Unification of Appearance Models and Non-rigid Registration

Abstract

Statistical models of appearance possess the ability to faithfully describe shape variation, as well as the variation in pixel intensities. Nonetheless, in order for such models to synthesise valid full appearances, it is essential that landmarks can be consistently identified in the training sets defining these models. One way of automatically achieving data landmarking is to establish an overlap across the training set, much as is already done in image registration.

We contend that the task of registration should seek a globally correct answer, rather than applying transformations that are reliant on the choice of a single reference image. The intrinsic power of appearance models, as well as Minimum Description Length considerations, allow us to define a fully automatic group-wise non-rigid registration scheme.

Various observations motivated us to investigate the innate bonds between non-rigid image registration and appearance models. Unification of the two will be mutually beneficiary and can entail a novel image analysis and alignment framework.

A complex and highly flexible application we constructed and called AART (Autonomous Appearance-based Registration Test-bed) provides proof of the premise above. It benchmarks genuine registration algorithms against other well-performing pair-wise algorithms and newly-conceived methods to illustrate the advantages gained by unifying registration and statistical modeling.

1. Introduction

Registration has become a vital pre-processing step in the analysis of bio-medical data where unpredictable yet restrained variations are inherent. In recent years there have been attempts to find an alternative to the arbitrary choice of a reference to carry out registration [1] and such endeavours presently continue. We base our research on the hypothesis that better paradigms exist for finding solutions to the registration problem and not only will these solutions be more precise, but also they will be globally optimal, suggesting that they may represent the *correct* answer.

Utilisation of registration methods for the establishment of overlap is made ever more appealing as it allows the construction of appearance models *automatically*. Thus far, good appearance model formations had the pre-requisite that good landmark points needed to be identified in the whole training data, thereby indicating where analogous regions lie. Following the successful work of Davies [2] on shape models, it is known that the construction of optimal statistical models can be conveniently treated as an optimisation problem. An objective function is defined which aims to find a model that is most concise and still fits all of its training data. The inference of the "goodness" of a model is assisted by the Minimum Description Length (MDL) criterion, relying on the assumption that simple descriptions are most preferable. In line with this approach, we will define a distinct way of representing an appearance model and the data encoded using that model. Once such messages can be composed, we shall seek a message whose length is minimal. Lower message lengths imply lower variation in the model and hence increased similarity across the whole data set. The optimisation problem itself is guided by a reparameterisation that sets the regime by which the shortest message is to be found.

2. Background

This inter-disciplinary research work is based on numerous areas with vague overlaps, yet only the three most predominant ones are explained in this section. The first two, broadly model-driven image analysis and image registration, are the subjects to be unified and mutually enrich one another, whereas the third which is MDL is strictly a heuristic technique that has been shown in the past to be the most powerful methodology for solving the similar problems which had been posed. The realisation of MDL is essential although alternatives do exist and are perpetually investigated under AART.

2.1. Active Appearance Models

The task of image analysis, especially in the bio-medical domain, must take into consideration the variation in shape and appearance of objects. The invariant presumption is that corresponding objects in all images are of one particular class so we can typify the contents of the image by training an entity that captures inter-subject variation as well as atrophies.

Statistical analysis of shapes [3] which obtains a model of deformation goes back a decade ago. The principles were later extended to sample the variation in pixel intensities (also commonly referred to as textures) to create a model of full variation that is able to synthesise full appearances [4] and their successful application to medical data has been frequently demonstrated [5]. The correlations between shape and intensity are learned using Principal Component Analysis [6] where much of the power of these principles lies.

The integrity of models breaks down if correspondences, annotated in the form of spatial landmarks, are inappropriately identified. Furthermore, the annotation process involves a preliminary segmentation process which highlights parts of the data where landmarks can and should be placed. Although this has become a solved problem in statistical modeling of shape, it is yet difficult to select good landmarks in images which strive to retain full appearances rather than contours or surfaces solely. Several attempts have been made to resolve the issue [7, 8, 9], but none was optimal or even quite satisfactory. Alignment has become the means by which this crucial limitation can be solved and the foundations of image registration assist in establishing this alignment.

2.2. Image Registration

In the medical domain, one of the more fundamental problems is the requirement for the setting of images in a state which makes them appear collectively similar [10]. This greatly simplifies the analysis of a group of images which bear common information, as in the case of brain slices fusion or comparison of patient data, either acquired using different modalities or collected at different time instances.

The problem is trivial if the difference is a rigid one – a difference due to rotation, scale and translation. More realistically, the problem is far more complex and images are inconsistent (primarily in the case of inter-subject registration) so affine and non-rigid transformations are required. In the case of non-rigid registration, transformation is merely unbounded. However, to avoid corruption and distortion of constituent finer parts of the image, limitations to their freedom and certain conditions must be met. Clamped-plate splines (CPS), which are based on Green's function, have proven to be a useful family of warps, allowing for highly flexible manipulation of images. Their attributes are reminiscent of those developed by Lötjönen and Mäkelä [11].

To drive transformation in the right direction and attain convergence, minimisation of the difference perceived in the images must be pursued. To measure discrepancies, or contrariwise, the similarity between two images, mean of squared differences (MSD) or mutual information (MI) [12] are traditionally used as metrics although new techniques are perpetually introduced [13].

Overall, the process of registration comprises the transformation of images followed by similarity measures, where transformations are chosen to iteratively maximise that similarity. Conventionally, a reference is selected in the process [14], but our contention is that this need not be the case if an optimal solution is sought. The technique according to which the registration problem will be solved is entirely described by the objective function as Section 4 illustrates.

2.3. Minimum Description Length

The notion of "information gained" and homologous uncertainty measures were initially proposed by Hartley. These were later extended by Shannon, a considerable amount of time before becoming applicable to registration [15]. Shannon quantified the amount of uncertainty and named it *entropy* – a very fundamental term in information theory.

The relevance of this information-theoretical principle to our research stems from the fact that excess of information that is encapsulated in messages is indeed measurable. Compression implies data simplification abilities and the less uncertainty is present in the data, the less information will be required to encompass the redundancy in that data. This therefore can be looked at as a measure of the *simplicity* of data, enabling prediction and detection of patterns. Minimum Message Length or Minimum Description Length (MDL), as defined by Rissanen [16], is a useful mechanism for discovering the complexity of some data and following Occam's Razor, models that are most concise, should also fit best.

We are then left with the decision of how to construct a message which describes model quality. In practice, such a description should not only consist of the model, but it also must contain all the training data with its relationship to that model.

3. Methodology

A significant step of any registration algorithm involves the transformation of a set of images which possess commonalities and constrained variations. Unlike existing registration schemes, where a single reference image is chosen for comparison with each of the other images in a pair-wise manner, we take a global scope of the problem. We believe that in order to find the global optimum, we ought to construct an appearance model from all the given images and transform them, one image at a time, while aspiring to minimise the resulting model's complexity.

This means that instead of a simple reference image, a richer entity of reference is used to define our target and this entity is dynamic.

It is clear that the model has a representative quality with respect to the images it encapsulates. Undoubtedly it results in a reference that is based on *all* images, generalisable to all images and should be superior to other reference types. For instance, if a mean reference is used, real data variation is lost and the reference becomes overly blurry – both problems that appearance models avert. The only drawback of the approach we take is that appearance models are quite slow to construct; time is dependent on the size of the set to be registered and yet such constructions form the very heart of the optimisation. In practical terms it means that model re-computation is expected to be frequent and for large numbers of images, the process will slow down considerably. When a multi-scale approach is applied to the model and the images, this discouraging aspect can be partially tackled.

4. The Objective Function

Objective functions define the means by which a solution is to be found. As explained before, efficiency is a reasonable concern so a sophisticated function that is prudent to construct the model more frequently than necessary must be employed. Our function needs to drive the search for transformations which increase similarity across all images. The different nature of the problem and the methods of solving it convey the ulterior goal somewhat differently than the vast majority of methods to date, resulting in the formulation below.

Let us denote a transformation function $W(\bullet, params)$ and the construction of an appearance model to be $Model(\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_n)$ where $\mathbf{x}_i$ are the images used to train that model. We seek a model that is more compact using the following (simplified) function

$$F_{obj} = MDL(Model(\mathbf{x}_1 ..., \mathbf{x}_i .., \mathbf{x}_n)) - MDL(Model(\mathbf{x}_1 ..., W(\mathbf{x}_i, params) .., \mathbf{x}_n))$$

where *params* should be found to minimise this expression for each image vector $\mathbf{x}_i$. A succinct description of our algorithm is as follows:

— Repeat
  — For each image vector $\mathbf{x}_i$,
    — Optimise F_{obj} by altering the values of *params*.
— Until convergence.

A *combination* of different objective functions, e.g. that which we have implemented for MSD, MI, normalised MI, wavelets and probability density functions alongside the model-based approach, produced encouraging results as well. We found that the performance of different methods varies depending on the data and the distance from convergence and AART can autonomously try a mixture of methods, thereupon hybrid objective functions are introduced. In practice, to indirectly and quickly evaluate MDL we obtain $\sum_{i=1}^{n} log(\lambda_i)$ where $\lambda_{1<i<n}$ are the n eigenvalues of the covariance matrix whose magnitudes are the greatest [17].

5. Experiments

AART was used to carry out all our experiments in a stable environment which is straightforward to handle. Outputs varied from data files to graphs, images and videos of several types. Long-running simulations provided the required comparative results which revealed the strengths of the algorithms proposed in this paper.

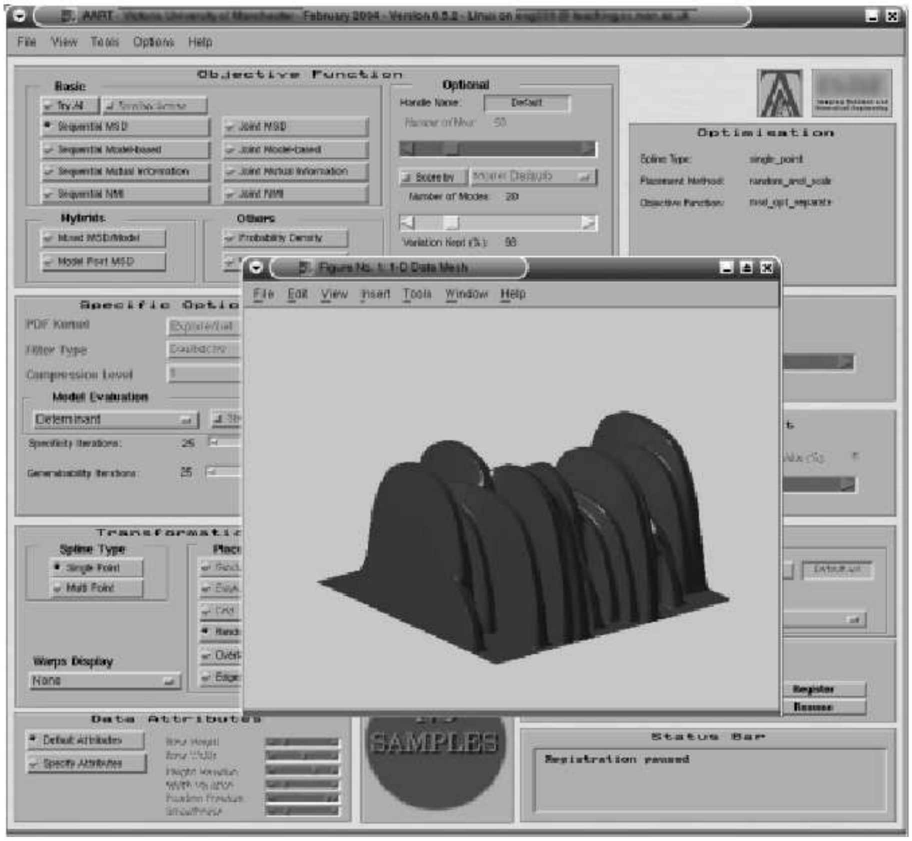

Fig. 1. AART investigation of the dynamic registration behaviour by displaying video sequences

The application outputs, in particular the many types of graphs and videos have been thoroughly analysed by a panel to derive conclusions regarding the behaviour of the objective functions and perform subsequent experiments accordingly. The exact objective functions and optimisation regimes for all methods are beyond the scope of this paper.

The following concepts are worthy of further elaboration. Some of these have never been used before so their potential can be truly comprehended for the first time.

Probability Density Function (PDF): Such functions describe the volume of data distributions. More uniform data, as we aspire to achieve across all images during registration, will result in lower such values. An exponential PDF was used in the experiments by default although over a dozen others are available in AART, including a Gaussian one.

Wavelets: As compression [19] is closely related to MDL, these can provide an accurate estimate of the complexity of data and abundance of patterns within that data. An extensive group of different wavelets are offered by the

application and, by default, Daubechy was used in the experiments. Computationally cheaper alternatives to the wavelets are Fourier and Hough transforms, but these have not yet been incorporated into AART.
Mutual Information: This strand of methods [18, 12] will analyse the peaks of image histograms. Normalised MI is currently one of the most robust and widely-used methods for 2-D data.

6. Results

Model-based algorithms result in appearance models whose determinant is by orders of magnitude lower than that which is measured at the start. The execution time they impose is inferior to MI, much as we expected all along, but is superior to that of MSD.

Since many of the statistics we gathered could indicate that we achieved what we had set our program to do, we intend to apply the same concepts to 2-D and 3-D data in the future. The principles remain unchanged and the only required extension is that of CPS to a higher-dimensional space – something we have developed already. The only foreseeable hindrances will then be the speed of execution and the diligent selection of knot-points for transformation.

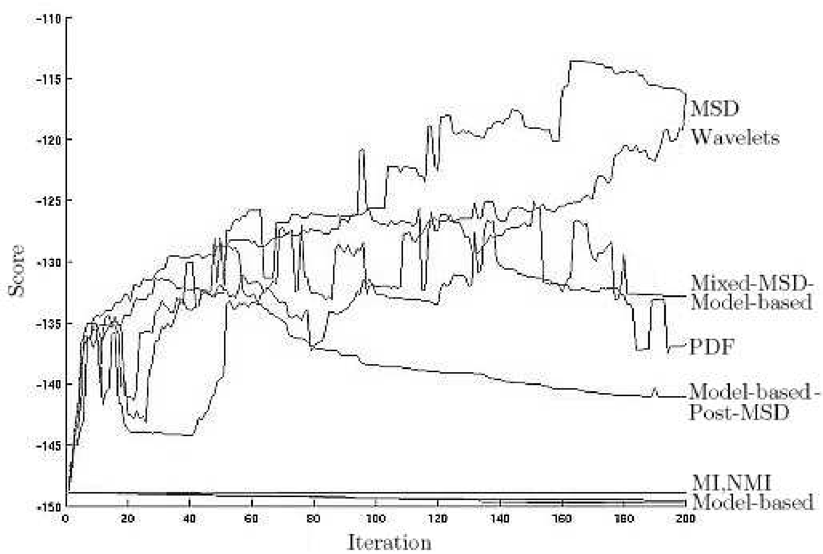

Fig. 2. Overlaid plots showing the quality of the appearance model resulting from each objective function.

For the results in Figure 3, the number of iterations was set to 50. For single-point transformations, the placement of the control point was random (both in location and magnitude) and for multi-point transformations the positioning of points was made random to abstain from data-bias or advantageous *a-priori* knowledge. The number of data instances was kept high at 20 in order to allow a substantial *group-wise* optimisation to be investigated. Objective functions based on mutual information remain flat simply due to the continuity of the data and the fact that it is one-dimensional. The table below shows the different values of $log \prod \lambda$. λ are the eigenvalues derived from the covariance matrix of the appearance model which was constructed from all 20 data instances. For completeness, differentiation is provided for optimisations which reparameterise over all dimensions at once (joint) or do so separately (sequential).

Objective Function	Single-point Warp	Multi-point Warp
PDF	-137.2658	-136.7145
Wavelets	-145.4988	-147.7877
Joint Model-based	-149.2192	-150.3197
Sequential Model-based	-148.7245	-149.9904
MSD	-143.0227	-149.0114
Joint MI	-142.3415	-136.0651
Sequential MI	-142.3712	-136.0651
Joint NMI	-142.3154	-142.3068
Sequential NMI	-142.3154	-142.3118
Model after MSD	-138.6823	-47.0961
Mixed Model/MSD	-129.3791	-105.3422

Table 1. Comparison of objective functions.

7. Conclusions and Contributions

We have illustrated not only that group-wise registration based on appearance model is possible, but also that it surpasses registration that is based purely on a reference image.

The current algorithms are being interpreted rather than compiled and no multi-scale approach is yet in use. The extension of the algorithms to 2-D and 3-D would require a long time to run, but remains practical. Compiled implementations should be available within months as well as heuristic optimisations that will make run-time more competitive with that of pair-wise approaches. Furthermore, the results have been shown to be better in a global sense and are not dependent on just one individual image.

Contributions of the work can be subdivided into three aspects:

1. It provides a benchmark environment and results for many methods, including several new ones.
2. Unprecedented model-based group-wise registration is introduced.
3. Automatic construction of increasingly better appearance model becomes practicable. Correspondences are obtained using techniques borrowed from image registration.

Acknowledgements. The project is funded by <<body>> and forms a part of the <<research>>. Contribution to this work and important discussions involved <<anon>> and <<anon>>. A significant proportion of the application's code were worked on by <<anon>>.

References

[1] S. Marsland, C. J. Twining, and C. J. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *proceedings of MICCAI 2003*, pages 771-779, Montreal, Canada, 2003.

[2] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525-537, 2002.

[3] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor. A unified framework for atlas matching using active appearance models. In *Proceedings of Information Processing in Medical Imaging*, Lecture Notes in Computer Science 1613:322-333, 1999.

[4] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In *Proceedings of European Conference on Computer Vision*, 2:581-595, 1998.

[5] M. B. Stegmann, B. K. Ersboll, and R. Larsen. FAME - a flexible appearance modeling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319-1331, 2003.

[6] I. T. Joliffe. Principal component analysis. In *Springer Series in Statistics*, Springer, New York, 1986.

[7] A. D. Brett and C. J. Taylor. A method of automated landmark generation for automated 3D PDM construction. *Image and Vision Computing*, 18(9):739-748, 2000.

[8] K. N. Walker, T. F. Cootes, and C. J. Taylor. Automatically building appearance models from image sequences using salient features. *Image and Vision Computing*, 20(6):435-440, 2002.

[9] A. Hill, C. J. Taylor, and A. D. Brett. A framework for automatic landmark identification using a new method of nonrigid correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):241-251, 2000.

[10] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes. Medical image registration. Boca Raton, Fla. ; London: CRC Press, 2001.

[11] J. Lötjönen and T. Mäkelä. Elastic matching using a deformation sphere. In *Proceedings of MICCAI 2001*, pages 541-548, 2001.

[12] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71-86, 1999.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986 - 1004, 2003.

[14] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8)1014-1025, 2003.

[15] S. K. Warfield, J. Rexilius, P. S. Huppi, T. E. Inder, E. G. Miller, W. M. Wells, III, G. P. Zientara, F. A. Jolesz, and R. Kikinis. An entropy measure to assess nonrigid registration algorithms for statistical atlas construction. In *Proceedings of MICCAI 2001*, pages 266-274,2001.

[16] J. R. Rissanen. Stochastic complexity in statistical inquiry. In *World Scientific Series in Computer Science*, Singapore, 1989.

[17] A. C. W. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by genetic algorithm. In *Information Processing in Medical Imaging*, 1997.

[18] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24:137-154, 1997.

[19] M. Rabbani and R. Joshi. An overview of the JPEG 2000 still image compression standard. *Signal Processing: Image Communication*, 17:3-48.