Automatic registration of temporal image pairs for digital subtraction angiography

G. S. Cox
G. de Jager (Member IEEE)

Department of Electrical Engineering
University of Cape Town, South Africa

ABSTRACT

Temporal Digital Subtraction Angiography (DSA) is used to visualize blood vessels in x-ray images. A DSA image pair consists of the mask image, which is a digitized x-ray taken before a contrast medium is injected into the bloodstream, and the live image, which is taken once the contrast medium has traversed the circulatory system and reached the blood vessels of interest. The mask image is then subtracted from the live image and ideally only the contrast enhanced blood vessels should remain. DSA has two main limitations. Firstly, gross patient motion and physiological events occur in the time that elapses between x-rays. Secondly, there are local and global differences in the mean gray-level at corresponding points in the live and mask images, excluding the variations introduced by the contrast media. To solve the motion problem, we take the approach of matching regions around control points in the live image in a search area around the approximately corresponding points in the mask image. In this way a motion vector field that describes the spatial offset to the best match position in the mask image (with sub-pixel accuracy) is constructed. The problem of mean gray-level disparity between the live and mask images is to a large extent overcome by the use of a match measure that is invariant to overall additive gray-level differences. Incorrect mismatches caused by the contrast media are avoided by using multiple sub-templates in the matching process. The sub-template method also allows the estimation of mean gray-level disparity between the mask and live images. The smoothed motion vector field and mean gray-level disparity estimates are used to perform an improved subtraction of the mask from the live image with a reduction in the artifacts that are a result of normal subtraction. Efficient best match search techniques are used to reduce the computational cost of the algorithm, at the expense of some difference image quality. Results are provided for simulated and actual DSA image pairs.

1 INTRODUCTION

Angiography is not always possible using x-ray images because the contrast between blood vessels and surrounding tissue is small. This problem can be solved by injecting a contrast medium directly into the blood vessel of interest. A less invasive approach is to inject the contrast medium into the bloodstream non-selectively, and wait for it to reach the blood vessel of interest via the circulatory system. However, in this case the contrast medium is diluted and the contrast is still too small for adequate visualization. In temporal Digital Subtraction Angiography (DSA) the increased x-ray attenuation of the blood vessels due to the contrast medium is used to isolate the blood vessels by subtracting an x-ray image of the same area before the contrast medium was administered. X-rays are taken before and after the contrast media is injected into the bloodstream. The *before* (or *mask*) image is then subtracted from the *after* (or *live*) image. Ideally, the resulting difference image should show only blood vessels.

In practice two problems limit the success of DSA using direct temporal subtraction. Firstly, the live image is spatially distorted relative to the mask image. This is due to gross patient motion and physiological events that occur in the time that elapses between x-rays. Secondly, there are global and local disparities between the mean gray-levels of the live and mask images. To some extent the quality of the difference image can be improved by averaging, or using multiple mask images and selecting the best mask for subtraction. In many cases, however, these methods are inadequate. Techniques that have been proposed to overcome the motion problem generally involve using template matching to determine registration parameters for corresponding regions in the two images. A global set of registration parameters, or registration parameters for zones in the mask image[1] can correct for gross patient motion, but are not accurate enough to deal with the local spatial distortions caused by physiological events. The successful correction of local spatial distortions has been achieved[2,3] by obtaining registration parameters for regions around control point pixels and interpolating these parameters for the rest of the pixels. Morishita and Yokohama[2] used this approach and proposed preprocessing of the live and mask images before registration. Tran and Sklansky[3] introduced the *exclusion template* for excluding the blood vessels from the matching process, and allowed for rotation of the template in addition to simple translation. Mean gray-level disparity has been dealt with by using global correction parameters for the images[1,3] and averaging in small corresponding regions of the live and mask images to find local correction parameters.[3]

This paper presents an algorithm for DSA that has the same basic framework as the techniques mentioned above. Different methods are used in key areas in order to improve performance. Registration with an offset invariant measure of match cancels the effect of additive mean gray-level disparity between the live and mask images. A *multiple sub-template* approach is introduced in order to ignore the contribution of contrast enhanced blood vessels during the registration. This allows accurate registration with automatic exclusion of blood vessels that mismatch due to the additional x-ray attenuation of the contrast medium in the live image. The multiple sub-template approach also allows the estimation of additive mean gray-level disparity.

2 ALGORITHM

The algorithm determines the components of a motion vector for each of the control point pixels, which lie on the intersections of a uniform grid over the live image. This vector indicates the displacement to the corresponding position in the mask image. The following procedure is used to find the motion vector for the control point with coordinates (x_l, y_l). The coordinates of the same position in the mask image are (x_m, y_m). The search radius R is defined as the maximum expected displacement of a pixel.

1. An $N \times N$ sub-image of the live image, with the control point at its center, is extracted as the template.
2. Template matching is performed in the area bounded by a circle of radius R and centered on the coordinates (x_m, y_m) in the mask image. A translation increment of one pixel is used. In addition the template can be rotated within specified limits at each position, but this adds considerably to the computational cost of the algorithm.
3. Interpolation in the match measure surface is used to determine the coordinates of the best match, (x_b, y_b), to sub-pixel accuracy.
4. The coordinates (x_b, y_b) are used to calculate the components of the motion vector.

The following procedure is then used to produce the final difference image:

1. The motion vector field is smoothed by removing vectors that are not consistent with their neighbors.
2. Motion vectors are found for pixels not on the grid using bilinear interpolation of the four nearest control points.
3. The motion vectors are used to perform a corrected subtraction of the mask from the live image.
4. The difference image is corrected for mean gray-level disparity effects between the live and mask images.
5. A *weighted averaging* smoothing method[4] is used to remove noise from the difference image.

3 MEASURE OF MATCH

Accurate registration requires that the measure of match is maximized for corresponding areas of the live and mask images. Corresponding pixels, however, may have different gray-levels due to local and global distortions of the mean gray-level. Global effects caused by the imaging system include additive variations caused by fluctuations in detector gain. Local additive effects are caused by diffusion of the contrast media. Previous DSA registration techniques have used the Stochastic Sign Change (SSC) and Deterministic Sign Change (DSC),[1] and the Sum of the Absolute Value of the Differences (SAVD)[3] measures of match. These measures are mean gray-level dependent and additive offsets will cause a mismatch where pixels actually correspond. The mean gray-level can be estimated before registration to remove differences between the live and mask images, but this is only successful where spatial distortion is small. The presence of the contrast media in the live image also complicates the estimation procedure. For this reason we use a measure of match that is invariant to additive offsets in the mean gray-level: the *variance* of the pixel by pixel differences between the template and the sub-image to which it is applied.

Consider the determination of a measure of match between the template and a particular template-sized sub-image. Let $\tilde{f}(x, y)$ and $\tilde{g}(x, y)$ give the pixel intensity as a function of position with the origin centered in the template and sub-image respectively. Then let

$$f_{ij} = \tilde{f}(in, jn)$$

and

$$g_{ij} = \tilde{g}(in, jn)$$

where n is is the distance between points on the discretization grid.

Notationally we define

$$\sum_A \equiv \sum_{i=-\frac{N}{2}}^{\frac{N}{2}} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}}$$

where the template and sub-image are $N \times N$ pixels.

The *variance of differences* (VOD) measure of match is given by:

$$M_{vod} = \frac{1}{1+\sigma^2} \tag{1}$$

where σ^2, the variance of the differences, is

$$\sigma^2 = \frac{\sum_A \left[(f_{ij} - g_{ij}) - \mu_d\right]^2}{n-1}. \tag{2}$$

Here μ_d is the mean of the differences and $n = N^2$ is the number of pixels in the $N \times N$ template. For a perfect match $M_{vod} = 1$ and for a mismatch $M_{vod} \rightarrow 0$.

The more efficient algebraic equivalent to equation 2,

$$\sigma^2 = \frac{n \cdot \sum_A (f_{ij} - g_{ij})^2 - (\sum_A (f_{ij} - g_{ij}))^2}{n(n-1)}, \tag{3}$$

is used for computing the variance.

4 SUB-PIXEL PRECISION

The registration initially determines a motion vector with x and y components that are an integral number of pixels. The need for greater precision requires that the coordinates of the corresponding peak in the match measure surface be found to sub-pixel accuracy. This can be achieved by translating the template by sub-pixel increments in the local region around the initial best match position.[3] This method is computationally expensive. We take an alternative approach by interpolating in the match measure surface to estimate the position of the peak to sub-pixel precision. Separate horizontal and vertical quadratic interpolation is used to refine the estimates of the x and y components respectively. Given the discrete match measure surface $M_{fg}(x, y)$ and the best match position (x_0, y_0), the estimated continuous coordinates $(\tilde{x}, \tilde{y})$ of the match peak are calculated as

$$\begin{aligned}
\tilde{x} &= x_0 - \frac{1}{2} + \frac{M_{fg}(x_0, y_0) - M_{fg}(x_0 - 1, y_0)}{2M_{fg}(x_0, y_0) - M_{fg}(x_0 - 1, y_0) - M_{fg}(x_0 + 1, y_0)} \\
\tilde{y} &= y_0 - \frac{1}{2} + \frac{M_{fg}(x_0, y_0) - M_{fg}(x_0, y_0 - 1)}{2M_{fg}(x_0, y_0) - M_{fg}(x_0, y_0 - 1) - M_{fg}(x_0, y_0 + 1)}.
\end{aligned}$$

This is illustrated in figure 1.

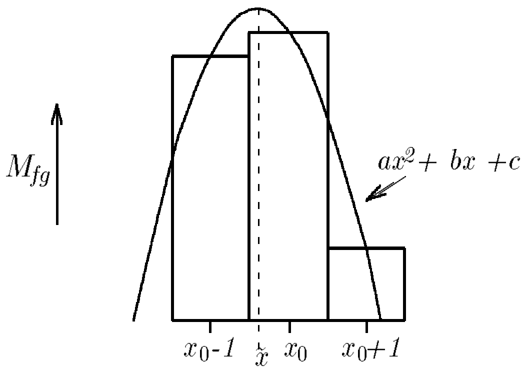


Figure 1: Quadratic interpolation for the best match x coordinate

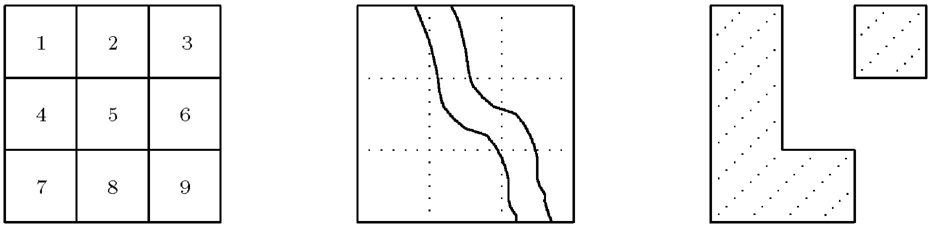
5 MULTIPLE SUB-TEMPLATES

Due to the contrast media, the blood vessels will have a higher intensity in the live image than in the mask image, causing mismatches for corresponding regions. It is therefore necessary to exclude the blood vessels during the registration process. The *exclusion template*, a binary image indicating where the contrast enhanced blood vessels are, has been proposed[3] for this purpose. This template is obtained by subtracting the mask from the live image before registration, and thresholding this difference image. The threshold is calculated as

$$T = \mu + n\sigma$$

where μ and σ are the mean and standard deviation of the pixel values in the difference image, and n is an empirically determined parameter. The positions of pixels that are set in the exclusion template are then ignored by the registration algorithm. However, as is the case with estimating mean gray-level variations before registration, this approach becomes more inaccurate with increasing spatial distortion.

We propose the use of selective sub-templates in the template matching as an alternative to the exclusion template. (The *exclusion* template, a binary image, should not be confused with the *matching* template, which is a sub-image used for template matching.) The matching template is applied as a tessellation of t smaller sub-templates. Referring to figure 2, the correlation measure is calculated separately for each sub-template and only the k best matching sub-templates are used to calculate the overall measure of match. Ideally, the t-k templates we ignore will contain any contrast enhanced blood vessels. With careful implementation, this approach need not add significantly to the computational cost of the template matching process. If the $\sum_A (f_{ij} - g_{ij})^2$ and $\sum_A (f_{ij} - g_{ij})$ terms of equation 3 are stored when calculating the measure of match for each sub-template, they can be summed for the retained sub-templates to calculate the final measure of match.



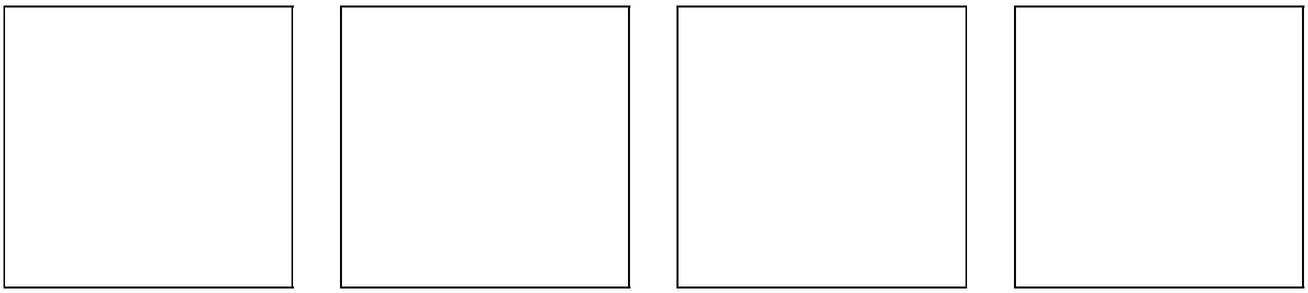
(a) Tessellated sub-templates. $t = 9$

(b) Blood vessel path.

(c) Final template. $k = 5$

Figure 2: Multiple sub-templates

Although we have overcome the problem of mean gray-level offsets in the registration by using a suitable measure of match, it is still necessary to estimate the disparity in order to remove it from the difference image. Sub-templates that are retained are unlikely to contain contrast enhanced blood vessels, and are ideal for mean gray-level estimation in the live image. During registration the mean gray-level of the retained sub-templates is stored. This information is used during the subtraction to remove any offset.



(a) Live image.

(b) Mask image.

(c) Difference image.

(d) Mean gray-level corrected difference image.

Figure 3: Mean gray-level disparity correction of the difference image

6 COMPUTATIONAL COST

Most DSA systems store a number of live and mask images and it may be necessary to repeat the subtraction with a different live-mask pair to obtain the best possible difference image. This re-processing should only take a few minutes,[5] but our method is computationally intensive, requiring approximately 34 hours on a Sun Sparcstation model 30 to prepare a motion vector field for the subtraction of a 512×512 live-mask pair. The simplest way to reduce this

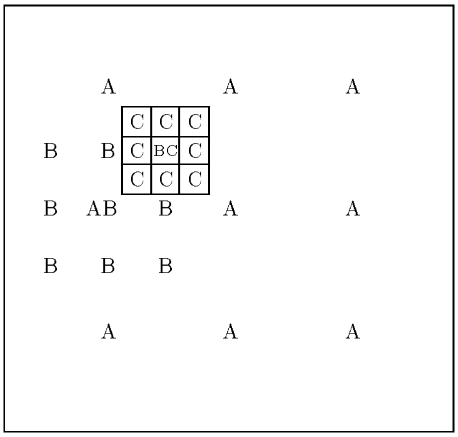


Figure 4: Three step search sequence

computational cost is by increasing the distance between control points and interpolating more of the vectors. For the particular study in figure 7(a) we found that a spacing of 5 pixels between control points was adequate. Larger spacing produced a significant loss of difference image quality. For a spacing of 5 pixels the computational requirement is reduced to 1.3 hours for a 512×512 pair. This is still unacceptable.

To further reduce computation a technique used in the coding of digital image sequences, the *three step search*, was implemented. Instead of exhaustively calculating a measure of match for every possible position of the live template (that is, the position of the center of the template) in the mask image, the search space is narrowed down in three steps. Figure 4 illustrates this procedure. The match measure is initially calculated at the positions labelled A, which directs the search to the positions labelled B, which in turn directs the search to the 3×3 region labelled C. With this scheme the measure of match is calculated 25 times, where an exhaustive search using a 15 pixel search radius would require over 700 calculations. This improvement reduces the 1.3 hour computation time to under one minute.

7 EXPERIMENTS

Test image pairs were obtained from a Philips Digital Vascular Imaging machine, printed onto x-ray film, and scanned by an XRS OmniMedia transmission scanner at 300 dots per inch. Corresponding regions in the image pairs were then extracted for experimentation. Two sets of experiments were conducted in order to evaluate the success of the different methods under investigation. The first applied the techniques to actual DSA image pairs, with subjective evaluation of difference image quality. Figure 7 shows two actual DSA studies.

The second set of experiments applied the techniques to simulated DSA image pairs where the parameters of the spatial distortion and mean gray-level disparity were predetermined. To quantify the results, a measure of success was defined based on the difference between the motion vector plot generated by the technique under evaluation and a 'perfect' motion vector plot generated with the same spatial distortion parameters used to create the simulated image. In the results presented here, the image in figure 8(a) is the mask image. A biquadratic mathematical warping function was used to obtain a spatially distorted live image. The coefficients of the biquadratic functions

$$x_i = a_0 + a_1 x_o + a_2 x_o^2 + a_3 y + a_4 x_o y + a_5 x_o^2 y + a_6 y^2 + a_7 x_o y^2 + a_2 x_o^2 y^2 \quad (4)$$

$$y_i = b_0 + b_1 x_o + b_2 x_o^2 + b_3 y + b_4 x_o y + b_5 x_o^2 y + b_6 y^2 + b_7 x_o y^2 + b_2 x_o^2 y^2 \quad (5)$$

were determined for a set of 12 tie-point pairs, which were in turn specified according to the spatial distortion required. The warping functions in equations 4 and 5 define an *inverse mapping*. That is, for each position in the output (warped) image (x_o, y_o), the coordinates of the corresponding position in the input image, (x_i, y_i), are calculated. In general x_i and y_i are not integers and bilinear interpolation was used to calculate the value at (x_o, y_o). Figure 5 is the motion vector plot of the biquadratic warp used for the results presented here. In order to simulate the effects of contrast

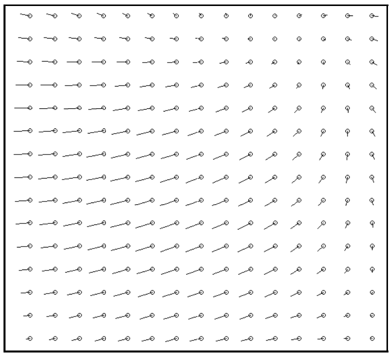


Figure 5: Motion vector plot for the simulated image pair (every tenth vector is shown).

enhanced blood vessels and gray-level offsets in the live image, the images in figure 8(b), 8(c) and 8(d) were added to the warped image in different combinations. A measure of success was based on the absolute magnitude of the differences between the x and y components of the motion vectors in the experimental and the calculated ('perfect') plot. These differences were added and an average value was plotted for each row and column.

Template sizes depend on the scale of the background structure, the size of contrast enhanced areas and the spatial distortion that is present in the image pair. Since we are not modelling any distortion within the template, its size should be small enough to neglect local distortion. Sub-template width should be in the order of the diameter of the larger blood vessels present in the live image. Template sizes of 33×33 and 21×21 were used, with 11×11 and 7×7 sub-templates respectively. Choice of the multiple sub-template parameter k, which specifies the number of sub-templates retained when calculating a measure of match, follows from the reasoning used when specifying template size. The second criterion for template size could state that contrast enhanced areas should, on average, occupy only A percent of the template. It follows that an upper bound for k is given by:

$$k < \left(1 - \frac{A}{100}\right) t.$$

However, the first criterion for template size implies that as many sub-templates as possible should be used to calculate a measure of match. We approximated A to be 20%, and values of 9 and 6 were used for t and k respectively. The maximum search radius was set to a value larger than the magnitude of the largest expected spatial distortion. Control point spacings of 1, 3, 5 and 10 pixels were evaluated.

8 DISCUSSION

Figure 9 shows graphs of the success measure for the columns of selected motion plot experiments. No smoothing of motion plot vectors was performed before the measure was calculated. Registration using the SAVD measure of match provided the most accurate motion vectors for the simulated image pair where the live image was only a spatially distorted version of the mask (figure 9(a)). This is to be expected, since VOD is analogous to SAVD, where the mean of the template and the sub-image it is applied to are made equal, and due to the spatial distortion the two original means would be calculated over slightly different information even in the best match position. Multiple sub-template VOD provides better performance than normal VOD since sub-templates can be ignored where spatial distortion within the template becomes significant. The mean gray-level independence of the VOD measure of match is illustrated in figure 9(b). Here the live image contains spatial distortion and the gray-level gradient shown in figure 8(b). This gradient is described by

$$g = x - N/2,$$

where x is the column number and N is the number of columns in the image. Figure 9(b) also gives some indication of the amount of gray-level offset that can be tolerated by SAVD. Considering the histogram of the mask image in figure 6, the dynamic range of the image is approximately 800 gray-levels. We can therefore estimate the offset tolerance of SAVD to be less than 5 percent of the dynamic range.

Figure 9(c) illustrates the advantage of the multiple sub-template approach. In the experiment shown, a crude simulation of different diameter blood vessels was added to the live image. These 'blood vessels' are shown in figure 8(d). The columns in figure 9(c) that indicate the greatest improvement over normal VOD correspond to the positions of the 'blood vessels' in the live image.

The effect of using different control point spacing, and relying on interpolation to provide the remaining vectors, is shown in figure 9(e). For the simulated image pair and the actual DSA studies we found that a 5 pixel spacing provided the best compromise between computational cost and difference image quality. However, as is the case for other algorithm parameters (eg. template size), this result is resolution dependent and a broader study is required in order to establish guidelines for specifying this parameter. Enlarging the control point spacing can have a smoothing effect, reducing the magnitude of errors while distributing them to neighboring vectors. This is true for the $S = 5$ case in figure 9(e). If the spacing is too large the vector field will not adequately describe the spatial distortion, which is the case for $S = 10$ in figure 9(e). Performance of the three step search was not adequate in any of the experiments. An example is shown in figure 9(f). This procedure often finds a local maximum, and not the global maximum on the match measure surface. An alternative fast search technique using a multi-scalar approach is being investigated. This method finds an approximate best match at a coarse resolution, and refines this approximation at successively higher resolutions.

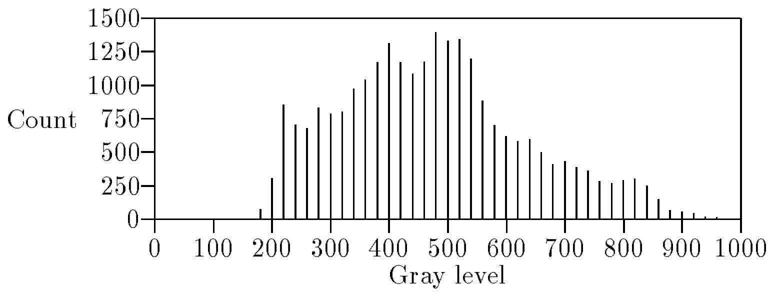


Figure 6: Gray-level histogram for the mask image of the simulated DSA pair

Figure 10 shows the difference images for three experiments with the simulated image pairs using multiple sub-templates and the VOD measure of match. For each experiment the live image, uncorrected subtraction and corrected subtraction are shown. Figure 10(c) is a binary image that indicates whether the center sub-template was among the sub-templates retained to calculate the overall measure of match at the best-match position. This image should indicate where contrast enhanced areas ('blood vessels') were encountered, which it does. Figures 10(g) and 10(k) are the estimated mean gray-level images, which correspond to figures 8(b) and 8(c) respectively.

9 CONCLUSIONS

Subjective evaluation of experiments on actual DSA pairs, two of which are shown in figure 7, leads us to the following conclusions:

1. In studies where the live image has a lot of background structure, spatial distortion, disparity in the background mean gray-level or any combination of these characteristics, corrected subtraction through registration offers a substantial improvement in difference image quality. This is illustrated by the experiment illustrated in figures 7(a) through 7(d).

2. In studies where there is relatively little background information, or where the live and mask image are already well matched in terms of spatial distortion and mean background gray-level, registration is unnecessary and an uncorrected subtraction is adequate. The latter is the case for the experiment illustrated in figures 7(e) through 7(h). Clearly the 'corrected' subtraction in figure 7 offers no improvement.

The techniques we propose are successful in providing a difference image that is corrected for spatial and mean gray-level differences between the live and mask images of a DSA pair. Use of the offset invariant VOD measure of match and a mean gray-level estimation procedure compensates for the mean gray-level disparity between the images. The multiple sub-template approach to template matching provides the basis for mean gray-level estimation and allows the registration procedure to ignore contrast enhanced areas in the live image. Although we have described a working system, we feel that the basic tools in the algorithm could be put to even better use. The information provided by the sub-templates, the match measure surface and the motion vector plot could be utilized by a more sophisticated system than the sequential procedure we have described. In addition, spatial distortion within the template should be catered for without incurring prohibitive computational cost. These issues are the subject of continuing research.

10 ACKNOWLEDGEMENTS

This work was partially funded by the Foundation for Research and Development (FRD) of South Africa. Technical assistance in the field of radiology and DSA image data was provided by the Radiology Department (in particular the head of department, Professor Steve Benningfield) at Groote Schuur Hospital, Cape Town.

11 REFERENCES

[1] A. Venot and V. Leclerc, "Automatic correction of patient motion and gray values prior to subtraction in digitized angiography," *IEEE Transactions on Medical Imaging*, vol. MI-3, pp. 179–186, Dec. 1984.

[2] K. Morishita and T. Yokohama, "Image registration method using adaptive nonlinear filter," *Systems and Computers in Japan*, vol. 19, pp. 41–50, Sept. 1988.

[3] L. V. Tran and J. Sklansky, "Flexible mask subtraction for digital angiography," *IEEE Transactions on Medical Imaging*, vol. 11, pp. 407–15, Sept. 1992.

[4] D. Wang and Q. Wang, "A weighted averaging method for image smoothing," in *Eighth International Conference on Pattern Recognition*, p. 981, Oct 1986.

[5] W. R. Brody, "Digital subtraction angiography," *IEEE Transactions on Nuclear Science*, vol. NS-29, pp. 1176–80, June 1982.

(a) Live image - 160×160 pixels.

(b) Mask image.

(c) Uncorrected subtraction.

(d) Corrected subtraction.

(e) Live image - 190×190 pixels.

(f) Mask image.

(g) Uncorrected subtraction.

(h) Corrected subtraction.

Figure 7: Subtraction results for two DSA studies.

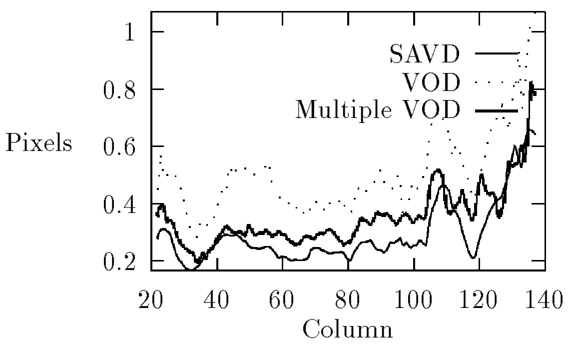
(a) Mask image.

(b) Gray-level gradient.

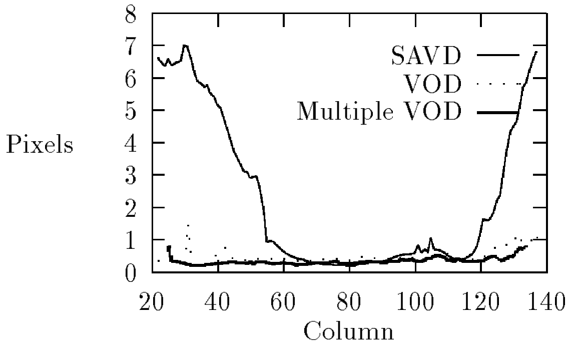
(c) Uniform gray-level offsets.

(d) Simulated areas of enhanced contrast ('blood vessels').

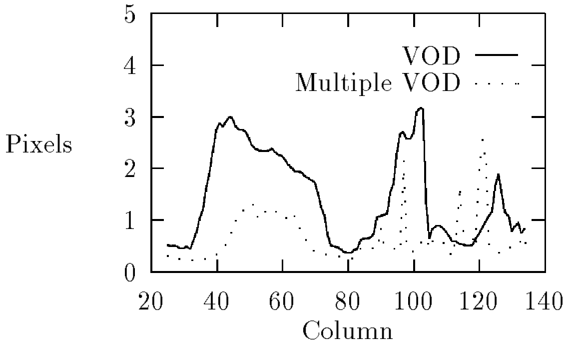
Figure 8: Image pair simulation



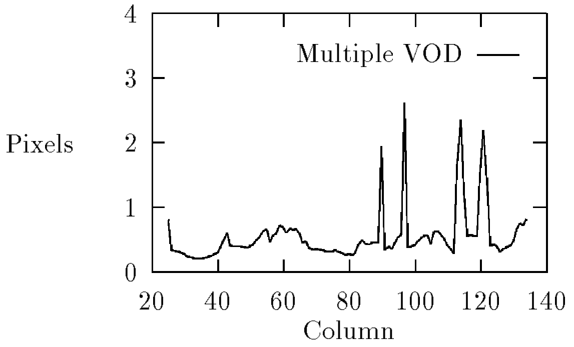
(a) Measure of match comparison for no gray-level disparity.



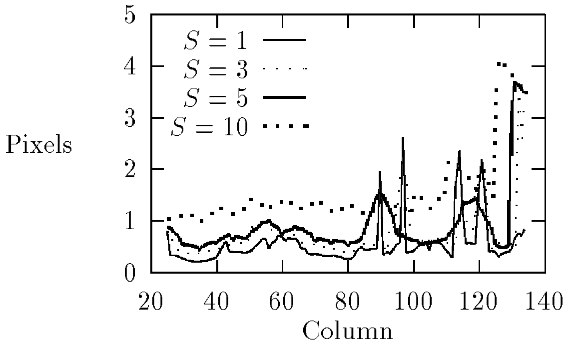
(b) Mean gray-level independence.



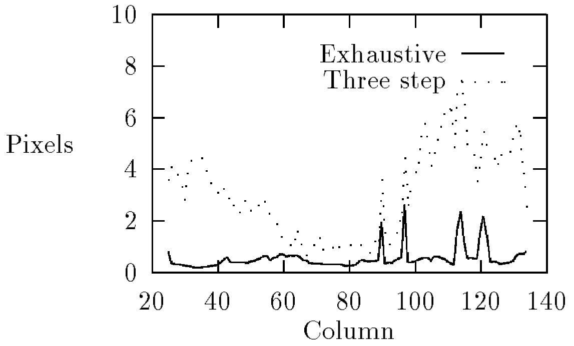
(c) Multiple sub-template performance with simulated blood vessels.



(d) Multiple sub-template performance with a gray-level gradient.



(e) The effect of control point spacing.



(f) Performance of the three step search.

Figure 9: Motion vector plot accuracy. In (a) the live image is a spatially distorted version of the mask. In (b) the live image is spatially distorted and has a gray-level gradient. In (c) the live image is spatially distorted and has simulated contrast enhanced blood vessels. In (d), (e) and (f) the live image is spatially distorted, has a gray-level gradient and has simulated blood vessels. Template and and sub-template sizes were 21×21 and 7×7 respectively.

(a) Live image: 'blood vessel' structures.

(b) Uncorrected subtraction.

(c) Multiple sub-template use. White areas indicate that the center sub-template was used.

(d) Corrected subtraction.

(e) Live image: 'blood vessel' structures and mean gray-level gradient.

(f) Uncorrected subtraction.

(g) Mean gray-level estimation.

(h) Corrected subtraction.

(i) Live image: 'blood vessel' structures and mean gray-level offsets.

(j) Uncorrected subtraction.

(k) Mean gray-level estimation.

(l) Corrected subtraction.

Figure 10: Results for simulated image pairs. Multiple sub-templates were used with the VOD measure of match. Template and and sub-template sizes used were 21×21 and 7×7 respectively.