Tina Memo No. 2002-004
Short versions presented at MIUA 2001, MIUA 2002, and ISMRM 2002

# A Novel Method for Non-Parametric Image Subtraction: Identification of Enhancing Lesions in Multiple Sclerosis from MR Images.

P.A. Bromiley, N.A. Thacker and A. Jackson

Last updated
8 / 1 / 2002




Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
University of Manchester, Stopford Building,
Oxford Road, Manchester M13 9PT, U.K.

# A Novel Method for Non-Parametric Image Subtraction: Identification of Enhancing Lesions in Multiple Sclerosis from MR Images

P.A. Bromiley, N.A. Thacker and A. Jackson
Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
University of Manchester, Stopford Building,
Oxford Road, Manchester M13 9PT, U.K.
paul.bromiley@manchester.ac.uk

**Abstract**

Image subtraction is used in many areas of medical image analysis to identify small changes between equivalent pairs of images. Typically only a subset of these differences will be of interest. One example is the identification of enhancing lesions in patients with multiple sclerosis (MS) from MRI scans of the brain. Such lesions can be identified by subtraction of scans taken before and after the injection of GdDTPA contrast agent, which highlights the lesions. However, the presence of the contrast agent also alters the global characteristics of the post-injection scan. Simple image subtraction highlights all differences regardless of their source, making interpretation of the difference image problematic. We introduce a new non-parametric statistical measure which allows a direct probabilistic interpretation of image differences. We expect this to be applicable to a wide range of image formation processes.

## 1 Introduction

Image subtraction is a common tool for the analysis of change in pairs of images, used in a wide range of circumstances. Most researchers will already be familiar with the difficulties of interpreting the resulting difference image [4]. Taking a simple subtraction between two images and identifying regions of change using a threshold is directly equivalent to forming a null hypothesis test statistic, with the assumption of a single distribution for the expected level of change due to uniform noise. In order for the technique to be used successfully great care has to be taken to ensure that the only differences between the two images are due to the biological mechanisms of interest. This may require realignment or pre-processing of the data in order to remove gross changes before a subtraction can be performed. The result can always be used immediately to identify regions of maximal change, but ultimately we would also like to be able to put a quantitative statistical interpretation on the significance of the observed change. The formation of such an interpretation using conventional statistics is generally prevented by the lack of a known statistical model of the expected scene contents or often even of the imaging process. However, most images contain sufficient data that in theory we might extract sensible models of data behaviour from the data itself. This approach has been used widely in recent image registration techniques [8], particularly in medical applications [9]. The technique generally referred to as maximisation of mutual entropy is in fact a boot-strap approach to the construction of a maximum-likelihood statistic [6]. It therefore seems reasonable to attempt to adapt these measures, and equivalent approaches, to the problem of image subtraction in order to investigate the possibility of obtaining quantitative and statistically well-defined measures of difference for arbitrary image pairs.

## 2 Method

### 2.1 Non-Parametric Image Subtraction

Non-parametric image subtraction bootstraps a model of data behaviour from the data itself in the form of a scattergram of a pair of images. Let the image plotted on the abscissa of the scattergram be called the first image, and that plotted on the ordinate the second image. The scattergram for two images of the same scene will show a ridge along the line $y = x$. Global changes to one of the images will result in the movement of this distribution as a whole: for example, a uniform offset to the grey-level values in the second image will move the ridge vertically

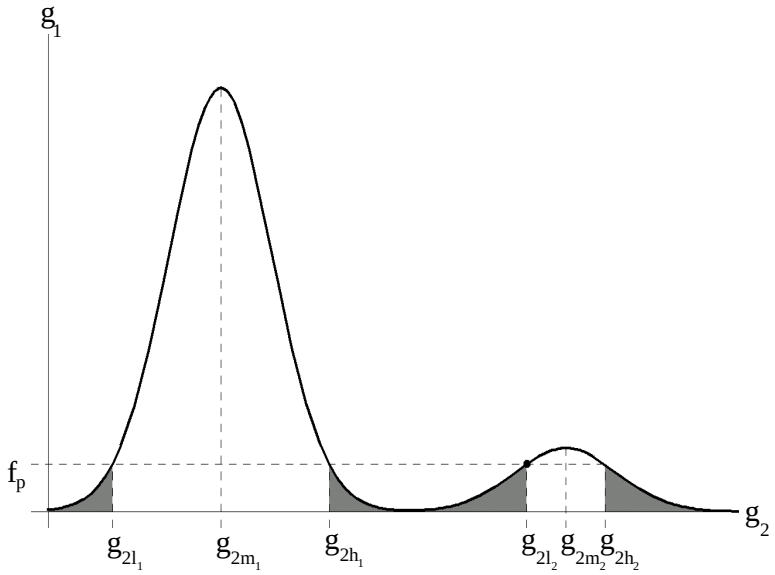



Figure 1: A vertical cut through the scattergram. The integration (the shaded region) is performed over all points lower than the value at the position defined by the original pixels ($f_p$).

in the scattergram. Conversely, localised differences between the two images will produce secondary distributions in the scattergram away from the main distribution. Taking a vertical cut through the scattergram identifies a set of pixels in the first image which all have the same grey-level value $g_1$. The distribution of data along this cut $f(g_2; g_1)$ gives the grey levels $g_2$ occurring at the corresponding pixels in the second image. If the scattergram is normalised along all vertical cuts, then these distributions become the probability distributions for the grey level value in the second image given the grey level value in the first,

$$\frac{f(g_2; g_1)}{\int_{-\infty}^{\infty} f(g_2; g_1) dg_2} = f_n(g_2; g_1). \tag{1}$$

The grey levels of corresponding pairs of pixels from the first and second images define coordinates in the normalised scattergram. An integration can then be performed along the vertical cut in the scattergram passing through that point, summing all of the values smaller than the value at that point, $f_p$. Let the limits for this integration be called $g_{2l}$ and $g_{2h}$. This follows directly from the original definition of a confidence interval, due to Neyman [1]. In addition, the ordering principle implicit in the integration results in the shortest possible confidence interval [3]. The result is the probability $\varepsilon$ of finding a more uncommon pairing of grey levels, given the grey level in the first image $g_1$, than that seen at the original pixel pair,

$$1 - \int_{g_{2l}}^{g_{2h}} f_n(g_2; g_1) dg_2 = 1 - P(g_{2l} < g_{2m} < g_{2h}; g_1) = \varepsilon, \tag{2}$$

where $g_{2m}$ is the mean grey level in the second image at pixels on this cut in the scattergram. The result of this calculation is used as the grey level for the corresponding pixel in the difference image. Since it depends on the mean grey level for the pixels on this cut in the second image, any process which results in a global alteration to the images, such as a change in the level of illumination, will be ignored.

Localised differences will result in secondary peaks along the vertical cut in the scattergram. In that case, shown in Fig. 1, the confidence intervals will span disjoint regions in the scattergram,

$$1 - \int_{g_{2l_1}}^{g_{2h_1}} f_n(g_2; g_1) dg_2 - \int_{g_{2l_2}}^{g_{2h_2}} f_n(g_2; g_1) dg_2 = 1 - P(g_{2l_1} < g_{2m_1} < g_{2h_1} || g_{2l_2} < g_{2m_2} < g_{2h_2}; g_1) = \varepsilon. \tag{3}$$

Since the integration is performed across both the main and secondary distributions, the result is the probability of obtaining a more uncommon pairing of grey levels than that seen at the relevant pixels in the original images based on all of the data from those images, including the data in regions showing localised differences. However, since the secondary peaks due to localised differences will be less significant than the main peak, all of the pixels contained within them will result in low probabilities. The maximum probability that will be assigned to any pixel in one of the secondary peaks will relate directly to the total number of pixels it contains compared to the

total number of pixels in the vertical cut.[1] The grey level values in the difference image therefore relate directly to the frequency of occurrence of the pairing of grey level values seen at the corresponding pixels in the original images. This is exactly the type of measure needed to identify outlying combinations of grey-level values in a fully automatic manner.

The non-parametric image subtraction technique produces a result in terms of a probability, and so both interpretation and further processing of the difference image are simplified and can be performed in a statistically rigorous manner. In addition, the grey level values in the difference image will have a uniform probability distribution, since the integration performed in the scattergram is directly equivalent to a probability integral transform. It follows that the distribution in the difference image is honest i.e. thresholding the difference image at some level $n$ will extract the $100n\%$ of the pixels that showed the most uncommon pairings of grey levels in the original images. The importance of this feature in relation to the work presented here is that knowledge of the expected distribution for the output provides a mechanism for self-test [5].

## 2.2 Renormalisation of Products of Uniform Probability Distributions

A standard technique exists to renormalise the probability distribution of the product of several quantities having flat probability distributions [2]. If n quantities $\omega$, each having a flat probability distribution, are multiplied to produce a product $P$,

$$P = \prod_{i}^{n} \omega_i, \tag{4}$$

then this product can be normalised to produce a new quantity $P'$, which has a flat probability distribution, using

$$P' = P \sum_{j=0}^{n-1} \frac{(-lnP)^j}{j!}. \tag{5}$$

This process is potentially nestable, providing a simple yet statistically principled method for data fusion. Extensive literature searches have failed to reveal an acceptable derivation or proof for this equation, and so a derivation is provided in Appendix A.

The probability renormalisation technique can immediately be applied to form a method for analysing spatial correlation in the difference image produced by non-parametric image subtraction. Low probability pixels in the difference image due to noise will be randomly scattered throughout the images, whereas low probability pixels due to localised differences will exhibit spatial correlation. Identification of this spatial correlation might therefore be expected to lead to a more accurate identification of localised differences. A new image can be constructed by forming the product of the grey-level value of each pixel in the original difference image with the grey-level values of its four nearest neighbours. Each pixel in this new image will therefore be the product of five quantities drawn from uniform probability distributions. The probability renormalisation technique can be applied to this image to generate a second "reflattened" image. The probability renormalisation technique assumes no spatial correlation (see Appendix A) and so only the randomly distributed pixels in the background will reflatten correctly. Low probability pixels situated in localised differences, and therefore surrounded by other low probability pixels, will form very low probability products which fail to renormalise correctly due to the presence of spatial correlation.

It is relatively straightforward to show that, if the probability distribution for the original difference image is composed of a background of random probabilities $P$ together with an additional contribution of lower probability pixels $\delta$ due to localised differences, as shown in Fig 2, then the localised differences must again lie below the other pixels in the product. In the absence of the additional contribution, the product of some number $i$ distributions $P_i$ would be

$$\bar{P} = \prod_{i} P_i \tag{6}$$

In the presence of the additional contribution of low probability pixels, the probability distribution for the background pixels $P'$ can be written as

$$P' = P(1-\delta) + \delta \tag{7}$$

and so the product of $i$ of these distributions becomes

$$\bar{P'} = \prod_{i} P'_i = \prod_{i} (P_i(1-\delta) + \delta) = \prod_{i} P_i(1 - \delta + \frac{\delta}{P_i}) \tag{8}$$

[1] It will also depend on the extent of the overlap, if any, between the main and secondary distributions in the scattergram. Thus differences cease to be detectable when the distributions overlap at the $1\sigma$ points or more.

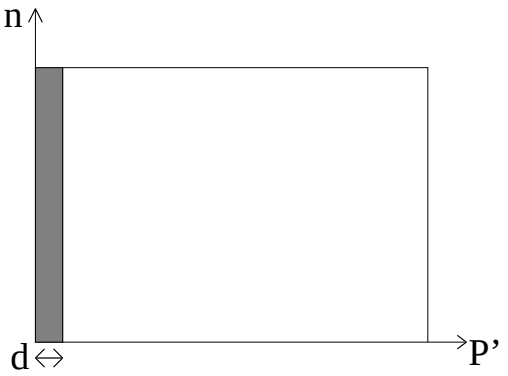



Figure 2: If localised difference pixels in the subtraction result (the shaded region) are assigned lower probabilities than background pixels, then they are guaranteed to still have lower probabilities than background pixels after application of the five-way reflattening technique.

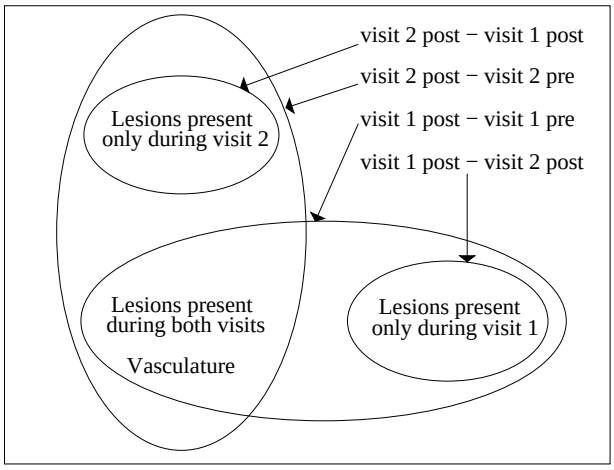



Figure 3: The enhancing features identified by various possible pairings of pre and post contrast agent injection images from two visits (1 and 2).

The term $-\delta + \delta/P_i$ must always be positive (since $P_i \leq 1$), and so

$$\bar{P}' \geq \bar{P} \tag{9}$$

The background pixels are pushed to higher probabilities in the presence of the localised differences, and so the localised difference pixels must lie at lower probabilities. Furthermore, the renormalisation process applies a simple scaling to the values, and so the localised difference pixels must still lie at the lower end of the probability distribution after renormalisation.

The probability distribution for the reflattened image will show a flat distribution for background pixels together with a spike at zero for localised difference pixels. In some circumstances this may lead to a method for volumetric estimation of the localised differences, through thresholding at the level of this spike and counting the number of pixels extracted.

## 2.3 Application to MS Lesion Identification

In order to demonstrate the applicability of the non-parametric image subtraction technique to identifying small but significant changes in medical images, it was applied to the problem of MS lesion identification. Pre- and post-contrast T1 weighted images were obtained in a single patient at each of five visits. Visits were spaced approximately two months apart over a one-year period. Imaging consisted of contiguous 3mm T1 weighted spin echo images (TR/TE, 650/12, FOV $50mm^2$, matrix $256^2$). The contrast agent injection consisted of 1mmols/kg of GdDTPA (Omniscan, Nycomed, Oslo, Norway). All volumes were coregistered to the pre-injection scans from the first visit, using the rigid co-registration software within Tina [7], and then re-sliced using normalised sinc

interpolation with a 5x5x5 kernel to produce the coregistered volumes. Binary masks covering the brain (grey matter and white matter) were produced for these images, using a simple gradient-based segmentation technique. This allowed the subtraction to be performed only on the tissues of interest.

The collection of both pre and post GdDTPA injection scans across five visits provided a total of ten MRI volumes, and subtractions using the new technique were performed on a variety of combinations of these images. Subtractions between the pre- and post-contrast scans from a single visit identified any tissues highlighted by the contrast agent. This included both enhancing lesions and vasculature. Subtraction between post-contrast scans from two visits identified any changes in the enhancing tissues between these visits. The subtraction routine is asymmetric: only enhancing tissues in the image plotted on the ordinate of the scattergram which are not present in the image plotted on the abscissa will be identified. Therefore, subtractions between post-contrast scans from two visits can separately identify either enhancing lesions which have arisen since the last visit, or those that have ceased to be enhancing, depending on which image is used on the ordinate of the scattergram. The possibilities for subtractions between various pairs of tissues are summarised in Fig 3.

# 3 Results

## 3.1 (Post-Contrast)-(Pre-Contrast)

Subtractions between the pre- and post-contrast injection scans for a single visit reveal all of the tissues which are highlighted by the presence of the contrast agent. This includes enhancing MS lesions and vasculature. Fig. 4 shows the pre- and post-contrast scans for a slice featuring two large MS lesions. The result of a simple pixel-by-pixel subtraction of the grey levels is also shown. Fig. 5a shows the result of applying non-parametric image subtraction to these images. The lesions can be seen as dark areas (low probabilities) against a background of random noise with a flat probability distribution. Visual interpretation of the difference image is difficult, but the main advantage of this technique lies in the fact that it provides a direct statistical measure of image differences. For instance, applying a threshold at the 1% level extracts the 1% of the pixels in the image showing the least common pairings of grey levels. The resulting image, Fig. 5b, clearly shows the lesions, together with the vasculature.

The renormalisation technique described above can be applied to the subtraction result to remove the smaller difference regions and any noise. The resulting image, Fig. 5d, can then be thresholded at much lower levels to extract difference regions. Fig. 5d shows the result of thresholding at the 0.001% level. The lesions and larger vascular regions are extracted.

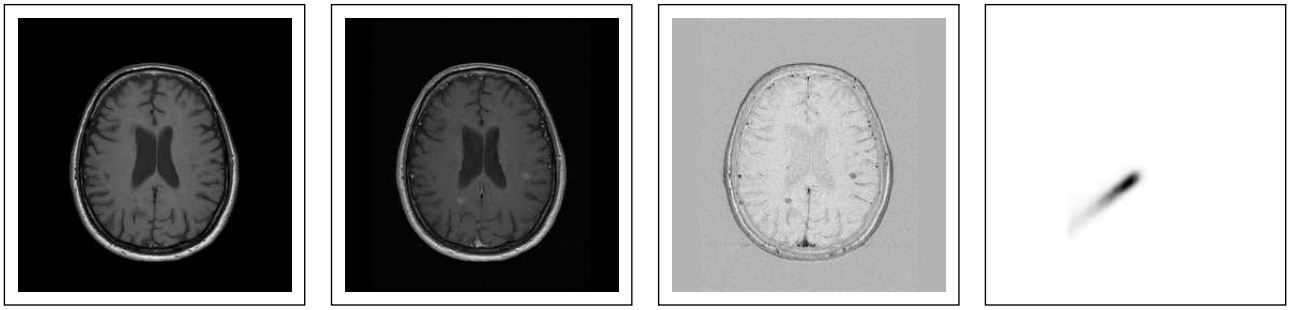
The technique is capable of detecting differences between the original image pair at the level of around $1.5\sigma$, which for these images is approximately 4 grey-levels in a dynamic range of 256 grey levels. Given that the human visual system is only capable of distinguishing 32 grey-levels (without image windowing), the technique should be significantly more sensitive than human inspection of the original images. In that case, the critical question becomes one of evaluation. Several approaches were adopted to demonstrate that the differences detected by the new technique represented genuine physical structures, rather than noise.

Firstly, there are several logical arguments to support the hypothesis that the pixels extracted at the thresholds used represent physical structures. For instance, 0.001% threshold applied to the 5x reflattened image extracts 99 pixels. Given that after application of the mask the image contains around 15000 pixels, we would expect this threshold to extract 0.15 background pixels. It is therefore safe to conclude, at least on statistical grounds, that none of the differences extracted at this level could be due to noise. In addition, the images were inspected by a radiologist, and the pixels extracted at this level were identified as either MS lesions or vasculature.

In order to demonstrate this more clearly, a minimum intensity projection (MIP) of several adjacent slices from around the slice shown in Figs. 4 and 5 was produced, and thresholded the 1% level (the same as applied to the subtraction result). If the pixels extracted using this threshold with the subtraction result represent physical structures, then pixels extracted using the same threshold applied to the MIP should show some spatial correlation, forming either clusters or linear structures depending on the orientation of the underlying structures. Fig 6 shows the resulting image, and it is clear that the majority of the pixels do indeed show spatial correlation: many obvious linear structures are present.

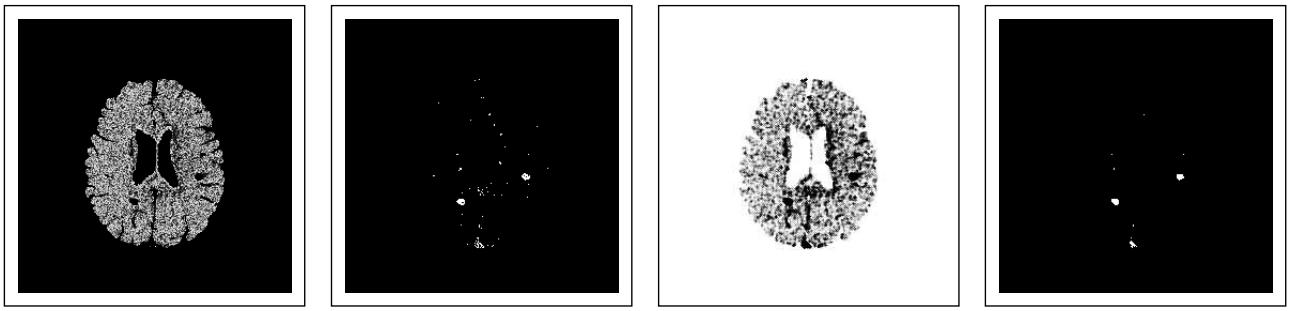
## 3.2 (Post-Contrast)-(Post-Contrast)

A common feature in all of the subtractions performed so far is that the vasculature is extracted together with any MS lesions present. However, if subtractions are performed between post-contrast scans taken during different visits then the vasculature, highlighted in both scans, will not represent a localised difference. In practice, slight



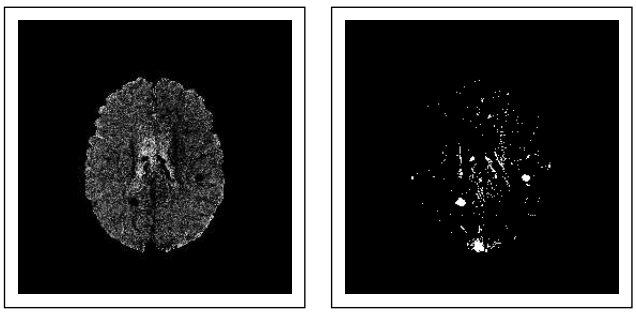
(a) (b) (c) (d)

Figure 4: MS data from the first visit: pre-contrast (a), post-contrast (b), simple subtraction (c) and scattergram (d).



(a) (b) (c) (d)

Figure 5: Subtraction results for pre- and post-contrast scans from the first visit: NPI subtraction (a), NPI subtraction threshold at 1% (b), 5x reflattening (c) and 5x reflattening threshold at 0.001% (d).



(a) (b)

Figure 6: Minimum intensity projection of the first visit pre- minus post-contrast subtraction results (a), and 1% thresholding of the MIP (b).

changes due, for example, to atrophy will result in the vasculature being identified to some extent, but the amount identified should be greatly reduced compared to the subtraction results shown in the previous sections. In addition, the results will show only the changes in the MS lesions, rather than identifying all of the enhancing lesions. Since the primary use of MRI scanning with MS patients is for monitoring the progress of the disease, either in relation to patient management or to therapeutic trials, this may be a more useful way in which to process the data. It is also important to remember that the subtraction technique is non-symmetrical: only new lesions (or lesion growth) in the image used on the vertical axis of the scattergram will be identified. Therefore, depending on the order of the subtraction it is possible to identify either new lesions that have appeared or lesions that have ceased to be enhancing during the interval between the scans.

To illustrate such subtractions, Fig. 7 shows two post-contrast scans from the same patient, taken approximately two months apart. Two enhancing lesions are present in the first scan and have ceased to be enhancing at the time of the second scan, and the second scan contains one new enhancing lesion. The result of a simple subtraction is also shown: the new lesion appears dark and the old lesions bright. Note that changes are identified in the

sagittal sinus at similar levels as the changes due to the lesions: this demonstrates that significant changes occur in the vasculature between visits, and so indicates that complete elimination of the vasculature from the NP image subtraction result may be impossible without resorting to a more complex technique, such as using an angiogram to identify vascular regions. Fig. 8 shows the result of an NPI subtraction between these scans with the later scan plotted on the vertical axis of the scattergram, so that only new lesions are identified. As stated above, there is some small amount of contamination by vasculature, but this disappears almost completely on application of the spatial correlation analysis technique. Fig. 9 shows the results of subtracting the images in the other order: in these images, only the lesions present in the first scan, which have ceased to be enhancing in the later scan, are identified. Again there is some small contamination with vasculature, but in general these images show clean detections of the development of MS lesions with negligible identification of background features.

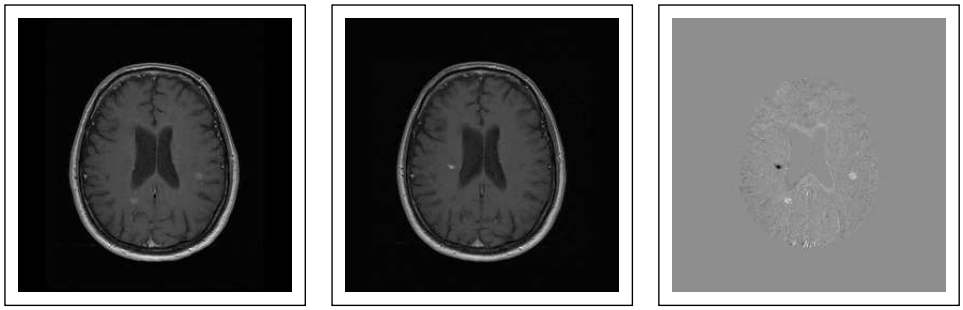


Figure 7: Slice 27 post contrast for the first (a) and second (b) visits, and simple image subtraction (c).

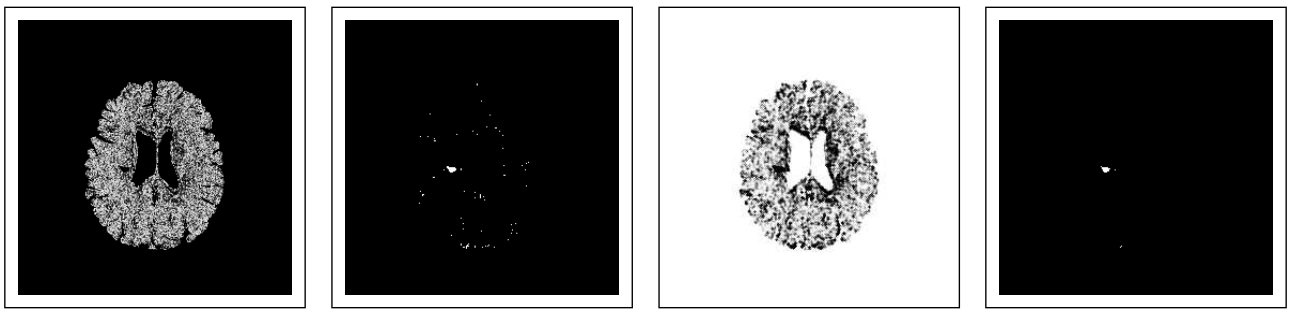


Figure 8: NP image subtraction (a), thresholding at 1% (b), 5x reflattening (c), and thresholding at 0.001% (d) for a subtraction of the post-contrast scan from the first visit from that from the second visit.

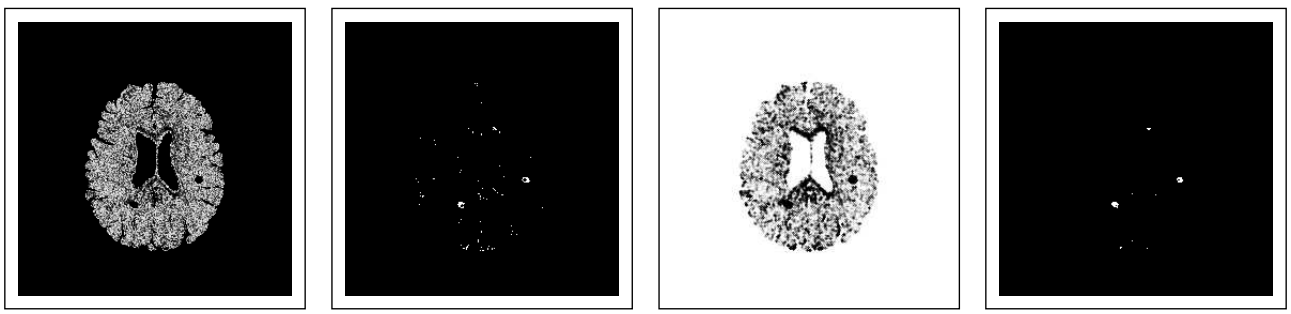


Figure 9: NP image subtraction (a), thresholding at 1% (b), 5x reflattening (c), and thresholding at 0.001% (d) for a subtraction of the post-contrast scan from the second visit from that from the first visit.

# 4 Conclusion

Identification of MS lesions in MRI scans of the brain is an important issue in relation both to tracking the progression of the (chiefly pre-clinical) disease, and to therapeutic trials. The lesions can be highlighted by injection of GdDTPA contrast agent which concentrates at the lesion sites and produces a strong signal in the scanner due to the presence of the paramagnetic Gd ion. However, the most widely used method for identifying lesions in these scans remains visual inspection.

This paper has presented a new technique for the subtraction of a pre-contrast scan from the post contrast-scan in a statistically rigorous manner, identifying significant changes between the scans by comparison to a model of data behaviour extracted from the data itself in the form of a scattergram. The technique is based on the standard formulation of a confidence interval and quotes a result in terms of a probability, and so is more statistically rigorous than a simple subtraction of grey levels, making the results more amenable to further analysis. In addition the probability distribution for the difference images is uniform (flat) and therefore honest. A standard technique exists for renormalising the product of several quantities, each having a flat probability distribution, and this has been exploited here to produce a spatial correlation analysis technique.

The new subtraction technique is superior in applications that require small but significant localised changes to be detected against a background of global changes between an image pair, and the identification of MS lesions in MRI scans is an example of such an application. We have shown that the new technique is capable of detecting either the MS lesions or changes in MS lesions over time, with a greater sensitivity than visual inspection.

This paper has illustrated the application of the basic non-parametric image subtraction technique to MS lesion detection, but several extensions are immediately apparent. Firstly, MS lesions occur only in white matter. Some segmentation technique could be used to identify the white matter, allowing other tissues to be ignored and thus refining identification of the lesions. Similarly, an angiogram could be used to identify and thus ignore the vasculature. Once the lesions positions have been identified an estimate of partial volumes in the pixels around the edges of the lesions could easily be obtained, since the grey-level values of these pixels would depend on a linear combination of the mean grey levels of white matter and MS lesion. This would allow accurate estimates of lesion volumes to be obtained. However, in order to assess the utility of the technique in clinical applications, further clinical work would be required to demonstrate a relationship between some parameter of the lesion distribution (e.g. position, volume or number), and some indicator of disease severity or progression.

# Acknowledgments

The authors would like to acknowledge the support of the DTI Medilink Scheme grant no. P169 (Smart Inactivity Monitor using Array Based Detectors (SIMBAD)) in the development of the technique described here, and the support of the EPSRC and the MRC (IRC: From Medical Images and Signals to Clinical Information) in obtaining the data used.

# References

[1] J. Neyman, "X-Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability.", *Phil. Trans. Royal Soc. London* **A236**, pp. 333-380, 1937.

[2] ALEPH Collaboration. "A Precise Measurement of $\Gamma_{Z\to b\bar{b}}/\Gamma_{Z\to hadrons}$." *Physics Letters* **B313**, pp. 535-548, 1993.

[3] G.J. Feldman and R.D. Cousins, "A Unified Approach to the Classical Statistical Analysis of Small Signals", *Phys. Rev.* **D57**, pp. 3873, 1998.

[4] J.V. Hajnal, I.R. Young, & G.M. Bydder. "Contrast Mechanisms, Functional MRI of the Brain." In *Advanced MR Imaging Techniques*, pp. 195-207. Martin Dunitz Ltd, London, 1997.

[5] I. Poole, "Optimal Probabilistic Relaxation Labeling." *Proc. British Machine Vision Conference 1990*, BMVA, 1990.

[6] A. Roche, G. Malandain, N. Ayache, et al. "Towards a Better Comprehension of Similarity Measures Used in Medical Image Registration." In *Proc. MICCAI 1999*, pp 555-565. Cambridge, 1999.

[7] N.A. Thacker, A. Lacey, E. Vokurka, et al. "TINA an Image Analysis and Computer Vision Application for Medical Imaging Research." In *Proceedings of the European Congress on Radiology*, 24-003, pp. s566. Vienna, 1999.

[8] P. Viola. *Alignment by Maximisation of Mutual Information.* PhD. Thesis, Artificial Intelligence Laboratory, MIT, 1995.

[9] J. West, J. Fitzpatrick, M. Wang, et al. "Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques." In *Journal of Computer Assisted Tomography* **21**(4), pp. 554-566, 1997.

# A Probability Renormalisation

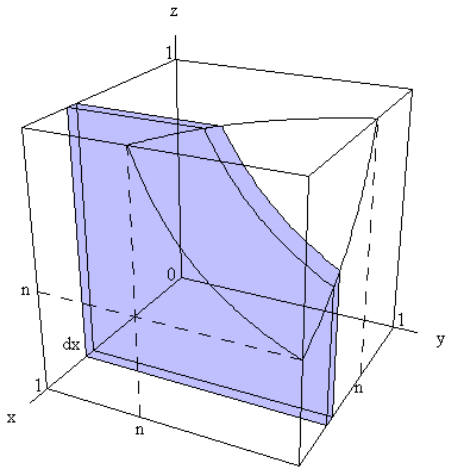



Figure 10: The sample space for the probability renormalisation in 3D, showing the element of integration (the shaded region) used to relate this to the 2D problem. The contour of constant probability is shown by the curved surface in the upper corner of the unit cube.

Given n quantities each having a uniform probability distribution $p_{i=1,n}$, the product $p = \prod_{i=1}^{n} p_i$ can be renormalised to have a uniform probability distribution $F_n(p)$ using

$$F_n(p) = p \sum_{i=0}^{n-1} \frac{(-\ln p)^i}{i!} = p + p \sum_{i=1}^{n-1} \frac{(-\ln p)^i}{i!} \tag{10}$$

The quantities $p_i$ can be plotted on the axes of an n dimensional sample space, bounded by the unit hypercube. Since they are uniform, and assuming no spatial correlation, the sample space will be uniformly populated. Therfore, the transformation to $F_n(p)$ such that this quantity has a uniform probability distribution can be achieved using the probability integral transform, replacing any point in the sample space p with the integral of the volume under the contour of constant p passing through this point, which obeys $\prod_{i=1}^{n} p_i = p = constant$. This can be expressed in terms of the volume of a hyper-region of one lower dimension by integrating over one dimension (let this be called x)

$$F_n(p) = p + \int_p^1 F_{n-1}(\frac{p}{x})dx \tag{11}$$

This is equivalent to dividing the integration into two regions using a plane perpendicular to the x axis which intersects the axis at $x = p$. Fig. 10 shows the element of integration that would be used in the 3D case, to relate the volume of the unit cube under the contour of constant probability to the 2D case.

Now, in the simplest case of $n = 1$, clearly $F_n(p) = p$, as no renormalisation is required. The solution for higher dimensions can then be derived by iterative application of Equation 13. This involves integration of terms in $(p/x)[-\ln(p/x)]^n$ which enter in the n=3 and higher cases. This integration can be performed using a simple substitution $x = pu$, $dx = pdu$

$$\int_p^1 (\frac{p}{x})[-\ln(\frac{p}{x})]^n dx = p \int_1^{1/p} (\frac{1}{u})[\ln u]^n du = p[\frac{1}{n+1}[\ln u]^{n+1}]^{n+1} = \frac{p}{n+1}[-\ln p]^{n+1} \tag{12}$$

Iterative application of Equation 13 therefore produces the series

$$F_n(p) = p - p\ln p + p\frac{(\ln p)^2}{2} - p\frac{(\ln p)^3}{6} + p\frac{(\ln p)^4}{24}.... \tag{13}$$

which can be written as

$$F_n(p) = p \sum_{i=0}^{n-1} \frac{(-\ln p)^i}{i!}. \tag{14}$$