

Quantitative Localisation of Manually Defined Landmarks

Hossein Ragheb
hossein.ragheb@manchester.ac.uk
Neil Thacker
<http://www.tina-vision.net/~nat>

ISBE, School of Cancer and Enabling
Sciences
University of Manchester
Manchester, UK.

Abstract

In this paper we propose a method for automatic placement of landmark points in 3D volumes, for use in morphological studies, which addresses the issue of quantitative estimation of measurement error. We are given a sample 3D volume of image slices (and a number of manually annotated landmark points) and develop an algorithm that is capable of estimating landmark locations in a similar but novel 3D volume. Problems are found which arise from the local minima generated when constructing likelihood functions using shifted noisy image patches, which must be addressed during both localisation and covariance estimation. The method is tested using Monte-Carlo in order to evaluate the quantitative validity of error estimates.

1 Introduction

In photogrammetry and computer vision localisation is generally constrained to purpose made patterns, affixed to the object ([1], pp. 5-10), with uncertainty assessed accordingly [5, 6, 7]. By contrast, our definition of landmark refers to a chosen location in a 2D image or 3D volume for the purpose of identifying corresponding locations in a second data set, in order to study natural shape variations. For automated location based upon statistical estimators, the accepted approach for assessment of errors is based upon the Cramer-Rao bound (CRB) [8], and this is the starting point for our work. Here we use smoothing in order to remove the effects of spatial noise in data, which (if ignored) can generate problems for optimisation, such as local minima. We therefore have to deal with these effects in order to obtain meaningful CRB estimates.

We assume we are given a number of reference mark-up locations within corresponding 3D volume data. We also assume that the problem of locating the corresponding object has been solved so that we are provided with the approximate orientation of the data and initial estimate of the landmark point. What is now required is an accurate measurement for the position of the corresponding landmark in the target block. Our approach is based on probability theory and template matching.

We use multiple reference examples of previous manual mark-up locations. The decision of how many reference candidates we need here should be based upon the resulting localisation accuracy. This requires a well-defined method to analytically extract the position of a single landmark point from the candidate positions available. One approach to search for landmark points in a 3D volume is to use its sub-volumes (*e.g.* $40 \times 40 \times 40$ data points) and reduce the 3D search problem to searching in three 2D images corresponding to the central

orthogonal planes. We expect information available for localisation to allow us to exceed sub-pixel accuracy (i.e. better than human repeatability) even for small regions, provided that there is at least some high contrast edge structure. It is important to note that error estimates must be valid, for instance landmark error estimates and corresponding variations along the minimum curvature direction must be quantitatively consistent.

Our approach is novel because: (a) we make use of direct measurement from a correctly scaled likelihood function (correctly scaled in this context means $-2\log L$ is distributed as a χ^2 [2]); (b) we compute quantitatively valid parameter covariances (location accuracy); (c) we show that sufficient accuracy is obtained using three orthogonal slices; (d) we use pre-processing in order to concentrate on where the information is, i.e. edges [2].

We focus on several important aspects of the problem to show that (a) without smoothing the data we cannot find a valid minima for the cost function; (b) using pre-processing we can obtain independence to grey-level scale (*e.g.* illumination and MR pulse sequence). We now discuss the main issues we must consider in order to deal with real-world data.

Scale: In this work apparent scale change is introduced due to developmental processes as well as the inherent differences in mature samples. However, unlike more general computer vision problems, the overall scale of a given structure in an image is largely fixed. We therefore assume that one overall approximate scaling of the data is sufficient to allow us to locate small sub-regions (*e.g.* a scale factor error of 5% will produce a negligible one-pixel shift in data at the edges of a region 40 pixels across).

Rotation: The accuracy of the automated estimate of initial orientation will need to be better than a degree in order to limit the effects of error on localisation (*e.g.* an error of half a degree will produce sub-pixel effects at the edge of a 40 pixel wide block) and its covariance. By matching to multiple candidate references we can however mitigate against large rotation errors using statistical criteria in order to select the examples with a good structural match. To quantitatively indicate how well sample data conforms to a predefined model, we plan to define a χ^2 based hypothesis. Such tests can also be used to identify missing landmarks.

Irregularities: For specific local bone structures there are a variety of shapes for which consistent landmark points need to be estimated. Manual annotation needs to have a clear strategy for defining landmarks on variable structures. Automation is therefore a challenging problem. One approach is to learn an appearance model [11] corresponding to the three orthogonal slices of each block. However, in our application we expect to have to work with a minimum number of reference examples ((insufficient to construct an accurate model), while at the same time topological differences between corresponding locations for identified structures are likely to be quite large, making appearance modelling inappropriate.

2 Methodology

We develop a matching algorithm for three orthogonal slices from the reference data block against the ones from the target data block. We use template matching based on the vertical and horizontal gradients corresponding to these image slices. This results in an algorithm which is robust to noise and imaging equipment settings. However, it is necessary to derive the method from appropriate statistical theory in order to define error covariances and so that meaningful χ^2 values can be used to confirm the goodness of fit.

We are given a reference and a target volume. In the reference volume, we are given a point for the manual landmark which is the origin of a small block ($40 \times 40 \times 40$ pixels). For this reference block we are given scaling and orientation values so that we can initially apply these to the target block. We then extract the corresponding re-scaled orthogonal slices, from the reference volume, in order to refine rough estimates of the landmark position.

Template Matching: According to [3] by applying the variational principle to the problem of matching two scaled noisy image patches, one can define the optimisation function

$$\chi^2 \propto \sum_n^N (\alpha I_n - \beta J_n)^2 \quad s.t. \quad \alpha^2 + \beta^2 = 1 \quad (1)$$

It follows that rather than using two scale factors α and β , one may use a single scale factor $\gamma = \alpha/\beta$. For similar patches it can be shown that $\gamma = \sqrt{B/A}$ where

$$A = \sum_n^N I_n^2, \quad B = \sum_n^N J_n^2 \quad \text{and} \quad C = \sum_n^N I_n J_n \quad (2)$$

To avoid lengthy execution times, one may expand the patch similarity measure and write

$$\chi^2 = [\sum_n^N (\gamma I_n - J_n)^2] / [\sigma^2(1 + \gamma^2)] = (\gamma^2 \sum_n^N I_n^2 + \sum_n^N J_n^2 - 2\gamma \sum_n^N I_n J_n) / [\sigma^2(1 + \gamma^2)] \quad (3)$$

$$\chi^2 = (\gamma^2 A + B - 2\gamma C) / [\sigma^2(1 + \gamma^2)] = 2(B - \gamma C) / [\sigma^2(1 + \gamma^2)] \quad (4)$$

When choosing the reference image patch to be J , it follows that B is the constant term while A is varying as the target image patch is moved around. Hence minimizing χ^2 in Eq. (4) is equivalent to minimizing

$$\chi_C^2 = (-2\gamma C) / [\sigma^2(1 + \gamma^2)] = [2/(1 + \gamma^2)] [(-\gamma C) / \sigma^2] \quad (5)$$

As this process requires adjustment to the assumed likelihood function (re-scaling the noise estimate) during optimisation, we can make a direct analogy to the method of Expectation Maximisation (EM). The standard proof of convergence for EM requires that we are not allowed to change the assumed likelihood distribution during the optimisation. Rather this should be fixed during the Maximisation step, and re-estimated during the Expectation step. Hence the $(1 + \gamma^2)$ term can be eliminated during optimisation¹. The term C works well when used in a conventional optimiser such as Simplex [9]. However, here rather than using gray-level image patches and least-square differences directly, their gradients in the horizontal and vertical directions are used. This reduces the dependency upon absolute scaling of the data making it more suitable for matching with MRI and CT datasets, though at the expense of reducing the capture range of the cost-function. In this case, C is the summation of the 6 dot products of the two-component gradient vectors originated from the 3 reference and target image patches. When using the analytic approach to compute the covariance, the constant terms are cancelled and χ^2 is automatically reduced to χ_C^2 .

Measurement Covariance: According to [2, 3], one may compute the inverse covariance matrix using the derivatives of χ where

$$C_\theta^{-1} = \sum_i (\nabla_\theta \chi_i)^T \otimes (\nabla_\theta \chi_i) |_{\theta=\theta_{max}} \quad (6)$$

However, in practice, as the contribution to the information (inverse covariance) matrix is assessed from each individual data point, this approach may be unstable in the presence of noisy data resulting in an unrealistic (over accurate) covariance. As a covariance is simply the second order shape of the total likelihood function, an alternative approach can be derived by observing the total change in χ^2 as a function of changes in the parameters ([10], page 11). This averages out the effects of noise giving more realistic estimates of parameter covariance. According to [4], the inverse covariance matrix C_θ^{-1} of parameters θ which is defined based on χ^2 is

$$\chi^2 = \chi_0^2 + \Delta\theta^T C_\theta^{-1} \Delta\theta \quad (7)$$

The χ^2 function used in the optimisation of parameters will be equal to χ_0^2 when there are no changes ($\Delta\theta$) in the optimum parameter values. For specific changes, however, we

¹If we were not to do this, the consequence would be an optimisation which had an optimal solution of $\gamma = 0$ i.e. a perfect match with infinite error on the re-scaled data.

can find linear equations between the unknown elements of the inverse covariance matrix C_θ^{-1} and the corresponding value obtained from the χ^2 function.

3 Experiments

We experiment with a CT scan of a Musculus mouse skull provided by our collaborators [12]. This consisted of 1003 image slices (656×656 pixels). We used five landmarks as reference points to test our measurement method. These were two well-constrained points at the end of two jaws, one point constrained well in vertical direction at the top of skull, and, two points constrained well in horizontal direction on the sides of skull. A novel 3D volume could be used to locate the landmarks corresponding to the examples given in the reference volume, but the systematic subjective error in landmark definition creates problems for defining the localisation gold standard with sufficient precision to accurately assess the covariances. Here we take the approach of using the same volume perturbed with Monte-Carlo noise to evaluate the results and covariances. In order to approximate use of a second data set, the level of added noise needs to be large enough so that the two noise fields are statistically independent. By varying the level of added noise we can also boost our test sample, 4 noise levels over 5 markup locations allows us to test the statistical method with 20 samples. We evaluate if the covariances estimated are a quantitatively valid summary of the localisation accuracy. We test how the optimisation works both when it starts from the answer (reference point) and from a rough estimate of the answer (up to 3 pixels away from the answer in each direction). We also change the volume orientation slightly providing more realistic data to investigate the behaviour of the method when the reference and test data come from two slightly different mouse skulls. We use a mean estimate of the image noise around these points ($\sigma = \sigma_o = 110$). To study the error residuals using Monte Carlo test we focus on the added noise that makes the ratio of χ^2/DoF about unity (DoF: degrees of freedom). Here $DoF = 40 \times 40 \times 6$. For Monte Carlo tests we adjust the χ^2 function by removing $\gamma^2 \sigma_a^2$ term from $\sigma^2(1 + \gamma^2)$ in Eq. (5), where $\sigma^2 = \sigma_o^2 + \sigma_a^2/2$, σ_o^2 is the variance for original noise and σ_a^2 is the variance for added noise.

In Fig. 1, we plot all error residuals $\Delta x/\sqrt{C_{xx}}$, $\Delta y/\sqrt{C_{yy}}$ and $\Delta z/\sqrt{C_{zz}}$ against standard deviation of the added noise for all five landmark points. As mentioned earlier Δx , Δy and Δz correspond to the smoothed image data with smoothing parameter η (in pixels) fixed at $\eta = 1$ while $\sqrt{C_{xx}}$, $\sqrt{C_{yy}}$ and $\sqrt{C_{zz}}$ correspond to non-smoothed image data $\eta = 0$ (Figs. 1.a, 1.b and 1.c). We observed that by adding noise at $\sigma_a = 600$ the ratio of χ^2/DoF becomes about unity. We then expect the error residuals observed in the Monte-Carlo study to fall in the range $[-2.5, 2.5]$ with their rms about unity. In Fig. 1, for $\sigma_a = 600$, when starting from answer the rms is 1.66 (1.a), when starting from a rough estimate of the answer the rms increases to 1.92 (1.c), and finally, when starting from the answer in the rotated block the rms increases again to 2.01 (1.b). In Fig. 1.d, we show that the minima cannot be found reliably when starting from a rough estimate of the answer if non-smoothed data are used. The plot corresponds to the same five landmarks used earlier. Here the error residuals are generally far too large to suggest that a good minimum was found. Hence by comparing this plot ($\eta = 0$) to that in Fig. 1.c, where the image data are smoothed for optimisation ($\eta = 1$), we can conclude that smoothing is necessary in order to find a minima.

4 Conclusions

The need for computing covariances together with any measurement has been overlooked in the appearance model literature. We proposed a method that fills this gap for quantitative landmark measurement and assessment. Our technique combines derivative and smoothing

operations to reduce the dependency on absolute grey-scale and spatial noise. We show empirically that the covariances obtained without smoothing are applicable to the results obtained with.

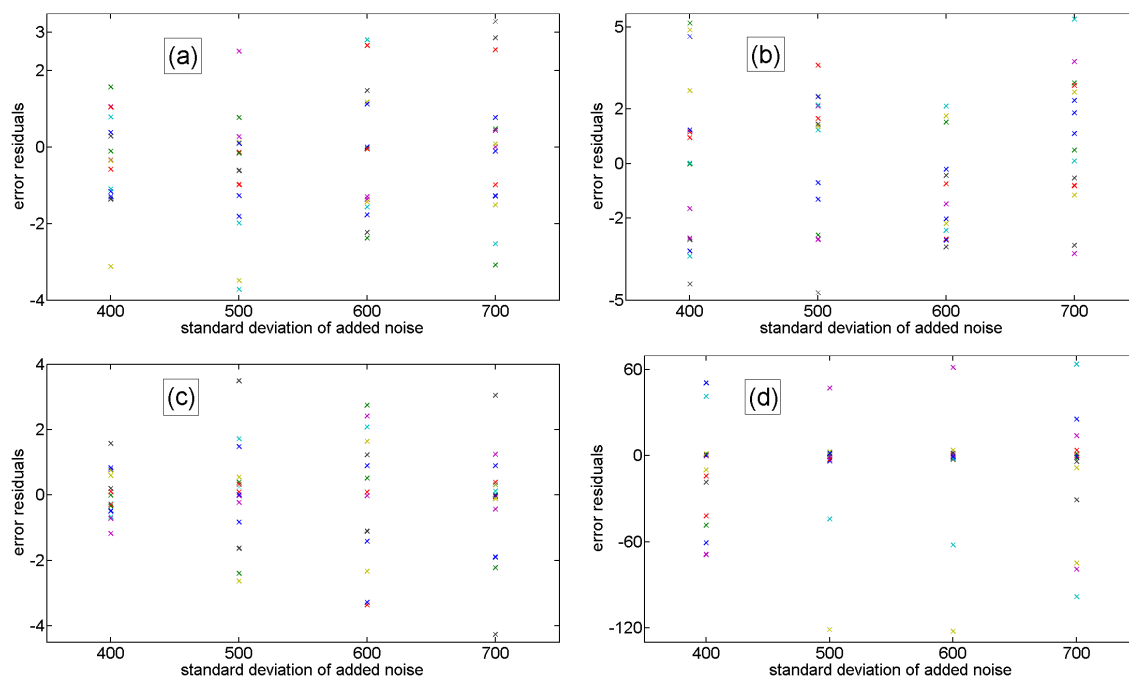


Figure 1: Monte Carlo test: error residuals against the standard deviation of added noise (5 landmarks), where (in a, b and c) optimisation is performed on smoothed data while covariances correspond to non-smoothed data; starting from the answer without (a) and with rotation (b) of the reference data block, and, starting from a rough estimate of the answer without rotation (c and d), while in (d) both optimisation and covariances correspond to non-smoothed data.

References

- [1] J.A. Gutierrez and B.S.R. Armstrong, Precision Landmark Location for Machine Vision and Photogrammetry. Springer, London, 2008.
- [2] P.A. Bromiley, M. Pokric and N.A. Thacker, Empirical Evaluation of Covariance Estimates for Mutual Information Coregistration. Proc. MICCAI, pp. 607-614, 2004.
- [3] N.A. Thacker, J.V. Manjon and P.A. Bromiley, A Statistical Interpretation of Non-Local Means. Proc. VIE 2008, pp. 250-255, 2008.
- [4] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes in C. Cambridge University Press, Cambridge, 1988.
- [5] R. Bansal, L. Staib, A. Laine, D. Xu, J. Liu, L. Posecion and B. Peterson, Calcul. of the Conf. Inter. for Trans. Model Param. in the Regis. of Med. Images. Med Im Anal, 13(2):215-233, 2009.
- [6] J.M. Fitzpatrick, J.B. West and C.R. Maurer Jr, Predicting Error in Rigid-body Point-based Registration. IEEE Trans Med Imaging, 17(5): 694-702, 1998.
- [7] I. Grachev, D. Berdichevsky, S. Rauch, S. Heckers, D. Kennedy, V. Caviness, N. Alpert, A Met. for Asses. the Acc. of Int. Reg. of the Hum. Br. Us. Anat. Land. NeuroImage, 9:250-268, 1999.
- [8] A. Liu, T. Kanungo, and R.M. Haralick, On the Use of Error Propagation for Statistical Validation of Computer Vision Software. IEEE Trans Pattern Anal Mach Intell, 27(10):1603-1614, 2005.
- [9] J.A. Nelder and R. Meade, A Simplex Method for Function Minimisation. Computer Journal, 7:308-313, 1965.
- [10] A. Stuart, K. Ord and S. Arnold, Kendall's Advanced Theory of Statistics Volume 2A: Classical Inference and the Linear Model. Arnold, London, 1994.
- [11] T.F. Cootes , G.J. Edwards and C.J. Taylor, Active Appearance Models. IEEE Trans Pattern Anal Mach Intell, 23(6):681-685, 2001.
- [12] The Max Planck Institute for Evolutionary Biology, www.mpg.de/english/portal/index.html.