Compact Rotation Invariant Descriptor for Non-Local Means

Nicholas Dowson[a] and Olivier Salvado[a]

[a]Australian e-Health Research Centre, CSIRO, Brisbane, Australia;

ABSTRACT

Non-local means is a recently proposed denoising technique that better preserves image structures than other methods. However, the computational cost of non-local means is prohibitive, especially for large 3D images. Modifications have previously been proposed to reduce the cost, which result in image artefacts. This paper proposes a compact rotation invariant descriptor. Testing demonstrates improved denoising performance relative to optimized non-local means. Rotation invariant non-local means is an order of magnitude faster.

Keywords: Image Filtering, Non-local means, Rotation Invariance, Feature Selection

1. INTRODUCTION

A trade-off remains between image acquisition time and noise levels for certain types of image, *e.g.* 3D Magnetic Resonance (MR) and Positron Emission Tomography (PET) images. It is often impractical to increase acquisition time beyond a certain limit, due to movement of the subject and the time constraints associated with any expensive shared resource. Hence, denoising filters are required.

Classical denoising approaches assume voxels should resemble their neighbours and enforce this by convolving with a kernel. The kernel is selected to trade-off retention of image structure with removal of noise. Algorithms that are adaptive to regional intensity gradients allow greater smoothing without an associated destruction of image structure. This is simply a tighter assumption: voxels should resemble *similar*, nearby neighbours. One popular approach is anisotropic diffusion proposed by Perona and Malik,[1] which relies upon "diffusing" intensities orthogonally to the local intensity gradient, thereby preserving edges and corners. The bilateral filter is a more generalized example of such an approach, independently proposed by Smith & Brady[2] and Tomasi & Manduchi.[3]

Recently, Buades *et al.*[4] and Awate & Whitaker[5] proposed the non-local means denoising method. Non-local means is similar to the bilateral filter, but with two key exceptions. First, the weighted sum of voxels over a much larger neighbourhood (possibly including the entire image) is computed. Second, the similarity of a small patch surrounding each voxel is used to weight their influence, rather than the voxel intensity itself. In contrast to previous approaches, the assumption is that images are self-similar and voxels should resemble other voxels of similar *structure*, anywhere within the image.

The disadvantage of non-local means is its prohibitive computational cost, making it impractical for use on large 3/4D data-sets. Coupe *et al.* proposed reducing the cost in three ways:[6] restricting matches to patches within a local region, writing overlapping patches of voxels so that only a subset of positions need be considered, and computing weights only if the mean and covariance of two patches are sufficiently similar. The first modification means that only a subset of available information is used for smoothing, while the second creates a plaid pattern of artefacts within the image.

This work proposes another approach to speed up non-local means, but without the negative side-effect of artefacts: a compact rotation invariant local descriptor. The descriptor is aligned along the axis of lowest local gradient. In addition, a more compact descriptor with only the most relevant features from a local patch is used, allowing the use of a hash-space similar that that proposed by Paris and Durand for the bilateral filter.[7]

The remainder of the paper is organized as follows. In Section 2 a background to non-local means is provided, followed by a description of the proposed modifications in Section 3. Some experiments are performed and discussed in Section 4, before concluding in Section 5.

Further author information: (Send correspondence to Nicholas Dowson). Address: Australian e-Health Research Centre, Level 7 - UQ CCR Building 71/918, Royal Brisbane and Women's Hospital, Herston, Qld, 4029, Australia. E-mail: nicholas.dowson@csiro.au

Medical Imaging 2010: Image Processing, edited by Benoit M. Dawant, David R. Haynor,
Proc. of SPIE Vol. 7623, 76233C · © 2010 SPIE · CCC code: 1605-7422/10/$18 · doi: 10.1117/12.844088

2. NON-LOCAL MEANS

Non-local means operates by computing the similarity of intensities of a patch of voxels surrounding a given location, $\mathbf{x}$, to a set of voxels at other locations, Y. Images of dimension D are treated as a function, f, of position $\mathbf{x} \in X : \mathbb{Z}^D$, where X is a set of locations in the image. Y may be a function of $\mathbf{x}$ such that $Y[\mathbf{x}] \subseteq X$. The image is assumed to have been obtained using a process which includes additive noise, *i.e.* $f[\mathbf{x}] = \hat{f}[\mathbf{x}] + n[\mathbf{x}]$, where n is the unknown additive noise, and $\hat{f}$ is the noise-free image. Square brackets are used to indicate that $\mathbf{x}$ is discretely defined.

The similarities between the intensities at a location $\mathbf{x}$ and all corresponding locations $\mathbf{y} \in Y$ are used to obtain a weight function $w_{\mathbf{xy}}$. The weights are used to obtain the non-local means estimate of intensity $\tilde{f}$ using a weighted sum:

$$\tilde{f}[\mathbf{x}] = \frac{\sum_{\mathbf{y} \in Y} w_{\mathbf{xy}} f[\mathbf{y}]}{\sum_{\mathbf{y} \in Y} w_{\mathbf{xy}}} \tag{1}$$

The weights are obtained from the intensities at each offset $\mathbf{p}$ within the set of offsets or patch, P, surrounding locations $\mathbf{x}$ and $\mathbf{y}$. The set of intensities defined by offsets from each location is concatenated into an $|P|$-dimensional vector, $\mathbf{f}[\mathbf{x}]$:

$$w_{\mathbf{xy}} = \exp\left(-h^{-2}(\|\mathbf{f}[\mathbf{x}] - \mathbf{f}[\mathbf{y}]\|^2_{2,k})\right) \tag{2}$$

where $||\cdot||_{2,k}$ is the L_2 Euclidean distance metric. The smoothing parameter, h, is used to vary the amount of smoothing that is applied. h should be chosen based upon the estimated noise variance, σ^2, and the number of offsets, $|P|$ used:[6]

$$h^2 = 2\sigma^2|P|\beta \tag{3}$$

where β is a tuning parameter that Coupe suggests leaving to one and $|\cdot|$ indicates the number of elements in a set.

Buades[4] and Coupe[6] use the set of offsets, P, defining a 3^D patch of offsets surrounding each location. To use pattern recognition terminology, the set of intensities at each offset, $\mathbf{p}$, may be treated as a set of *features*, each of which is partly correlated with the central intensity, $f[\mathbf{x}]$. Non-local means uses a radial basis function with a Gaussian kernel to obtain the intensity of maximum likelihood.

Converting the intensity of a voxel and its 26 neighbours into a 27-component vector is a convenient formulation. However, the orientation of a local image structure is arbitrary relative to the image axes. Likewise, structures are arbitrarily orientated to the location of intensities within each vector. Hence, the similarity of two regions within the image that resemble each other but of different orientations will not be exploited. For example, a horizontal edge and vertical edge, which are identical in all other respects, may have a low co-weight even though they are the same structure at different orientations. Similarly, mirrored structures will also have low co-weights. This implies that much of the self-similar to be found in images are not being explored. Some examples of rotation variance for otherwise identical structures are shown in Figure 1. A rotation invariant descriptor may be used to rectify this deficiency.

3. ROTATION INVARIANT NON-LOCAL MEANS

Considering a 3D voxel and its adjacent neighbours, the gradient may be directly computed along 13 possible axes, as shown in Figure 1. Three of the axes are aligned with the image axes, as shown in Figure 2. Rotational invariance may be obtained by adopting some convention for orientating local image patches, that allows similar structures of arbitrary orientation to be placed in the same coordinate system.

This work treats the axis of lowest gradient as the primary axis, with the positive direction in the direction of increasing intensity. Similarly the *orthogonal* axis with the next lowest gradient is selected as the secondary axis, leaving only one possible selection for the tertiary axis. The three axes just described are collectively referred to as the *local axes*. Other conventions could have been used, but the lowest gradient axis is assumed to give the most information about the central voxel intensity.

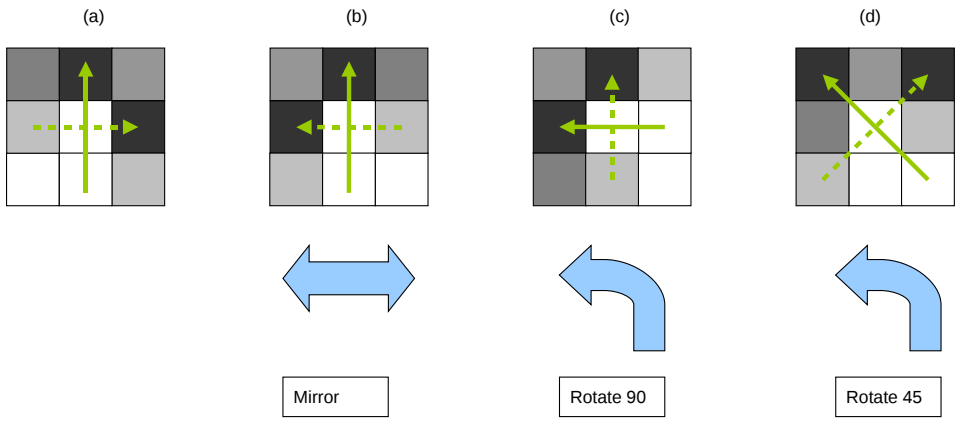


Figure 1. Examples of an otherwise identical 3x3 descriptor at different orientations for a 2D image patch. The orientation with the largest gradient is shown by the thin solid arrows and the orthogonal gradient with the next highest by the dotted arrow. Intensity is depicted by gray level.

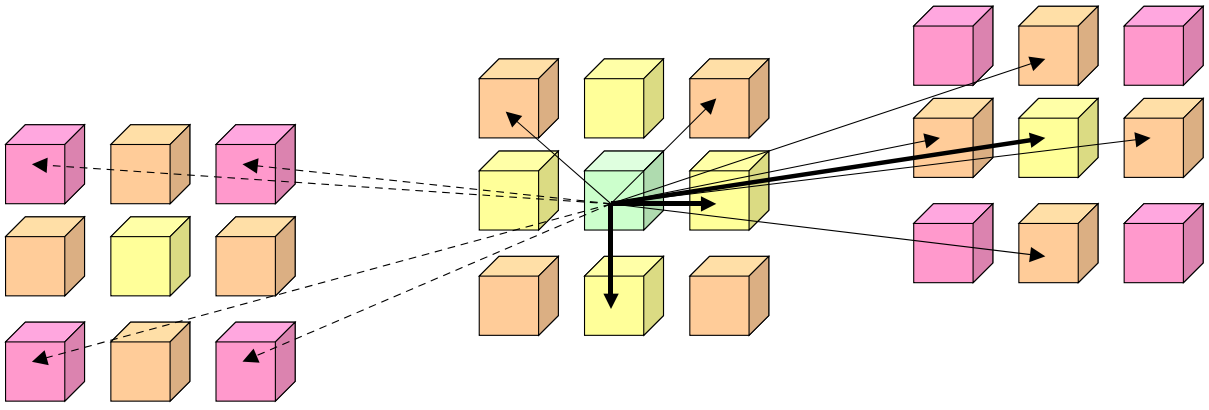


Figure 2. 13 axes (ignoring direction) are possible when considering the adjacent 26 voxels to a central voxel.

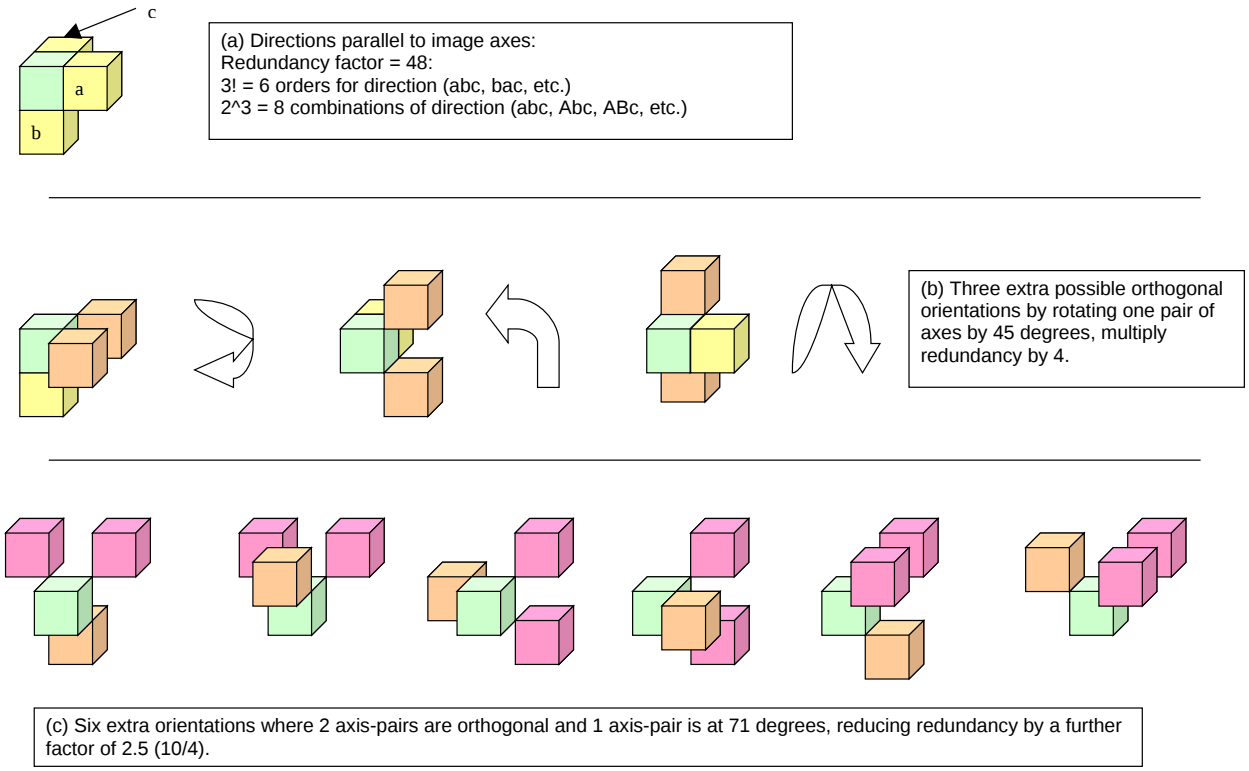


Figure 3. Triples of orientation axes. (a) Allowing orientations parallel to the image axes gives a redundancy factor of 48. (b) Allowing one pair of axes to rotate by 45° increases this factor to 192.

If only those axes parallel to the image coordinate system are considered, the redundancy within the descriptor-space is reduced by a factor of 48: 6 permutations or ways of ordering the three axes (3!) multiplied by the 8 combinations of gradient directions (2^3). This is illustrated in Figure 3a.

The redundancy factor is defined as follows. Assuming that the number intensities is a finite, N_i, the number of possible descriptors is N_i^{27}. If the values are re-ordered according to the convention described in the directly preceding paragraph, the number of possible descriptors is reduced to $N_i^{27}/k_{\text{redundancy}}$, where $k_{\text{redundancy}} = 48$ in this case.

Considering the additional six axes obtained by rotating a single pair of axes by 45°, gives three further ways to select orthogonal axes. This reduces the redundancy by a further factor of 4: $\frac{1+3}{1}$. In this case, case all possible selections of axis triples are orthogonal, as shown in Figure 3b.

To obtain further redundancies the requirement for orthogonal axes must be relaxed. If the enforcement of orthogonality is softened to allow one pair of axes at an angle of 71°, a reduction by another factor of 2.5 is obtained, as there are an additional six ways to select "almost" orthogonal axes: $\frac{4+6}{4}$. The six additional orientations are shown in Figure 3c.

3.1 A compact descriptor

Assuming that the local axes are independent, because they are orthogonal, the axes can be assumed to make separate predictions about the content of the noisy central voxel. Ignoring the central voxel, the two intensities may be transformed into an equivalent representation: an axis-gradient and an axis-mean. Although the axis-gradient describes the local voxel structure, experiment shows that the axis-mean better predicts the intensity in the central voxel. Hence the axis-gradients are neglected.

The remaining intensities are on axes that are not orthogonal to the local axes, and are thus correlated with the intensities on the local axes. This implies that much of this data is redundant. However, this assumes a linear intensity model with no cross-terms, which is seldom the case in practice. Hence, rather than neglecting this data entirely, a weighted mean of axis-means is calculated. The weighting is based upon the axis-gradients, Δ_a, where a is the axis index, μ_a is the corresponding axis-mean, and f is the value of the central voxel:

$$\bar{\mu} = \frac{f + \sum_{\forall a \in \text{axes}} e^{-\frac{\Delta_a^2}{h^2}} \mu_a}{1 + \sum_{\forall a \in \text{axes}} e^{-\frac{\Delta_a^2}{h^2}}} \tag{4}$$

This weighting method is chosen, because axes with low gradient axes are generally better predictors of the intensity in the central voxel. However, certain complex image structures, such as a narrow homogeneous region sandwiched between two regions of similar intensity, may be smoothed out by this assumption. This necessitates the inclusion of the central voxel's intensity, f, in (4).

The final descriptor consists of four features, rather than the original 27:[4] $(\mu_1; \mu_2; \mu_3; \bar{\mu})$. The small descriptor allows a hashing approach similar to that proposed by Paris[7] to be used, as the 4D hash space can fit into memory. Standard non-local means directly computes w_{xy} for each pair of voxels in the image, an $O(N_x^2 N_p)$ process, where N_x and N_p are the number of voxels and descriptor features respectively, which is expensive. Hashing takes the approach of storing two discretely represented $\mathbf{f}$-images, which are N_p-dimensional: one for the numerator in (1) and one for the denominator. The numerator hash is incremented by $f[\mathbf{x}]$ at location $\mathbf{f}[\mathbf{x}]$ for each voxel $\mathbf{x}$ in the 3D image. Similarly, the denominator is incremented by 1 for each voxel. The two hash spaces are convolved and divided. Finally, linear interpolation is used to extract the smoothed $\tilde{f}$ values from the divided hash-space. This makes for an approach that is $O(N_x)$ and primarily uses linear operations.

4. EXPERIMENTS

Validation against existing algorithms is performed using several images:

- A phantom T1-weighted MRI of a healthy brain from the brain-web data base[8] to which noise is added
- A real T1-weighted MRI acquisition of a knee.
- A real T1-weighted MRI scan of a brain.

The phantom T1-weighted MRI is used for a quantitative comparison between non-local means with the new descriptor and other denoising algorithms. Experiments using both additive Gaussian and Rician noise were performed. Eight levels of noise were added, ranging from 1% to 15% in steps of 2%. The "real" images are used for qualitative comparisons. For the real images, noise is estimated from the variance, σ_g, of a background region.

The noise in MR images has a Rician distribution.[9] This is removed using the standard approach of denoising the square-intensity image, using co-weights from the intensity image, subtracting twice the noise variance and computing the square-root:[10]

$$\tilde{f} = \sqrt{\tilde{f}'^2 - 2\sigma_r^2}, \tag{5}$$

where σ_r is the standard deviation of the Gauss noise independently added to the real and imaginary parts of the complex MR signal. If σ_g is the measured standard deviation of the magnitude MR image, then $\sigma_r = \sqrt{2 - \frac{\pi}{2}}\sigma_g$.[9,10]

Four methods are compared: the proposed method, standard non-local means[4] with an 11^3 local neighbourhood, optimized non-local means,[6] and the bilateral filter.[2,3,7] A neighbourhood is necessary to complete the standard non-local mean tests within approximately one hour per image. All methods were implemented in C++. For rotation invariant non-local means, an h of $\frac{3}{4}\sigma_g$ gave optimal results.

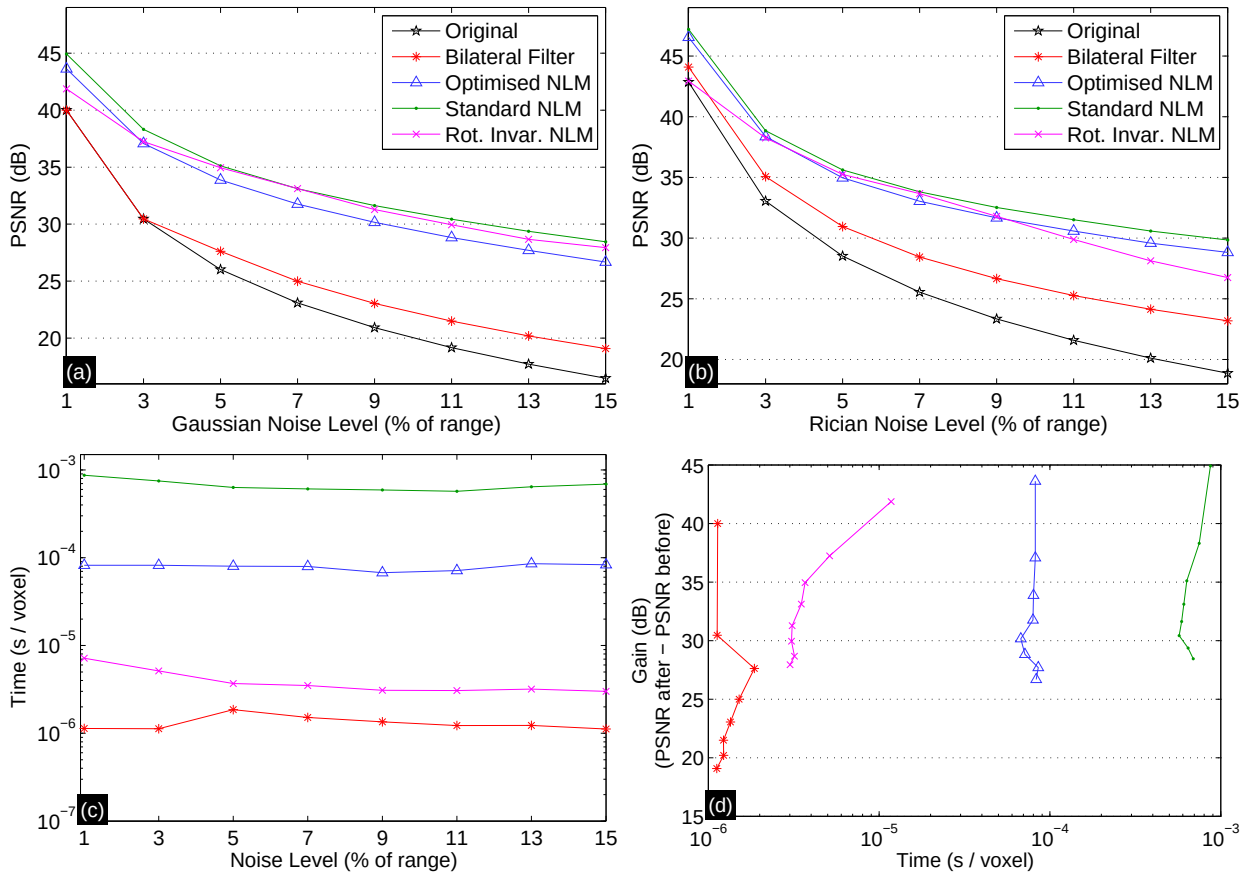


Figure 4. **(a)** PSNR for Gaussian noise. **(b)** PSNR for Rician noise. **(c)** Execution times for denoising methods with Gaussian noise. **(d)** PSNR gain (Difference between PSNR before and denoising) versus time for Gaussian noise. Legends are identical in (a)-(d).

The quantitative denoising results are given in Figure 4 in terms of Peak Signal to Noise Ratio (PSNR) along with the speed of each algorithm. PSNR is measured as follows:

$$\text{PSNR} = 10 \log_{10} \frac{\max_{\mathbf{x} \in X} f[\mathbf{x}]^2}{\frac{1}{|X|} \sum_{\mathbf{x} \in X} (\tilde{I}[\mathbf{x}] - f[\mathbf{x}])^2} \tag{6}$$

For Gaussian noise, optimized non-local means performs worse than standard non-local means due to the plaid artefacts it generates, but optimized non-local means is an order of magnitude faster. The bilateral filter is faster still, but the use of only local information and a single descriptor results inferior performance.

In most cases, rotation invariant non-local means performs nearly as well as standard non-local means despite using a much shorter descriptor consisting of 4 instead of 27 elements. It also outperforms optimized non-local means, because plaid artefacts are not generated. The exception is at the 1% noise level, where the relative performance of rotation invariant non-local means worsens. The drop in performance at low noise levels occurs because the descriptor is an over-simplification of local image structures. Complex image structures resemble noise to short descriptors and are smoothed when they should not be, offsetting the positive results from denoising truly noisy parts of the image. In addition, the resolution of the hash space was restricted to a lower than the optimal value in order to fit into memory, due to the low amount of smoothing.

Similar results are obtained for Rician noise, except the relative performance of rotation invariant worsens from noise levels from 11% and greater. This occurs because changes in intensity range induced by non-local means make the method prone to incorrectly estimated Rician bias.

As shown in Figure 4c optimized non-local means is an order of magnitude faster than standard non-local means. Rotation invariant non-local means is a further order of magnitude faster than optimized non-local means, in some cases nearly reaching the speed of the bilateral filter. The speed improvements arise because a short feature descriptor is used, which allows the algorithm to scale approximately linearly with the number of voxels within an image. The trade-off between speed and denoising performance (or lack thereof) is shown in Figure 4d. The algorithm is slower for low noise noise images, because the hash space is larger, increasing the time taken for convolution.

In the qualitative results shown in Figure 5, artefacts are not apparent in the rotation invariant approach. The difference images show how denoising occurs near strong edges with little blurring of image structures, despite using a short 4 element descriptor.

5. CONCLUSION

A compact rotation invariant descriptor has been proposed for non-local means. The short descriptor allows a hashing approach to be used, allowing significant speed ups: taking 20s to filter a 10 megavoxel 3D image versus 10minutes using optimized non-local means and 2hours using standard non-local means with neighbourhoods.

The rotation invariance of the descriptor means that denoising performance is generally better than optimized non-local means, reaching that of standard non-local means in some cases. In addition, the rotation invariant descriptor is less prone to artefacts, because local patches align with direction of the local edge, rather than the image axes.

REFERENCES

[1] Perona, P. and Malik, J., "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **12**, 629–639 (July 1990).

[2] Smith, S. M. and Brady, J. M., "Susan - a new approach to low level image processing," *International Journal of Computer Vision* **23**, 45–78 (1997).

[3] Tomasi, C. and Manduchi, R., "Bilateral filtering for gray and color images," in [*Proc. Int'l Conf. on Compouter Vision*], 839–846 (1998).

[4] Buades, A., Bartomeu, C., and Morel, J., "A non-local algorithm for image denoising," in [*IEEE Int'l. Conf. on Computer Vision and Pattern Recognition*], **2**, 60–65 (June 2005).

[5] Awate, S. and Whitaker, R., "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in [*Computer Vision and Pattern Recognition*], **2**, 44–51 (June 2005).

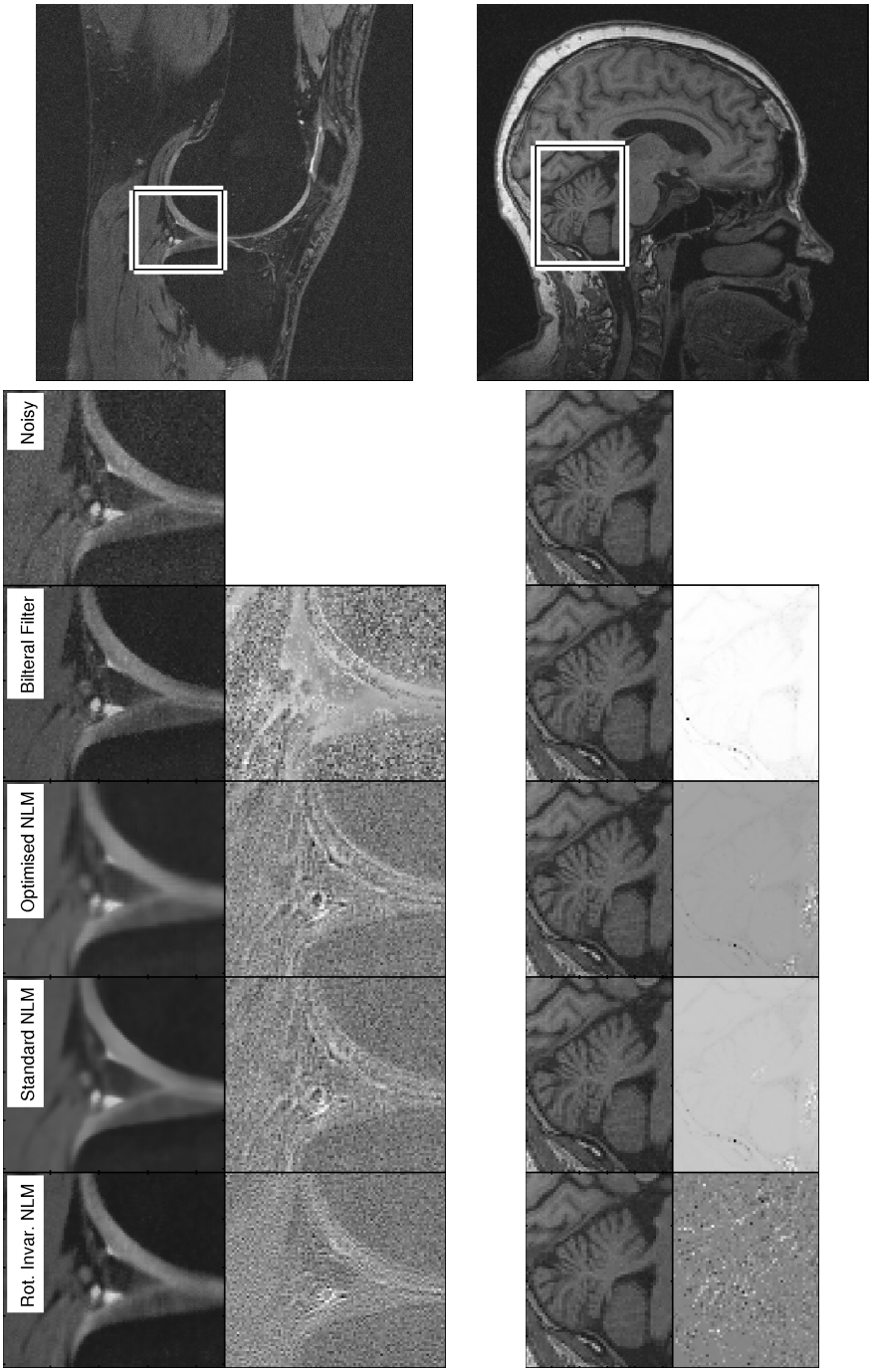
[6] Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C., "An optimized blockwise nonlocal means denoising filter for 3D magnetic resonance images," *IEEE Trans. on Medical Imaging* **27**, 425–441 (April 2008).

[7] Paris, S. and Durand, F., "A fast approximation of the bilateral filter using a signal processing approach," *International Journal of Computer Vision* **81**, 24–52 (2009).

[8] Collins, D., Zijdenbos, A., Kollokian, J., Sled, N., Kabani, C., Holmes, C., and Evans, A., "Design and construction of a realistic digital brain phantom," *IEEE Transactions on Medical Imaging* **17**, 463–468 (March 1998).

[9] Henkelman, R. M., "Measurement of signal intensities in the presence of noise in MR images," *Medical Physics* **12**, 232–233 (March 1985).

[10] Wiest-Daessle, N., Prima, S., Coup, P., Morrissey, S. P., and Barillot, C., "Rician noise removal by non-local means filtering for low signal to noise ratio MRI: Applications to DT-MRI," in [*Proc. of Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*], **2**, 171–179 (September 2008).



(a) Knee MRI (Rician) $\sigma = 7.8\%$ (b) Brain MRI (Rician) $\sigma = 0.6\%$

Figure 5. A qualitative comparison between various denoising methods for 3D images: (a) T1-weighted MRI of a knee, (b) a T1-weighted MRI of a brain. The top row shows one slice from the unfiltered image. Rows below show zoomed regions in the filtered image and the corresponding difference between the filtered and unfiltered images. Difference images are independently scaled to cover image range.