Physica Medica xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Physica Medica

journal homepage: http://www.physicamedica.com

Original paper

Can the non-pre-whitening model observer, including aspects of the human visual system, predict human observer performance in mammography?

R.W. Bouwman [a,*], R.E. van Engen [a], M.J.M. Broeders [a,b], G.J. den Heeten [a,c], D.R. Dance [d,e], K.C. Young [d,e], W.J.H. Veldkamp [a,f]

[a] *Dutch Reference Centre for Screening (LRCB), Radboud University Medical Centre, The Netherlands*
[b] *Radboud Institute for Health Sciences (RIHS), Radboud University Medical Centre, The Netherlands*
[c] *Department of Radiology, Academic Medical Centre (AMC), The Netherlands*
[d] *National Co-ordinating Centre for the Physics of Mammography (NCCPM), Royal Surrey County Hospital, United Kingdom*
[e] *Department of Physics, University of Surrey, United Kingdom*
[f] *Department of Radiology, Leiden University Medical Centre (LUMC), The Netherlands*

ARTICLE INFO

Article history:
Received 13 May 2016
Received in Revised form 8 November 2016
Accepted 15 November 2016
Available online xxxx

Keywords:
Model observer
Mammography
Contrast sensitivity function
Clinical image quality

ABSTRACT

Purpose: In mammography, images are processed prior to display. Current methodologies based on physical image quality measurements are however not designed for the evaluation of processed images. Model observers (MO) might be suitable for this evaluation. The aim of this study was to investigate whether the non-pre-whitening (NPW) MO can be used to predict human observer performance in mammography-like images by including different aspects of the human visual system (HVS).

Methods: The correlation between human and NPW MO performance has been investigated for the detection of disk shaped objects in simulated white noise (WN) and clustered lumpy backgrounds (CLB), representing quantum noise limited and mammography-like images respectively. The images were scored by the MO and five human observers in a 2-alternative forced choice experiment.

Results: For WN images it was found that the log likelihood ratio (R^2_{LR}), which expresses the goodness of fit, was highest (0.44) for the NPW MO without addition of HVS aspects. For CLB the R^2_{LR} improved from 0.46 to 0.65 with addition of HVS aspects. The correlation was affected by object size and background.

Conclusions: This study shows that by including aspects of the HVS, the performance of the NPW MO can be improved to better predict human observer performance. This demonstrates that the NPW MO has potential for image quality assessment. However, due to the dependencies found in the correlation, the NPW MO can only be used for image quality assessment for a limited range of object sizes and background variability.

1. Introduction

In digital mammography (DM) image processing is applied to make images suitable for assessment on diagnostic monitors, and to improve diagnostic accuracy. This processing may involve non-linear steps which take into account the characteristics of breast images [1,2] and differs between manufacturers and software versions. Several studies have been conducted showing that image processing affects cancer detection [1–3] and emphasize the need for objective measurements of image quality (IQ) on processed images. However, methods to evaluate IQ on processed images objectively are lacking and therefore subjective evaluations incorporating human observers, such as proposed by van Ongeval et al. [4], are used today.

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; CIQ, clinical image quality; CLB, clustered lumpy background; d′, detectability index; DICOM, digital imaging and communications in medicine; DM, digital Mammography; HVS, human visual system; IQ, image quality; MO, model observer; NM, noise mask; NPW, non-pre-whitening model observer; PC, proportion correct; PIQ, physical image quality; PM, probability map; WN, white noise; 2-AFC, 2-alternative forced choice.

[*] Corresponding author at: Dutch Reference Centre for Screening (LRCB), Radboud University Medical Centre, PO Box 6873, 6503 GJ Nijmegen, The Netherlands.

E-mail address: r.bouwman@lrcb.nl (R.W. Bouwman).

http://dx.doi.org/10.1016/j.ejmp.2016.11.109
1120-1797/

Current methodologies to assess IQ, referred to as physical image quality (PIQ), use acquired images of simple homogeneous phantoms and require a strictly monotonic relation between dose and pixel value which is constant over the image [5,6]. As a result, image processing is not applied to images used for PIQ and images scored by radiologists might be different in quality. In order to objectively assess the IQ of processed images, which we refer to as clinical image quality (CIQ), a new methodology is required. We believe that this new methodology requires the development of a clinically realistic phantom which takes into account the complexity and characteristics of a real breast and methods of assessing the quality of processed images obtained with the phantom. In this study we focus on the latter.

Model observers (MO) have been suggested for the evaluation of image quality in medical imaging [7,8]. MOs are computer models which can be used to perform a particular task, for example the detection of a mass in a mammogram. There are two types of MOs: mathematical and statistical [9]. The first type uses signal transfer and noise properties obtained from linear system theory metrics. The second type of MO is based on the statistical properties of pixel values in images. Thus, statistical MOs are solely based on the image data and do not make any assumptions about the relationship between dose and pixel value. This is important since this relationship might be unknown after applying image processing. When measuring CIQ, it is important that the outcome of the MO is related to human observer performance. So before MOs can be introduced for CIQ analysis, the correlation between human and MO performance needs to be known and validated for different tasks and system characteristics.

The non-pre-whitening (NPW) MO is one of the statistical MOs that is a candidate for CIQ assessment. Rolland and Barrett [8] demonstrated that the NPW observer fails to predict human observer performance in lumpy backgrounds. Burgess [10] subsequently demonstrated that a better prediction could be obtained if a spatial frequency filter is applied which mimics the response of the human eye, the eye filter. In the literature several approximations of the eye filter can be found [10–19]. The human visual system (HVS) however, cannot be fully characterised by only using an eye filter. Avanaki et al. demonstrated that inclusion of the psychometric function [20] improved the match in absolute performance between human and MO and that the inclusion of a masking filter further improved this match [21]. In the current study we have investigated whether NPW MOs incorporating various HVS aspects have potential to be used for CIQ analysis by evaluating the correlation between human and MO performance. For this purpose their performance was investigated for a simple detection task in simulated white noise (WN) and clustered lumpy background (CLB, [22]) images. These type of images were chosen to represent images from a quantum noise limited ideal system and clinical realistic structures respectively. Since different formulations of the NPW MO are used throughout this manuscript, we use NPW as a generic term for all the MOs considered. Although the ultimate aim of the authors is to evaluate the NPW MO for use in image quality analysis of processed images, neither defining image quality nor proposing the methodology to evaluate image quality is part of this study. Once the NPW MO has proven its ability to predict human observer performance sufficiently well, image quality can be defined and a methodology to evaluate image quality can be proposed.

2. Materials and methods

2.1. Non-pre-whitening model observer

In detection experiments the observer has to decide on the presence or absence of a signal (for example a mass) by dividing the images into two classes: signal present (class 1) and signal absent (class 2). The goal of a detection experiment is thus to determine the separation between the two classes. This separation is expressed by the detectability index (d') and is estimated from the decision variables (λ) assigned to the images of both classes. For linear MOs, λ is described by a linear transformation between an observer template ($\boldsymbol{w}$) and the image vector ($\boldsymbol{g_i}$) via:

$$\lambda_i = \boldsymbol{w}^t \boldsymbol{g_i} + \varepsilon \quad (1)$$

where two dimensional images are written as a one dimensional vector (bold symbols) to reduce the computational complexity, i represents the image class, t is the transpose, and ε an optional additive model of internal noise. The addition of internal noise degrades the performance of the MO and could be used when matching model and human observer performance [9]. The inclusion of internal noise is beyond the scope of this paper and has not been evaluated in this study.

Among the statistical MOs the NPW model observer requires the least computational power and is therefore attractive for IQ analysis. The NPW MO correlates the image with the expected signal, meaning that the observer template ($\mathbf{w}$) equals the signal ($\mathbf{s}$), to be detected ($\boldsymbol{w}_{NPW} = \boldsymbol{s}$) [9].

2.2. Human visual system

In medical imaging, human observers view clinical images on a diagnostic workstation which translates pixel values to presentation values and luminance in a standardized way [7,23]. The perceived luminance (brightness) of the image is transferred to the human observer by the eye and to correctly predict human observer response the processes that take place in the human visual system (HVS) should be included in the MO. In this paper three different processes that occur in the HVS are evaluated: (1) the contrast transfer of the human eye, further referred to as the eye filter, (2) the relationship between the perceived signal and detection, further referred to as the probability map (PM) and (3) masking of the contrast arising from fluctuations in luminance in the background, further referred to as noise masking (NM). These three processes are further explained in the next subsections. However, these three processes do not fully describe the HVS and further processes are involved [24,25]. These are difficult to model and as a first practical step have been omitted from the current study.

2.3. Eye filter

The eye filter is based on the human contrast sensitivity function [9] and is defined and applied in the spatial frequency domain. The decision variable, λ, after inclusion of the eye filter, E, is estimated via:

$$\lambda_i = \left[\boldsymbol{E}^t \cdot \boldsymbol{E} \cdot \boldsymbol{s}\right]^t \cdot \boldsymbol{g}_i \quad (2)$$

where $\boldsymbol{s}$ is the signal and $\boldsymbol{g}_i$ the image vector with pixel values converted to luminance values. The luminance is calculated from the pixel values using the display function of the monitor [7]. In the literature several approximations of the eye filter can be found. A brief overview of 11 different approximations used in this study is given below. Fig. 1 shows the 11 approximations of the eye filter, each normalized to its peak response. It is worth noting that we have limited ourselves to 11 approximations, which have been used in combination with MOs in the past, but that more eye filters can be found in the literature.

Several eye filters use a fixed mathematical expression which does not take into account the perceived luminance nor the process of signal transfer to the fovea of the eye. These eye filters

Please cite this article in press as: Bouwman RW et al. Can the non-pre-whitening model observer, including aspects of the human visual system, predict human observer performance in mammography? Phys. Med. (2016), http://dx.doi.org/10.1016/j.ejmp.2016.11.109