known that the data in a magnitude MRI image are Rician distributed [2]:

\[
p(m|A, \sigma) = \frac{m}{\sigma^2} e^{-\frac{m^2 + A^2}{2\sigma^2}} I_0 \left( \frac{A m}{\sigma^2} \right) \epsilon(m),
\]

with \(I_0\) denoting the 0th order modified Bessel function of the first kind, \(A\) the noiseless signal level, \(\sigma^2\) the noise variance, and \(m\) the MR magnitude variable. The unit step Heaviside function \(\epsilon(\cdot)\) is used to indicate that the expression for the PDF of \(m\) is valid for non-negative values of \(m\) only. When the signal to noise ratio, defined as \(A/\sigma\), is zero, the Rice PDF simplifies into a Rayleigh PDF:

\[
p(m|\sigma) = \frac{m}{\sigma^2} e^{-\frac{m^2}{2\sigma^2}} \epsilon(m).
\]

2.2. Previously reported, histogram-based noise variance estimation methods

Magnitude MR images generally contain a large number of background data points. Hence, the histogram of such images often shows a background mode that is clearly distinguishable from the signal contributions in the histogram. Therefore, automatic and robust noise variance estimation methods have been reported that exploit this background mode along with the knowledge that the noise-only contribution represents a Rayleigh distribution [9, 7, 8]. In this section, these methods are reviewed. Next, in subsection 2.3, a new method is described based on ML estimation.

2.2.1. Maximum of the background mode of the histogram

From the Rayleigh PDF, given in (2), the noise variance can be estimated by searching for the value of \(m\) for which the Rayleigh PDF attains a maximum [9]:

\[
\frac{\partial p}{\partial m} = 0 \implies \hat{\sigma} = m_{\text{max}}.
\]
2.2.2. Brummer

In the work of Brummer et al., a noise variance estimation method is presented in which the Rayleigh distribution is fitted to a partial histogram using least squares estimation [7]:

\[
\hat{K}, \hat{\sigma}_{Br} = \arg\max_{K,\sigma} \sum_{i=0}^{f_c} \left( h(f) - K \frac{f_c}{\sigma^2} e^{-(f^2/2\sigma^2)} \right)^2
\]

(4)

where \( K \) is the amplitude and \( \sigma \) the width of the Rayleigh distribution that is fitted to the histogram \( h \). The cutoff \( f_c \) is defined as \( f_c = 2\sigma_{Br,0} \), where \( \sigma_{Br,0} \) is an initial estimate of the noise level. Brummer’s method specifies that the position of the first local maximum of the low-pass-filtered gray-value histogram is to be used as the initial estimate. In our implementation of Brummer’s method, we used Chang’s estimate (2.2.3) as an initial value.

2.2.3. Chang’s noise variance estimation method

In order to improve robustness of the noise variance estimation method described in 2.2.1 and 2.2.2, Chang et al. proposed a procedure to smooth the histogram prior to estimation [8]. Thereby, a Gaussian smoothing kernel \( K(y) = 1/\sqrt{2\pi \sigma \exp(-y^2/2)} \), was used. The smoothing width \( h \) was set to \( h = 1/0.68\sigma_0^{1/5} \) in which \( \sigma_0 \) is the sample standard deviation and \( n \) the sample size. The smoothed histogram of the image intensity data \( x_i \) is searched for the location of the first local maximum:

\[
\hat{\sigma}_{Ch} = \arg\max_{\sigma} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\sigma - x_i}{h} \right)
\]

(5)

2.3. New noise variance estimation method

In this subsection, a new noise variance estimation method will be described based on ML estimation. Let \( \{l_i\} \) with \( i = 0, \ldots, K \) denote the set of boundaries of histogram bins. Furthermore, let \( n_i \) represent the number of observations (counts) within the bin \([l_{i-1}, l_i]\), which are multinomial distributed. Then, the joint PDF of the histogram data is given by:

\[
p(\{n_i\} | \sigma, \{l_i\}) = \frac{N_K!}{\prod_{i=1}^{K} n_i!} \prod_{i=1}^{K} p_i^{n_i}(\sigma)
\]

(6)

with \( N_K = \sum_{i=1}^{K} n_i \) the total number of observations within the partial histogram and \( p_i \) the probability that an observation assumes a value in the range \([l_{i-1}, l_i]\). The likelihood function is obtained by regarding the \( n_i \)’s as constants and \( \sigma \) as variable. For Rayleigh distributed observations, \( p_i \) is given by

\[
p_i(\sigma) = \frac{f_{l_i} l_i m \exp\left(-\frac{m^2}{2\sigma^2}\right) \, dm}{\sum_{i=1}^{K} f_{l_{i-1}} l_i m \exp\left(-\frac{m^2}{2\sigma^2}\right) \, dm}
\]

(7)

When all bins are adjacent it is easy to show that (7) simplifies to

\[
p_i(\sigma) = \left( e^{-\frac{m^2}{2\sigma^2}} - e^{-\frac{m^2}{\sigma^2}} \right) \left( e^{-\frac{m^2}{2\sigma^2}} - e^{-\frac{m^2}{\sigma^2}} \right)^{-1}
\]

(8)

The ML estimate of \( \sigma \) is then found by minimizing \(-\ln L\) with respect to \( \sigma \):

\[
\hat{\sigma}_{ML,K} = \arg\min_{\sigma} \left[ N_K \ln \left( e^{-\frac{m^2}{2\sigma^2}} - e^{-\frac{m^2}{\sigma^2}} \right) - \sum_{i=1}^{K} n_i \ln \left( e^{-\frac{m^2}{2\sigma^2}} - e^{-\frac{m^2}{\sigma^2}} \right) \right]
\]

(9)

2.3.1. Selection of the number of bins

Notice that finding the ML estimate (9) comes down to fitting a (discretized) Rayleigh PDF to (the left) part of the image histogram. This raises the question how to select the number of bins \( K \) that constitute this part. Generally, a more precise estimate (lower variance) will be obtained if the number of bins \( K \) that constitute this part. However, incorporating bins with counts that can not be attributed solely to noise but also to signal contributions will introduce a bias (reduce accuracy). Hence, as a selection criterion for \( K \), a combined measure of the bias and variance of the estimator \( \hat{\sigma}_{ML,K} \) was chosen. This criterion is derived as follows.

**Variance**

A measure of the variance of \( \hat{\sigma}_{ML,K} \) was constructed from the Cramér-Rao lower bound (CRLB), which is a lower bound on the variance of any unbiased estimator. It is known that the ML estimator is consistent and asymptotically most precise (i.e., it attains the CRLB asymptotically), so a useful measure of the variance is given by

\[
\sqrt{\text{Var}(\hat{\sigma}_{ML,K})} = -\left( \frac{\partial^2}{\partial\sigma^2} \ln L(\sigma|\{n_i\}) \right)_{\sigma=\hat{\sigma}_{ML,K}}^{-1}
\]

(10)

where the term on the right hand side is known as the inverse of the observed Fisher information. This estimate of the variance was observed to be reliable only when a sufficient number of bins was taken into account. In our implementation, this number was chosen such that at least the maximum of the histogram was included.

**Bias**

A measure of the bias was found by quantifying the difference between the Rayleigh distribution fitted using the first \( K \) bins of the histogram and the actual bin counts in the histogram. The histogram bin counts \( n_i \) are multinomial distributed. Furthermore, the marginal distribution of the number of counts in each bin is a binomial distribution with parameters \( N_K \) and \( p_i \). This means that the expected value of \( n_i \) is \( p_i N_K \) and its variance is \( p_i (1 - p_i) N_K \). However, since in general \( N_K \) is large (and \( p_i \) is small), the binomial distribution can be approximated by a normal distribution
with expectation value and variance both equal to $p_iN_K$. Under the null hypothesis ($H_0$) that the observations in all bins are Rayleigh distributed, $p_i$ is given by (7). Next, consider the test statistic:

$$
\lambda_K = \frac{1}{2} \sum_{i=1}^{N} \frac{(f_{i,K} - n_i)^2}{f_{i,K}}, \quad (11)
$$

with $N$ the number of bins in the histogram and $f_{i,K} = p_i(\hat{\sigma}_{ML,K})N_K$. It can be shown that, under $H_0$, $\lambda_K$ is approximately $\chi^2_{K-2}$ distributed (i.e., chi-squared distributed with $N - 2$ degrees of freedom). Notice, that a large value of $\lambda_K$ indicates a bad fit and thus hints at the presence of a bias in our estimate of $\sigma$. Therefore, $\lambda_K$ will be used as a bias measure.

It is reasonable to assume that for the bins for which $i > K$ the counts due to the underlying, noiseless signal may outnumber those due to the background noise only. Since contributions from the underlying signal can only increase the bin counts $n_i$, the actual bin counts will likely be significantly higher than the counts predicted by the fitted Rayleigh distribution. If we exclude the bins $i$ with $i > K$ for which $n_i > f_{i,K}$ from (11), we obtain the modified test statistic:

$$
\hat{\lambda}_K = \frac{1}{2} \sum_{i=1}^{K} \frac{(f_{i,K} - n_i)^2}{f_{i,K}} + \sum_{i=K+1}^{N} \frac{[\max(0, f_{i,K} - n_i)]^2}{f_{i,K}}, \quad (13)
$$

The first term of (13) is known as Pearson’s test statistic, which is asymptotically $\chi^2$ distributed with $K - 2$ degrees of freedom under $H_0$. The second term of (13) is $\chi^2_M$ distributed under $H_0$, with

$$
M = \sum_{i=K+1}^{N} \epsilon(f_{i,K} - n_i), \quad (14)
$$

Since both terms are independent, $\hat{\lambda}_K$ is approximately $\chi^2_{K-2+M}$ distributed under $H_0$. Hence, the statistic

$$
\hat{b} = \frac{\lambda_K - (K - 2 + M)}{\sqrt{K - 2 + M}}, \quad (15)
$$

has approximately a standard normal distribution under $H_0$. The statistic (15) will be used as a measure of the bias.

**Selection criterion**

Finally, both measures of bias and variance given in (10) and (15), respectively, are combined into a single criterion that selects the optimal number of bins $K$:

$$
\hat{K} = \arg \min_K \left[ \hat{b} + \text{Var}(\hat{\sigma}_{ML,K}) \right]. \quad (16)
$$

3. EXPERIMENTS AND DISCUSSION

Experiments were designed to compare the performance of the noise variance estimators discussed in subsection 2.2 to that of the newly proposed method presented in subsection 2.3. As a performance measure, the root-mean-squared-error (RMSE) was used.

**Simulated noise-only images**

First, the performance of the estimators was compared using simulated, integer valued Rayleigh distributed data (corresponding to noise-only magnitude MR images), with different noise levels $\sigma$. The size of the image was $181 \times 80$. In Fig. 1, the RMSE of the different estimators is shown as a function of $\sigma$. At low noise levels, the proposed ML based noise variance estimation test clearly performs best in terms of the RMSE, partially due to the fact that the ML based estimator correctly accounts for the discreteness of the data. Due to the adaptive number of bins used, our estimator also has the lowest RMSE when the noise level is larger. In that case the RMSE of the ML based estimator is approximately half of the RMSE of the second best, which is Brummer’s estimator.

**Simulated 3D MR image**

Next, Rician distributed data with varying $\sigma$ were generated from a noiseless 3D MR image obtained from the simulator [10]. The size of the image was $181 \times 217 \times 60$. The results are shown in Fig. 4. As is clear from the figure, the proposed ML based estimator generally performs best in terms of the RMSE, although all estimators are biased when the noise level approaches the signal level.

**Simulated 3D MR image with ghost**

Finally, the robustness of the noise variance estimators in the presence of a ghost artefact was tested. The ghost was generated by circularly shifting the original 3D image in one direction over half the image size in that direction and reducing the intensities to 5% of the original. This ghost was added to the original image before adding Rician distributed noise. In Fig. 4, the results are presented. The change in the histogram of the noise free image which resulted from adding the ghost is mainly concentrated in the range 10 - 150. The ghost seems to slightly affect the noise variance estimation for all noise variance estimation methods. However, also in this case, the proposed ML based estimator generally performs best in terms of the RMSE.

4. CONCLUSIONS

In this paper, previously proposed noise variance estimation methods that employ the image histogram were reviewed and a new method was proposed based on Maximum Likelihood (ML) estimation. Simulation experiments showed that the ML based estimator outperforms the previously described estimators in terms of the root mean squared error.

**Acknowledgements**

The authors thank Robert Bos from Delft University of Technology (The Netherlands) for useful discussions. The research of D. Poot is funded by the I.W.T. (Institute for Science and Technology - Belgium). J. Sijbers is a post-doctoral fellow of the F.W.O. (Fund for Scientific research
5. REFERENCES


