Quality Improvement Study

Medicine®

OPEN

A no-reference respiratory blur estimation index in nuclear medicine for image quality assessment

David Morland, MD, MSc[a,b,c,*], Paul Lalire, MD, MSc[a], Sofiane Guendouzen, MSc[d], Dimitri Papathanassiou, MD, PhD[a,b,c], Nicolas Passat, PhD[c]

Abstract

Few indexes are available for nuclear medicine image quality assessment, particularly for respiratory blur assessment. A variety of methods for the identification of blur parameters has been proposed in literature mostly for photographic pictures but these methods suffer from a high sensitivity to noise, making them unsuitable to evaluate nuclear medicine images. In this paper, we aim to calibrate and test a new blur index to assess image quality.

Blur index calibration was evaluated by numerical simulation for various lesions size and intensity of uptake. Calibrated blur index was then tested on gamma-camera phantom acquisitions, PET phantom acquisitions and real-patient PET images and compared to human visual evaluation.

For an optimal filter parameter of 9, non-weighted and weighted blur index led to an automated classification close to the human one in phantom experiments and identified each time the sharpest image in all the 40 datasets of 4 images. Weighted blur index was significantly correlated to human classification ($\rho = 0.69$ [0.45;0.84] $P < .001$) when used on patient PET acquisitions.

The provided index allows to objectively characterize the respiratory blur in nuclear medicine acquisition, whether in planar or tomographic images and might be useful in respiratory gating applications.

Abbreviations: 18F-FDG = 18F-Fluorodeoxyglucose, MTV = metabolic tumor volume, PET = positron emission tomography, SUL = standardized uptake value normalized to lean body mass, SUV = standardized uptake value, TLG = total lesion glycolisis.

Keywords: gating, PET, respiratory blur

1. Introduction

Nuclear medicine is a domain in full expansion, broadening the spectrum of its applications. The administered activity of radiopharmaceuticals is kept as low as possible to limit patient radiation exposure. Imaging times are increased accordingly to compensate low count rates and decreased signal-to-noise ratio. A complete nuclear medicine acquisition thus takes usually several minutes and is subject to respiratory blur. Diaphragm motion amplitude can reach 10 cm during respiratory cycle[1] and this phenomenon may have clinical consequences. The concentration of radiopharmaceutical within a given structure is in particular spread out over a larger area leading to a respiratory blur and an underestimation of lesion uptake.

A wide range of indexes has been developed to characterize radiopharmaceuticals uptake in nuclear medicine[2]: intensity of the uptake (SUV, SUL), volume (MTV, TLG) and more recently textural parameters.[3] Few indexes are however available for image quality assessment once the image acquired, particularly for respiratory blur assessment. This question might be of interest with the development of respiratory gating, particularly on PET/CT systems.

Motion blur, caused by the relative motion of a structure during image capturing, has 2 main components: angle and amplitude. In this paper, we focus exclusively on amplitude estimation as the predominant axis of respiratory motion is craniocaudal.

A variety of methods for the identification of blur parameters has been proposed in literature,[4] mostly on photographic pictures. Edge detection techniques are widely documented whether with first or second order derivative, Sobel operator,[5] Canny detector[6] or wavelet transform methods.[4] However these methods are very sensitive to noise.[7] Nuclear medicine acquisitions have a lower signal-to-noise ratio than photographs which can mislead the edge detection. In their review article, Tiwari and al[4] tested 2 methods based on frequency domain (radon transform method and cepstral methods) that showed a similar sensitivity to noise adding, making them unsuitable to nuclear medicine image evaluation.

Editor: Majid Assadi.

All procedures performed in study involving human participants were in accordance with the ethical standards with the 1964 Helsinki Declaration and its later amendments.

The authors have no funding and conflicts of interests to disclose.

[a] Médecine Nucléaire, Institut Godinot, [b] Laboratoire de Biophysique, UFR de Médecine, Université de Reims Champagne Ardenne, [c] CRESTIC EA 3804, Université de Reims Champagne Ardenne, [d] Physique Médicale, Institut Godinot, Reims, France.

[*] Correspondence: David Morland, Médecine Nucléaire, Institut Godinot, 1 rue du general Koenig CS80014, 51726 Reims Cedex, France (e-mail: david.morland@reims.unicancer.fr).

How to cite this article: Morland D, Lalire P, Guendouzen S, Papathanassiou D, Passat N. A no-reference respiratory blur estimation index in nuclear medicine for image quality assessment. Medicine 2019;98:48(e18207).

Received: 21 June 2019 / Received in final form: 13 October 2019 / Accepted: 31 October 2019

http://dx.doi.org/10.1097/MD.0000000000018207

In this paper we test an automated estimation index to assess the respiratory blur in nuclear medicine images, whether planar or tomographic.

2. Methods

2.1. Blur index calculation

2.1.1. Thresholding. Let I be the original image of size $m\,x\,n\,x\,p$. The first step is the creation of a binary mask M of I whose value is 0 when the voxel value is less than 100 counts. The threshold was determined on ad hoc basis. This step is performed in order to avoid any influence from pixels/voxels outside the phantom or patient body on the blur index.

2.1.2. Blur index calculation. Blur index is based on the article of Crete and al[8] and adapted to accept 2D and 3D nuclear medicine images. A blurred image B is created using a low-pass filter h of length L in the direction of the respiratory movement (z axis in 3D images, y axis in 2D image). B is cropped to be of the same size as I. The adequate L parameter is evaluated in the first part of this article.

$$h = \frac{1}{L}[1, \ldots, 1] \quad (1 reapeated\, L\, times)$$
$$B = h \times I$$

Index calculation: The absolute difference images DI and DB studying the variations of neighboring pixels are initialized as all zeros matrix of $m\,x\,n\,x\,p$ and then computed as follow:

$$DI(i,j,k) = Abs(I(i,j,k) - I(i,j,k-1)) for\, i = 1\, to\, m,\, j = 1\, to\, n,\, k = 1\, to\, p-1$$
$$DB(i,j,k) = Abs(B(i,j,k) - B(i,j,k-1)) for\, i = 1\, to\, m,\, j = 1\, to\, n,\, k = 1\, to\, p-1$$

In order to analyze the variations of the neighboring pixels after the blurring effect, an image DV is created: if the variation is low then the original image was already blur.

$$DV(i,j,k) = Max(0,\, DI(i,j,k) - DB(i,j,k)) for\, i = 1\, to\, m,$$
$$j = 1\, to\, n,\, k = 1\, to\, p$$

The comparison of the variations from the initial picture is computed as follow, using the binary mask M:

$$sI = \sum_{i,j,k=1}^{m,n,p} DI(i,j,k) \times M(i,j,k)$$
$$sV = \sum_{i,j,k=1}^{m,n,p} DV(i,j,k) \times M(i,j,k)$$

The final index *Blur* varies from 0 (sharp) to 1 (blurred) and is given by:

$$Blur = \frac{sI - sV}{sI}$$

2.1.3. Blur index ponderation. High intensity lesions can artificially decrease the blur index by increasing the contrast at the interface of the lesion and the background. This phenomenon can theoretically decrease the blur index when patients exhibit lots of high intensity voxels. An estimation of the number of high intensity voxels is given by the ratio R between the number of voxels exceeding a predetermined threshold (expressed as a percentage P of the maximum intensity value in the original image I) and the number of voxels corresponding to the patient (voxels whose value is 1 in the mask M).

$$Weighted\, blur = R \times Blur$$

2.2. Experiments

2.2.1. Filter calibration and high intensity voxel threshold (P) determination. The length L of the low pass filter was determined based on a numerical 2D simulation. A moving disk was simulated on a 128 × 128 matrix with pixel size of 2 × 2 mm (motion length 20 mm, 10 cycles per second). Two dynamic acquisitions consisting of 200 frames of 1 second were generated: 1 with and 1 without disk movement. Poisson noise was added to each frame. Blurred random images were obtained by summing 40 randomly selected frames in the 200 frames available in the simulated moving acquisition. Static random images were obtained by summing 40 randomly selected frames in the 200 frames available in the simulated non-moving acquisition.

Five hundred datasets of 4 images were reconstructed, each composed of 1 static random image and 3 blurred random images. The sharpest image was then identified based on the blur index calculated for different values of L.[3,5,7,9,11,13] Sharpest image identification was considered successful if it identified the static random image.

This process was tested for several disk diameter values (5 mm, 10 mm, and 20 mm) and several pixel intensities (2, 5, and 10 times the background whose value had been fixed to 1 arbitrary unit).

The threshold P was determined based on 30 consecutive PET/CT performed for clinical purpose. All were acquired on a Discovery 710 system (General Electrics, Milwaukee, WI) after an intravenous injection of 3 MBq/kg of 18F-FDG. An external observer was asked to rate the respiratory blur from 1 (absent) to 5 (major blur). Correlation between perceptual blur and weighted blur index were computed using Pearson correlation coefficient for all threshold value (ranging from 0% to 100% SUVmax).

2.3. Phantom experiments

2.3.1. Input data. We used a dynamic thorax phantom (Model 008A, Computerized Imaging Reference System, Inc.) with a spherical insert of 8 ml (2.5 cm of diameter) filled with 20 MBq of [99mTc] Pertechnetate (gamma camera acquisition) or 20 MBq of [18F] FDG (PET/CT acquisition). The phantom was positioned at the center of the field of view: the center of the insert was located 65 mm right from the center. Motion length of the phantom was set to 20 mm at 10 cycles per minute.

2.3.2. Image acquisitions. For gamma-camera acquisitions, 200 images of 1 second were acquired in planar mode on a 128 × 128 matrix using a Symbia T2 system (Siemens Medical Solutions, USA). For PET acquisitions, a 200 second volume was acquired in list-mode on a Discovery 710 system (General Electrics, Milwaukee, WI) and reconstructed to obtain 200 frames of 1 second (OSEM: 24 subsets and 2 iterations, no attenuation correction, reconstructed slice thickness of 3.27 mm, Butterworth post-filter with 6.4 mm cut-off).

Random blurred images were obtained by summing 40 randomly selected frames in the 200 frames available in each acquisition. Thirty datasets of 4 random blurred images were reconstructed for gamma-camera acquisition and another 30 datasets for PET acquisition.

2.3.3. Images rating. Each dataset of 4 images was sorted in ascending order of blurriness (from 1 to 4) by an independent observer –considered as the gold standard– and afterwards along the automated weighted and non-weighted blur index. An error

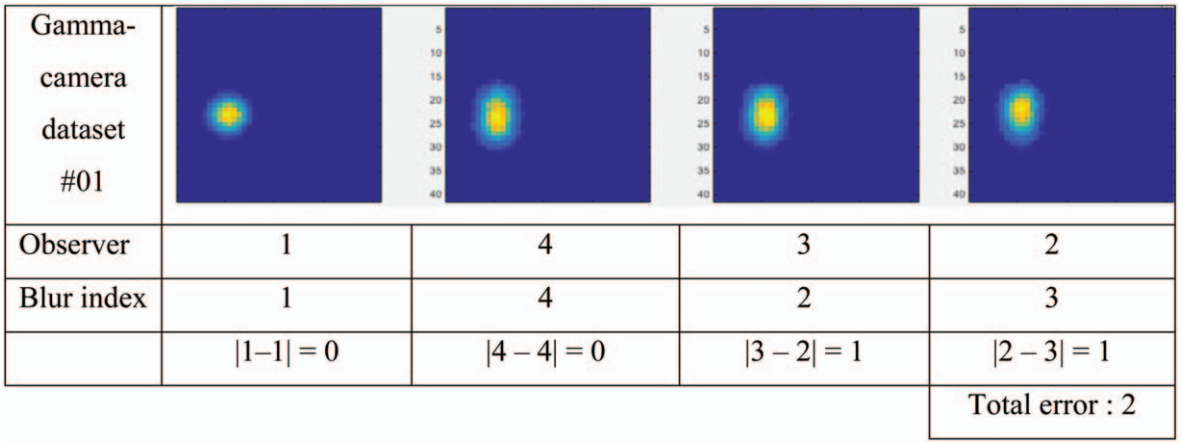

Figure 1. Example of a dataset rating with error score calculation between human and automated rating. This dataset is sorted in ascending order of blurriness (from 1 to 4) based on observer evaluation and blur index. An example of error score calculation between the 2 methods is provided.

score was calculated between the 2 ratings, corresponding to the sum of the absolute difference of rank of each image, as shown below. By construction, this error score is always even and ranges from 0 to 8 with 0 (perfect agreement), 2 (slight disagreement), 4 (mild disagreement), 6 (subtotal disagreement), 8 (complete disagreement). An example is provided in Figure 1.

2.3.4. Patient experiments. Thirty consecutive PET/CT, different from those used for threshold determination, were retrospectively selected. All were acquired on a Discovery 710 system (General Electrics, Milwaukee, WI) after an intravenous injection of 3 MBq/kg of 18F-FDG. An external observer was asked to rate the respiratory blur from 1 (absent) to 5 (major blur). Correlation between perceptual blur and blur index were computed using Pearson correlation coefficient on R software.[9]

3. Results

3.1. Filter calibration

We report the rate of successful identification of the sharpest image on simulated acquisitions in Table 1. We retained an optimal L value of 9 with a mean of 88.7% of successful identification throughout all simulations.

Optimal threshold P was 42% of the SUVmax ($\rho=0.64$) as seen in Figure 2.

3.2. Phantom experiments

No ranking differences were seen between weighted and non-weighted blur index.

For gamma-camera 2D acquisitions: mean error score value was 0.5 with perfect agreement (error score 0) for 30/40 datasets and slight disagreement (error score 2) in the remaining 10. No mild, subtotal or complete disagreements were noted. The sharpest image was always concordant between visual and automated ranking.

For PET 3D acquisitions: mean error score value was 0.6 with only slight disagreement (error score 2) in 12/40 datasets. The sharpest image was again always concordant between visual and automated ranking.

3.3. Patient experiments

Non-weighted blur index was not correlated to perceptual blur ($\rho=0.08$ [−0.28;0.43] $P=.64$). Weighted blur index was significantly correlated to perceptual blur ($\rho=0.69$ [0.45;0.84] $P<.001$), Figure 3.

4. Discussion

Low signal-to-noise ratio relative to photographs is the main concern in blur estimation in nuclear medicine acquisition. We

Table 1

Success rate of sharpest image identification for various blur index L-parameter values.

	L=3	L=5	L=7	L=9	L=11	L=13
Diameter: 5 mm/Intensity: 2	30.8%	37.0%	59.4%	68.8%	48.2%	50.4%
Diameter: 5 mm/Intensity: 5	12.2%	30.2%	92.6%	92.0%	87.6%	82.6%
Diameter: 5 mm/Intensity: 10	10.0%	13.0%	100%	100%	99.6%	99.6%
Diameter: 10 mm/Intensity: 2	9.0%	13.6%	48.8%	79.2%	85.4%	86.2%
Diameter: 10 mm/Intensity: 5	1.0%	2.0%	92.2%	100%	100%	100%
Diameter: 10 mm/Intensity: 10	0.0%	0.0%	98.2%	100%	100%	100%
Diameter: 20 mm/Intensity: 2	53.4%	55.8%	21.6%	58.6%	47.8%	75.4%
Diameter: 20 mm/Intensity: 5	22.0%	57.8%	65.8%	100%	100%	100%
Diameter: 20 mm/Intensity: 10	2.2%	19.4%	42.6%	100%	100%	100%
Total	**15.6%**	**25.4%**	**69.0%**	**88.7%**	**85.4%**	**88.2%**

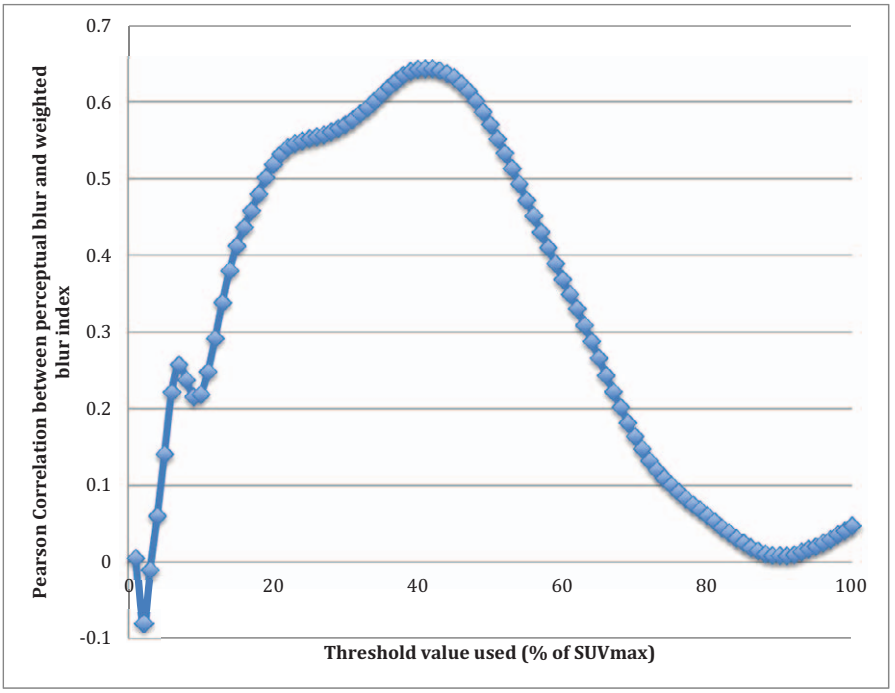

Figure 2. Pearson correlation between perceptual blur and weighted blur index along threshold value (P).

adapted the blur index published by Crete and al[8] by using a threshold eliminating the area of the image with the lowest signal-to-noise ratio and taking into account the number of high intensity voxels. The thresholding excludes most of the voxels of the background, where no signal is expected: the background can indeed increase the blur index value as it is a large area of low contrast between neighboring voxels. By construction, the blur index is sensitive to high intensity voxels which can artificially decrease its value due to the high contrast between the lesion and the neighboring voxels. An estimation of the ratio of high intensity voxels over total patient voxels was proposed to compensate this phenomenon. High intensity voxels were defined as voxels exceeding 42% of the SUVmax of the image based on a first training dataset. Pearson correlation between perceptual blur and weighted blur index decreased when using a higher percentage presumably because of the low number of voxels selected. With a lower percentage, low intensity voxels are also selected and lead to a decreased Pearson correlation as well.

In phantom experiments, the revised non-weighted and weighted blur index led to an automated classification close to the human one with perfect agreement in the order in between 70% and 75% of the datasets. The remaining discrepancies gathered only slight

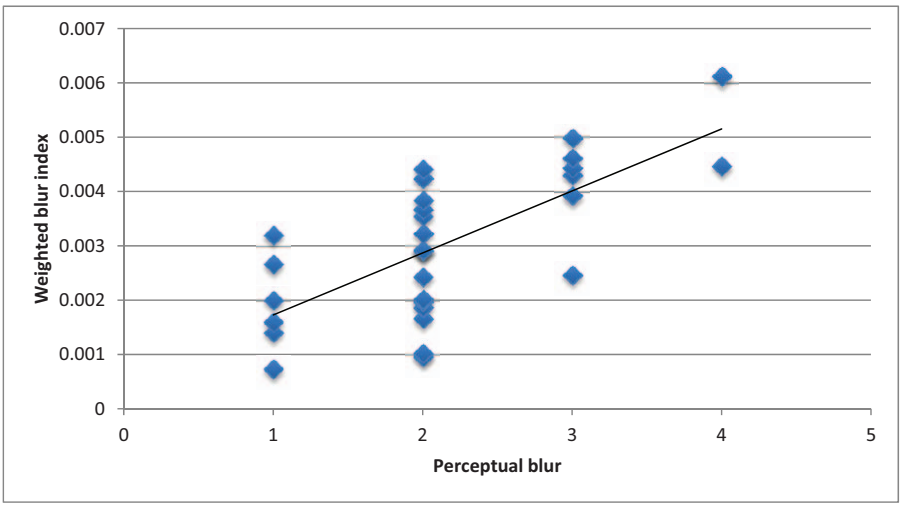

Figure 3. Relationship between weighted blur index and human based blur evaluation (perceptual blur).

disagreements, that is to say a permutation of 2 consecutive images relative to the human ranking. The sharpest image was always concordant. The high performance of these indexes to select the sharpest arrangement of frame in phantom experiments might be promising to propose a derived respiratory gating algorithm.

In patient experiments, non-weighted blur index did not show any correlation with perceptual blur: this is probably mainly due to the interpatient variability. The number of high intensity lesions varies from one patient to another by contrast to the phantom experiments in which only 1 high intensity lesion was tested. Once this correction made, weighted blur index was significantly correlated to perceptual blur ($P < .001$).

Respiratory blur is a cause of image degradation in nuclear medicine.[10] The provided weighted blur index allows to objectively evaluate its severity. This fully automated index can be a first step toward a machine-learning based blur estimation which is a field of mounting interest.[11–13]

5. Conclusion

The provided index allows to objectively characterize the respiratory blur in nuclear medicine acquisitions, whether in planar or tomographic images and might be useful in respiratory gating quality assessment.

Author contributions

Conceptualization: David Morland.
Data curation: David Morland, Paul Lalire, Guendouzen Sofiane.
Formal analysis: David Morland, Paul Lalire, Guendouzen Sofiane.
Methodology: David Morland.
Project administration: Dimitri Papathanassiou, Passat Nicolas.
Software: David Morland, Paul Lalire.
Writing – original draft: David Morland, Paul Lalire.
Writing – review & editing: Guendouzen Sofiane, Dimitri Papathanassiou, Passat Nicolas.
David Morland orcid: 0000-0001-8738-4841.

References

[1] Wang H, Lu T, Liing R, et al. Relationship between chest wall motion and diaphragmatic excursion in healthy adults in supine position. J Formos Med Assoc 2009;108:577–86.
[2] Nagamachi S, Wakamatsu H, Fujita S, et al. Which FDG PET/CT index (SUVmax, metabolic volume, total lesion glycolysis) is the most reliable to predict therapeutic effects in cancer therapy? J Nucl Med 2010;51 (supplement 2):565–1565.
[3] Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? J Nucl Med 2015;56:1642–4.
[4] Tiwari S, Shukla VP, Singh AK, et al. Review of motion blur estimation techniques. J Image Graph 2014;176–84.
[5] Ong E, Lin W, Lu Z, et al. A no-reference quality metric for measuring image blur. In: Seventh International Symposium on Signal Processing and Its Applications, 2003 Proceedings. 2003. 1:469–72.
[6] Marziliano P, Dufaux F, Winkler S, et al. A no-reference perceptual blur metric. In: Proceedings International Conference on Image Processing. 2002. p. III-III.
[7] Ferzli R, Karam LJ. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans Image Process 2009;18:717–28.
[8] Crete F, Dolmiere T, Ladret P, et al. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In: Human Vision and Electronic Imaging XII [Internet]. International Society for Optics and Photonics; 2007 [cité 4 janv 2019]. p. 64920I. Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6492/64920I/The-blur-effect–perception-and-estimation-with-a-new/10.1117/12.702790.short.
[9] R: a language and environment for statistical computing [Internet]. [cité 4 janv 2019]. Available at: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.
[10] Song C, Yang Y, Qi W, et al. Motion-compensated image reconstruction vs postreconstruction correction in respiratory-binned SPECT with standard and reduced-dose acquisitions. Med Phys 2018;45:2991–3000.
[11] Kong B, Wang X, Li Z. Cancer metastasis detection via spatially structured deep network. In: Niethammer M, Styner M, Aylward S, et al., éditeurs. Information Processing in Medical Imaging. Springer International Publishing: Boone, NC, USA; 2017. p. 236–248
[12] Kong B, Sun S, Wang X. Invasive cancer detection utilizing compressed convolutional neural network and transfer learning. In: Frangi AF, Schnabel JA, Davatzikos C, éditeurs. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer International Publishing: Granada, spain; 2018. p. 156–164
[13] Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. Eur Radiol 2019;29:6191–201.