Non-Parametric Image Subtraction

P. A. Bromiley, N.A.Thacker and P. Courtney
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
paul.bromiley@man.ac.uk

July 12, 2001

Abstract

This document describes the underlying theory behind the technique of non-parametric image subtraction. The context of this work, details of the implementation, and examples of the application of the technique to some relevant problems are given in [1] [2] and [3].

1 Introduction

Image subtraction is a standard tool in many areas of machine vision. At the simplest level, subtraction of two images can be used to idenitify the differences between them so, for example, image subtraction could be applied to a video sequence to identify motion of objects in the scene. In practice, there are two main probelms with simple pixel by pixel subtraction. Firstly, it identifies all differences between a pair of images regardless of their source. Typically only a subset of these due to a specific physical mechanism or mechanisms will be of interest. This often takes the form of significant localised differences against a background of unimportant global differences e.g. an image sequence may show motion of features but also show global illumination changes. In addition, the pixel intensities in the resultant difference image are arbitrary measures of difference in units of grey levels, and so have no objective meaning. The non-parametric image subtraction technique addresses these problems by using the image data itself to generate a model of the data distributions in the original images. This allows the technique to take account of the global differences between the two images, and identify only the more significant, localised differences. Furthermore, it returns an answer in terms of a well-defined statistical quantity, a probability, simplifying both interpretation and further processing of the difference image.

2 Method

The first step in non-parametric image subtraction is to produce a scattergram of the two images. Let the image plotted on the abscissa 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, an increase in the level of illumination between the first and second image will move the ridge vertically in the scattergram. Conversely, localised differences between the two images will produce secondary distributions in the scattergram away from the main distribution, as shown in Figure 1. 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 occuring at the corresponding pixels in the second image. If the scattergram is normalised along all vertical cuts, then these distributions become the probability distribution for the grey level value in the second image given the grey level value in the first,

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

The next step is to take corresponding pairs of pixels from the original first and second images, and use their grey levels to find their coordinates in the normalised scattergram. An integration is then 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, as shown in Figure 2. 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 [4]. In addition, the ordering priciple implicit in the integration results in the shortest possible confidence interval [6]. The result is the probability ε 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,

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

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, as shown in Figure 3. In that case, the confidence intervals will span disjoint regions in the scattergram

$$1 - \int_{g_{2l_1}}^{g_{2h_1}} f_n(g_2; g_1) dg_2 - \int_{g_{2l_2}}^{g_{2h_2}} f_n(g_2; g_1) dg_2$$

$$= 1 - P(g_{2l_1} < g_{2m_1} < g_{2h_1} || g_{2l_2} < g_{2m_2} < g_{2h_2}; g_1) = \varepsilon.$$

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. It will also depend on the extent of the overlap, if any, between the main and secondary distributions in the scattergram, although this factor is not further explored here. 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 distribution of grey level values in the difference image will be flat. For example, 30% of the pixels in the vertical cut lie in the outermost 30% of the distribution, and so will be assigned the value 0.3 or less: 20% of the pixels will be assigned the value 0.2 or less, and so on. 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. In addition, a standard technique exsists to renormalise the probability distribution of the product of several quantites having flat probability distributions [5]. If n quantities ω, each having a flat probability distribution, are multiplied to produce a product P,

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

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

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

This process is potentially nestable, providing a simple yet statistically principled method for data fusion.

3 Extensions to the Method

The main advantages of the technique as described so far are the ability to measure differences in terms of well-defined units (a probability), taking into account and ignoring global differences between the images, and producing a difference image with a flat probability distribution, simplifying further processing. However, several variations on the method immediately suggest themselves. If the locations of the differences between the two images are known, the scattergram could be generated from regions of the image not featuring the differences. The integration in the scattergram would then be performed only over the mean distribution, as shown in Figure 4,

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

Pixels in secondary distributions resulting from localised differences would produce very low probabilities that related to the distance between the mean distribution and the secondary distribution. Each pixel in the resulting difference image would give the probability of obtaining a more uncommon pairing of grey levels (than that seen at the corresponding pixel pair in the original images) due to noise on the mean distribution. This is the kind of difference measure that we might intuitively desire, since the result would relate to the difference in intensity between the mean distribution pixels and the localised difference pixels in the original images. The bootstrapped technique would retain the advantages of ignoring global differences and producing an output in terms of a probability. However, since the pixels in the secondary distributions would not be used in generating the scattergram, the distribution in the difference image would be flat only for pixels drawn from the mean distribution. Localised differences would result in a spike at low probability. If the difference image were then thresholded at some level n higher than this peak, it would extract all of the pixels corresponding to localised differences together with $100n\%$ of the pixels from the remainder of the image. Volumetric analysis could therefore be performed on the differences.

The identification of the non-parametric image subtraction technique with Neyman's construction of confidence intervals also suggests extensions to the method, particularly the use of one-sided confidence intervals. If the secondary distributions are known to lie on one side of the mean distrbution in the scattergram, then the construction of a one-sided confidence interval, extending the integration to infinity on the side nearest the secondary distributions, as shown in Figure 5,

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

allows the incluison of additional data from the mean distibution. Localised differences are therefore assigned lower probabilities than with the double-sided technique, making their identification easier.

In addition the renormalisation technique and flat probability distribution of the difference image lead to an immediate extension to the basic method. If a threshold is applied to the difference image produced by the basic technique, then some predictable number of pixels will be extracted e.g. a threshold of 0.1 will extract 10% of the pixels in the image. Some proportion of these will be localised difference pixels, and some proportion background pixels. The proportions of each will be determined by the relative numbers of pixels in the localised differences and in the remainder of the image, as well as by the exent of any overlap between the two distributions on the relevant vertical cuts in the scattergram. The important difference between the two populations will be that the background pixels will be randomly distributed over the image, whereas the localised difference pixels will be spatially correlated. This implies that some method for analysing the spatial correlation of low-probability pixels in the difference image could improve the identification of the localised differences. A simple spatial correlation measure has been investigated as follows. A new image can be produced where the grey level of each pixel is the product of the grey level of the corresponding pixel in the difference image with the grey levels of the four pixels arranged in a cross around it i.e. the pixels above, below, to the left and to the right. This is equivalent to forming the product of five images each having a flat probability distribution, so the distribution of the resulting image can be renormalised (reflattened) using the renormalisation equation given above. However, the renormalisation equation is only valid if there is no spatial correlation, and so the renormalised image will have a flat probability distribution only for pixels drawn from the mean distribution. Localised difference pixels, being spatially correlated and thus surrounded by low probability pixels, will form very low probability products which will not re-flatten correctly. Therefore, the probability distribution of the image showing the renormalised five-pixel product will feature a flat distribution for non-difference pixels together with a spike close to zero containing the localised difference pixels. The localised difference pixels can then be extracted at lower thresholds than would be used for the original difference image, leading to less contamination by background pixels. In addition, the number of background pixels below the threshold will be known, again offering the possibility of volumetric analysis of the localised differences.

4 Conclusion

This report outlines the statistical theory underlying the technique of non-parametric image subtraction. The main advantages of the technique as opposed to a simple pixel-by-pixel image subtraction have been identified as:

- The technique ignores global differences between the images, and identifies only localised differences.
- The result is given in terms of a probability, a well defined statistical quantity, simplifying interpretation and further processing of the difference

image.

- The difference image has a flat, and therefore honest, probability distribution, providing a simple route to data fusion and other further processing.

Several possible extensions to the technique have been identified, including bootstrapping the technique from regions of the images not featuring localised differences, in order to allow volumetric analysis of the differences; the construction of one-sided rather than two-sided confidence intervals; and a technique for spatial correlation analysis in the difference image using the renormalisation equation, which also provides a route to volumetric analysis.

We are currently investigating the extensions to the basic method identified in this report, and are aquiring medical datasets (consisting of MRI scans of MS patients with enhancing lesions taken before and after the injection of GdDTPA contrast agent) for more extensive testing on realistic images. In addition, the source code for non-parametric image subtraction has been integrated into TINA, our research software environment, which is freely available for download from our website:

http://www.niac.man.ac.uk/Tina

References

[1] P.A. Bromiley, N.A. Thacker, & P. Courtney. "Non-parametric Image Subtraction using Grey Level Scattergrams." In *Proceedings of the British Machine Vision Conference 2000*, pp. 795-804. BMVA, 2000.

[2] P.A. Bromiley, N.A. Thacker & P. Courtney "Non-parametric Image Subtraction using Grey Level Scattergrams." *Image and Computer Vision*, BMVC 2000 Special Edition, In press.

[3] P.A. Bromiley, N.A. Thacker, & P. Courtney. "Non-parametric Image Subtraction for MRI." In *Proceedings of Medical Image Understanding and Analysis 2001*, In press.

[4] 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.

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

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

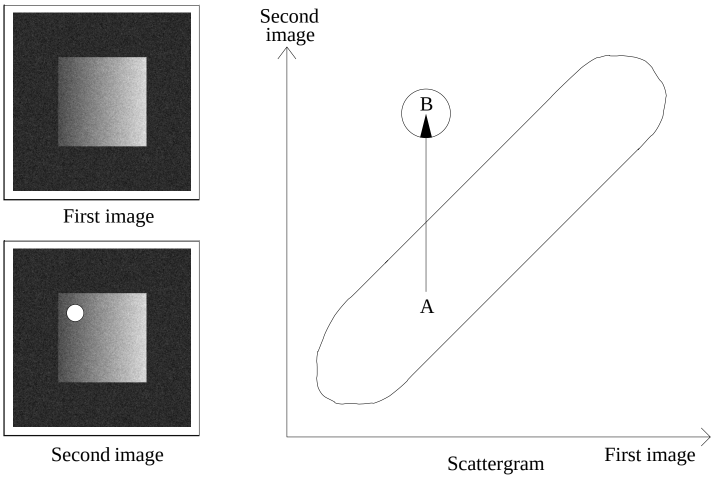

Figure 1: If two images feature some localised difference, this will generate a secondary distribution B. Global differences will result in overall movement of the mean distribution A.

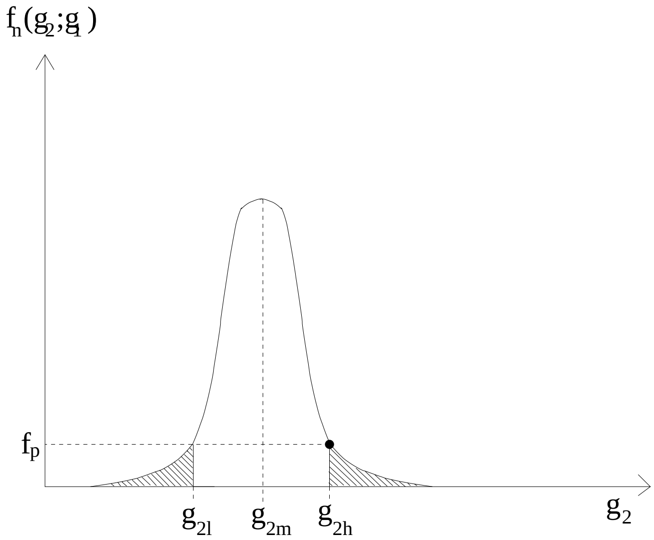

Figure 2: A vertical cut through the normalised scattergram for two images without localised differences shows a unimodal distribution due to noise. For any pair of corresponding pixels from the original images (the black point), the integration (the shaded region) is performed across all values smaller than f_p, the value at the point defined by the original image pixels.

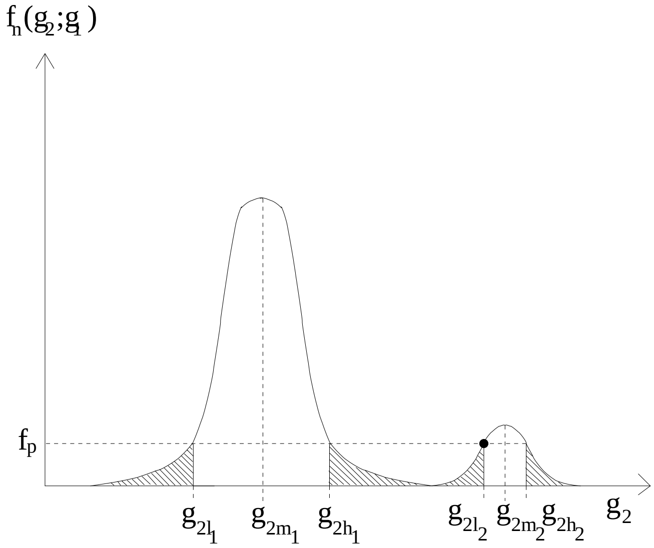

Figure 3: A vertical cut through the normalised scattergram for two images with local differences will have a multimodal distribution, so the confidence intervals will be disjoint.

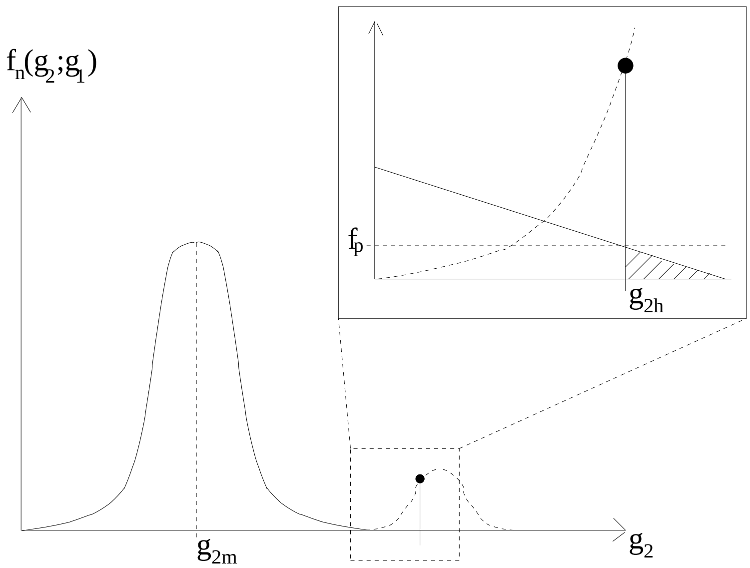

Figure 4: If the scattergram is bootstrapped from regions of the image without localised differences, then the probability is computed from the mean distribution alone, so points in regions showing localised differences result in lower probabilities.

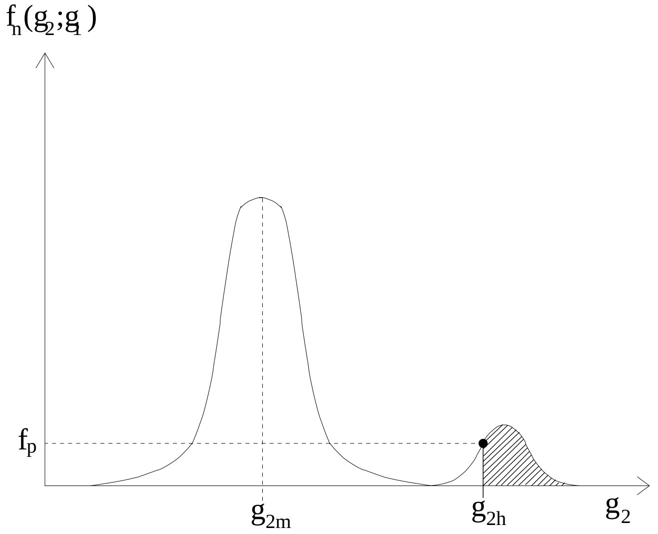

Figure 5: The identification of the technique with Neyman's construction for confidence intervals suggests the use of one-sided confidence intervals as an extension to the basic method. The integration (the shaded region) is then performed from infinity to the point defined by the pixels from the original images (the black point).