Tina Memo No. 2006-006
Internal Report.

Tutorial: A Critical Assessment of the Use of Riemann Manifolds for Shape Analysis.

N.A. Thacker.

Last updated
25 / 9 / 2006

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

A Critical Assessment of the Use of Riemann Manifolds for Shape Analysis.

N.A.Thacker

The book "The Statistical Theory of Shape" [?] describes the basic approaches now taken by the Kendall School of shape analysis and covers a subset of the material also to be found in Kendall et. al.'s"Shape and Shape Theory" [?]. It explains the motivation for, and the approaches to computing statistical tests on measured landmarks. It discusses the fundamental problems associated with this and also solutions described in terms of Riemann manifolds. These manifolds are used in general relativity as a way of encoding the change in oriented space distances (expresses as a 'metric') as a function of the local gravitational field. The key point is that the mathematical machinery exists to apply non-linear transformations to these spaces while maintaining meaningful definitions of physical distance.

This review compares these approaches with conventional methods for statistical analysis. As a consequence, though the page by page mathematical descriptions of the approach are of course beyond reproach (though perhaps overly "set theoretic" for a general audience), I have some (I would say substantial) misgivings regarding the practical use of these techniques in scientific studies. Despite the heavy presentation style, this book provides a good opportunity to summarise the appropriate criticisms with respect to what many people would regard to be a fair introduction to the topic.

This work has been done for the purposes of documenting my opinions for collaborators and PhD students. I have no intention at this time to send this material for publication, but it will be made available from our web pages (www.tina-vision.net).

Background

The introductory chapter in [?] lays out the motivation for an approach to the analysis of shape, and provides some examples of the kind of data sets researchers and scientist may like to ask statistical questions about. One specific example is the analysis of primate skull shape. Here, the aim is to model the variations in the location of sets of structurally corresponding points across multiple images which display skull shape variation. Though other examples are given later in the book, we might agree that this is a good starting point for us to consider the issues involved.

Before we start I think it is necessary to explain what I understand to be the main scientific problem embodied in the analysis of shape, such as the skull example. This will then make it easier to identify potential problems with respect to the methods suggested here. The scientific method as we understand it is about trying to identify and test theories of the world around us. The methodology for application of these tests is necessarily probability theory and statistics

For many scientific theories we have plausible candidates which can be suggested which satisfy our prior expectations of a solid theory, conservation of energy, dimensional consistency etc. Leaving aside (for now) null hypotheses for comparison of class distributions, for problems such as the analysis of shape the problem can often be considered more as a data mining exercise than the test of a specific parametric theory. We would like our analysis to identify for us the major characteristics which demand a scientific explanation. For shape, this is acknowledged as a complex and non-trivial problem. The fundamental question we need to ask is; **Is shape theory, and in particular the use of geometric approaches such as Riemann manifolds, a justifiable way to address scientific interpretation of data?**

My idea of a solution would involve being able to take a dataset of measured landmarks and to automatically identify and parameterise the main characteristics. We would expect this process to be limited by, and have to take account of, the information present in the measured data. A conventional statistical analysis of shape would involve the prior construction of a shape model which would then be fitted to observed data in order to test its adequacy up to the measurable limits of the data. The main problem with shape analysis is probably that we start with a set of measured points, and we have no parametric theory. In accordance with Occam's razor, we seek the model which has the least complexity. Unfortunately this is not a sufficiently well defined aim to allow us to evaluate models. What is needed here instead is a quantitative definition of the data analysis task, and indeed an approach has been identified which does result in the appropriate behaviour. **In statistical terms this can be restated as saying we want the model with the best predictive capabilities on new data drawn from the same distribution.** This test is the classical statistical problem of model selection which embodies a key part of the scientific method. This is also the criteria adopted by many in the machine learning, pattern recognition and neural network communities as the fundamental basis for learning systems.

The Book

The first chapter of this book explains the basic problems of shape analysis associated with the need to represent shape in a way which is independent of location, orientation, and often, but not always scale. The conventional approach involves what is called a "Procrustes" alignment which determines estimates of parameters which can align a given shape to a standard "space". The variations in shape across the sample of data can then be analysed once the data has be transformed into this normalised space. The most common method for applying these processes is often referred to as point distribution modelling using Principle Component Analysis. This is consistent with an assumed orthogonal linear model for parametric shape. Standard extensions which may produce more meaningful scientific interpretations would include Independent Component Analysis. These models do not have any intrinsic hierarchy for theoretical validity, they should be strictly compared on the basis of their predictive capabilities for any data set as described above.

Chapter 2 of the book explain the Riemann interpretation of shape spaces in terms of non-linear re-parameterisations of the original data and the characteristics of these manifolds. It then suggests the idea of Bookstein co-ordinates as a scale invariant way of representing shape. The next few chapters identify various terms, definitions and properties of Riemann manifolds. For the purpose of understanding the fundamental suitability of Riemann manifolds for shape analysis the definitions of differential geometry and properties are actually irrelevant. The one exception here is the requirement of differential continuity. If we define the shape analysis problem as that of quantifying the difference between variations in a set of corresponding points such as with the skull shape analysis example, then the the property of differential continuity is immediately violated in many practical data sets. For example, the analysis of shape in medical image data would immediately run into trouble at organ boundaries (that represent mathematical discontinuities which can move in opposite directions under physical forces), with the inclusion of tumours (which add additional structure where there was none before) and regions of cell death (which looses structure). Clearly, such cases often provide the datasets which we are most interested in being able to interpret. Similar problems occur with projections of 3D objects, the most obvious being self occlusion which can eliminate corresponding points in some viewpoints. Unfortunately, no amount of subsequent mathematical rigour can make up for inappropriate initial assumptions. Quite obviously;

Point 1: Real data does not always satisfy the requirements of Riemann geometry.

It is only much later in the book chapter 4, with the introduction of the Helmert matrix, that we begin to see how the theory matches up to a practical approach to data analysis. In particular it is suggested that simply subtracting the mean from a set of data is actually an approximation to a formal geometric approach involving the use of Helmert matrices to re-project the data into a pre-shape space. This is lent further credibility by observing the characteristics of Gaussian IID data distributions in the projected space.

Indeed, the discarded first row of a Helmert matrix computes nothing more than a rescaled estimate of the mean. In addition, the various analyses on the properties of distributions projected using Helmert matrices are only special cases of the more general, and much simpler, observation that distributions are preserved under rotation into any orthogonal space. This includes construction of principle axes as advocated in point distribution models (above). So by discarding the first term in the new shape vector, and taking the new orthogonal data vector using the Helmert matrix as suggested, the conventional approach and geometric approach are equivalent. However, this is the first point in the book at which it is possible to identify a gulf between a statistical test of a theory and Riemann geometry as the theoretical basis for data analysis.

From the point of view of statistical analysis, we would interpret the calculation of a shape mean as being an attempt to remove the arbitrary location of the object from the shape measurement. We would probably all agree to begin with that this is not a simple problem and there may in fact be no meaningful way to define an equivalent location for two different shapes in the absence of the correct parametric model. To quote from Bookstein's criticism of the VBM[1] approach to analysis of brain deformation ".. imprecise alignment of a dataset leaves the true variability (deformation) of the data unknown". However, using the statistical approach with a parametric model, we would expect this calculation to be done in the way which gives the

[1]Voxel Based Morphometry.

most stable alignment estimate. The mean is only the appropriate way to do this if we have selected a model which defines the mean as a parameter and also have IID Gaussian errors on the landmark locations (not the data distribution, the errors (see above)). In particular if the characteristic localisation error varies for different landmarks then the most accurate localisation is obtained with the appropriate statistically weighted mean, not a simple mean. If you believe that this is the basis for what you are doing then the use of Helmert matrices in the re-projection of the shape space immediately becomes inappropriate for achieving the most predictive location of a shape in the general case. Moreover, the geometric approach ignores one very important issue, orientation. In summary;

Point 2: Use of Geometry as the principle for data normalisation conflicts with the scientific principle of using statistics for parameter estimation.

In the next chapter a method is proposed for the use of Bookstein coordinates in the analysis of shape. This representation has been analysed in order to determine how a distribution of variation will transform into this co-ordinate space. Adjustments can then be made to estimated densities using the methods of projective geometry described in the earlier chapters in order to account for the transformation process as applied to the shape distribution (eg: biological variability).

One limitation in the work as described here is that no account is taken of the way measurement precision affects the stability (precision) of the resulting representation. Loosing sight of the errors in a measurement process precludes the possibility of assessing the validity of any parametric model for shape (see below). In particular, this prevents solution of problems such as shape analysis of skulls mentioned in the first chapter. As Bookstein has illustrated, some questions can be answered with this approach. However, the full machinery of differential geometry was not necessary to achieve this task. The same result can be justified from a statistical standpoint with more conventional methods for transformation of random variables using the Jacobian. Moreover, if the original data vector (X) has the property of uniform independent errors on location measurement, then the density transformation applied in Bookstein's work is directly related to the process of error propagation. In fact the metric used to compute the volume on the Riemann manifold takes the place of what we would have obtained by defining the statistical similarity between measured vectors in terms of a measurement covariance. Perhaps statistical similarity in Bookstein coordinates may be computable using the methods of differential geometry (p 48). I think it is possible to show that this is not the case, and the argument doesn't need specialised knowledge of differential geometry or set theory. Indeed, the set theoretical definition of probabilities on manifolds (p 118) does not protect us from this problem. My reasons for this opinion will now be covered in the next few sections and are summarised by points 3-5.

Composition of a shape vector of Bookstein co-ordinates for more complex shapes, as described in this chapter (p 150), introduces correlations between the components ($z_1 - z_{n-2}$) in the latter part of the shape vector, due to the measurement errors in the common landmark points (x_1, x_2). Thus any meaningful statistical similarity calculation can no longer assume independence between the z's, as required if we try to use the metric on a Riemann manifold to encode measurement error. We could argue that Bookstein coordinates is simply a bad choice for the representation of shape, and try to select Riemann spaces which have statistical independence, but this misses the point. If this approach had been fundamentally the correct way to solve the shape similarity problem then the theory of geometry would have been capable of identifying the problem of correlation without having to make comparisons with conventional statistics and would also contain the machinery for a solution.

In other words;

Point 3: Interpretations of shape based soley on arguments from differential geometry cannot be expected to be statistically valid in the general case.

At this point, advocates for inherent theoretical validity of differential geometry and the use of Riemann spaces might begin to argue that forcing a statistical interpretation on shape construction is a matter of faith rather than scientific necessity. I would respond by saying; If we have accepted that understanding the effects of the transformation process on the probability density for shape variation is important, then not treating the probability density for measurement with equal rigour is inconsistent. As has been noted many times before, probability theory (and quantitative statistics) is the only self consistent theory for all data analysis. People should therefore not be too surprised when others try to interpret what they do in these terms.

In applying the principle of theory selection using generalisation (as described in the introduction) a distinction must be made between the information embodied in a data set due to its (biological) variability and the variability due to measurement noise. The former is what we should be able to predict with our theory and the latter is needed in order to assess the validity of any model description. Though both of these processes appear to cause broadening of observed distributions it can be a mistake to simply combine them and treat them as one combined distribution[2]. Inevitably this results in making it impossible to *uniquely* determine the best parametric description, ie: it becomes an ill-posed problem.

For the skull shape example, if we have several competing models for shape variation, we can choose to model the density distribution of data, but can not identify how many distinct factors cause measureable variations, or select between alternative models which describe them. In short;

Point 4: Transformed measurements that have errors which are not explicitly modelled cannot be used as the basis for model selection.

Although I would consider this issue to be of fundamental importance if we are aiming ultimately to build systems capable of shape recognition, some may regard this last point as minor, as it does not pertain to the problems that differential geometry is applied to. So let us consider now the specific clam that the constructs of Riemann geometry can be used to define meaningful defnitions of shape difference. There are some observations which need to be made regarding the idea of using the metric to define statistical similarity. Firstly, if there is a genuine one to one mapping between a measured input data space and the derived Riemann space then, if we apply the process of error propegation in order to construct the metric, any computed distances on the Riemann manifold would actually be the same as would be computed in the original data space (where for uniform measurement errors the shortest cost path is simply a straight line). This means that performing the calculation on the manifold would just complicate what would otherwise be a very simple calculation. Normally, statisticians use transforms to equal variance spaces specifically to avoid this problem. However, a representation (for a measurement of fixed error) which has scale invariance is a many to one mapping, not one to one. This prevents the tensor in the new space from ever being interpreted directly as a statistical distance. It is my belief that constructing a representation with pure scale invariance, though the Holy Grail for many workers in this area, is impossible under these circumstances. Transformations to achieve invariance need to be chosen with far more care if we are to build systems to solve tasks such as shape recognition. We have already seen how a good represenation may require the property of statistical independance (in order to avoid correlation), if we also demand homogemous propegated errors then we would have no use at all for differential geometry. It was this kind of reasoning which led our group to develop Pairwise Geometric Histograms (see this web site) a decade ago.

Although differential geometry allows us to begin to understand the important problem of changes in probability distributions, as a theory it does not guarantee statistically valid measures of similarity between measurements. Far from supporting the need for such a complex approach to shape analysis, these arguments would appear to imply differential geometry has great potential to be mis-used for constructing ad-hoc shape similarity measures. This is to be contrasted with the use of these techniques in general relativity, where they provide the only meaningful approach for measuring physical (non-statistical) distances. Unfortunately, theoretical legitimacy of a mathematical approach in one scientific area does not guarantee validity in another.

Point 5: The definition of a Riemann space with pure scale invariance prevents the direct use of the metric for calculation of a measurable difference in shape.

In the final chapters of the book the various methods of analysis presented are illustrated on examples. As examples of the utility of these approaches these are certainly good choices, but not necessarily for the reasons you might expect. At no point is any attempt made to apply differential geometry to a dataset such as the skulls in chapter 1.

One example is in the analysis of Ley lines, for the regularity in alignment of archaeological sites. To be fair to the author several limitations of the scientific validity of the null hypothesis as posed (that the regularities in alignment are consistent with random placements of monuments) are briefly mentioned. However, it might be reasonable to also have observed that any practical construction of man-made structure would probably have to take into account the local terrain. Any null hypothesis which assumes completely random construction is probably over simplistic [3]. The same argument applies to a later example, the analysis of post-holes. When we accept that the same questions could be answered much more easily by Monte-Carlo,

[2]Though this might seem an obvious point, unfortunately the multitude of ad-hoc methods for analysis which follow from this error have been the focus of many publications in high citation journals over the last few decades. Thus, there appears to be substantial evidence that this issue is easily overlooked.

[3]"the third one (castle) burnt down, fell over and then sank into the swamp" implies a successful castle construction in a swamp of about 1/4.

with unlimited scope for realistic modelling of a sampling process (castles in swamps etc.), the complex mathematics associated with construction of analytic functions looks clever but pointless.

Point 6: Hypotheses for the generation of any statistical data sets can generally be more realistically constructed and more easily tested using Monte-Carlo.

Another example is in the analysis of brooches. The book explains in some detail the process of determining a warp field between images, but gives little indication of how these warps are processed to generate a shape analysis. One observation is however, that if the process of manufacturing brooches is dependant upon various technologies which limit the minimum spatial size of a structure, then analysis of shape in a scale invariant manner would loose a very important aspect regarding the technology of fabrication.

Point 7: Practical examples of use appear to illustrate over simplistic models, driven by the available mathematics rather than a consideration of the scientific question.

Summary

Returning now to our initial question:

Is shape theory, and in particular the use of geometric approaches such as Riemann manifolds, a justifiable way to address scientific interpretation of data?

Though this topic appears to offer mathematicians who have developed skills in the field of set theory and differential geometry an opportunity to apply them, for the rest of us the various points made above would appear to caution against any suggestion that differential geometry is automatically the appropriate body of theory for the analysis of shape. In particular, a close focus on the mathematics of Riemann geometry may misguide people into believing that the theoretical basis for the analysis they perform is geometric rather than statistical. It would be wrong, for example, to say that use of PCA to model distributions is only justifiable as an approximation to a Riemann tangent space. Rather, PCA should be considered as a hyperplane fit to the data on the assumption of uniform independent localisation errors. Under the scientific method, the model can only be rejected if it does not account for the data up to the limits of measurement. If we adopt differential geometry as our underlying theory and ignore the statistical nature of this scientific test, the most interesting questions relating to data behaviour cannot be answered. For the skull example, this would correspond to not being able to identify either the minimum number of key variables or a specific theory which concisely accounted for the observed variations in shape. Those questions which can be addressed might have been more easily and more correctly addressed with Monte-Carlo.

It does not logically follow from the criticism of this book that all use of Riemannian manifolds in the context of shape spaces are rubbish. Some may wish to continue to use the formalism and terminology of differential geometry and set theory to describe their work. It may even be possible to bolt on extensions in order to correctly deal with some of the issues discussed above and this may give us shortcut solutions to some problems. Fundamentally however, it will always be the statistical interpretation (and in particular the need to treat our input data as measurements) which tells us if what is being done is correct. In order to do this, the proponents of this theory will have to explain clearly the assumptions necessary to apply Riemann geometry a as genuine statistical theory of shape. In particular we need to know how the 'metric' can be selected in accordance with conventional notions of statistical measurement.

References

1. C. G. Small, The Statistical Theory of Shape, Springer, 1996.
2. D.G. Kendall, et. al. Shape and Shape Theory, Wiley, 1999.

Comments from Colleagues

I have circulated this among various of my colleagues in order to get some feedback. I have made some consequent amendments, but this section covers opinions which I feel deserve an airing and my responses.

Chris Rose:

*You state that probability theory and statistics are *the* methods by which scientific hypotheses should be tested. I don't disagree with this. However, it's clear that many in the field either don't look at their work as being about the formation and testing of hypotheses, or don't agree that the way to approach such work is to treat it as data analysis. It might be worth briefly stating why probability and statistics are the right tools to use, because this then dictates the terms of reference of your review.*

In fact it is very difficult to summarise briefly the kind of historical philosophical considerations which surrounded the formation of modern science without misrepresenting the arguments and making the issue appear trivial. For a student level introduction of the basic position I refer you to Simon Coupes recent document regarding "Sound Scientific Principles" on these web pages. In the space available I can only say this, the quantities which most directly summarises our ability to interpret the value of any data in an experiment to test a theory are the conditional probabilities which relate the selected hypothesis and data.

You talk about the interpretation of PCA: it might be a good idea to clearly state that (if I've understood correctly) the Riemann geometric approach in the book is an approximation to a principled statistical approach (rather than PCA being an approximation to a principled Riemann geometric approach—i.e. the other way around).

I will when I believe that advocates of the use of Riemann geometry for shape analysis are attempting to achieve such an approximation. Currently, I don't have evidence for this, but see below.

You might also want to comment on the title of the book: is the author really presenting a statistical theory of shape, or a geometric one which neglects the statistical (i.e. imprecise) nature of measurement?

I would say this captures the substance of the main conclusions.

Chris Klingenberg:

Point 2: I think this is a problem related to the one about consistency of Procrustes estimates raised by Lele (1993) and answered by Kent and Mardia (1997), and briefly summarised by Dryden and Mardia (1998, section 12.1).

Ian Dryden's group has worked with weighted Procrustes methods of the sort you are talking about at the bottom of p. 3 (poster at LASR conference 2005). They have used a set of my fly wing data with that method and find the differences to the usual least-squares Procrustes methods vanishingly small (personal communication).

In my experience the measurement error is usually negligible relative to the variation in the data. So the influence of measurement error on the results will be small in practice. Moreover, because the total amount of variation is small in most biological data sets, the transformation from the raw coordinate data is pretty much the same regardless of the covariance structure (e.g., covariance-weighted methods differ very little from least-squares superimposition).

References:

Dryden, I. L., and K. V. Mardia. 1998. Statistical shape analysis. Wiley, Chichester.

Kent, J. T., and K. V. Mardia. 1997. Consistency of Procrustes estimators. Journal of the Royal Statistical Society B 59:281-290.

Lele, S. R. 1993. Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. Mathematical Geology 25:573-602.

I've no doubt that for the class of problems addressable there will be many instances where measurement error is small. However, the issue here isn't so much the absolute magnitude of these processes but the relative scaling as we move across the space. Unless this is done in a non-arbitrary manner there can be no suggestion that distances computed across the space are meaningful measurements of shape difference.

... you also seem to rely on conjecture: "I think it is possible to show that this is not the case." Further down, where you argue from the correlation of errors in Bookstein coordinates (the reason why Bookstein himself has mostly given up on Bookstein coordinates) to the general invalidity of Riemann shape spaces, the same thing applies. Here, what carries the weight of your argument from the one case to the general class is the phrase "but this misses the point".

I didn't think I was relying on conjecture, I've added a few more sentences to try to link the logical arguments together better, and what "point" was being missed. I don't know if what you say about Bookstein is true, but I have to say I have a lot of respect for his opinion and really would have expected him to notice the correlation problem given his other work. Do you think anyone in this field would ever publish a retraction under these circumstances? I don't expect so, and this is why I think we have to be extremely careful that we don't just accept anything published in computer vision on face value (complicated mathematics) or received wisdom (popularity). The original authors may ultimately be aware of the limitations of the techniques they publish, but the rest of us will generally be left unaware of them unless we figure it out ourselves.

I agree with much of your criticism of the present book, but I don't think it logically follows that all uses of Riemannian manifolds in the context of shape spaces are therefore rubbish. You are worried about the perception that "the theoretical *basis of the analysis they perform is geometric rather than statistical." I agree if your objection is against the "rather than", but I disagree if you object to the inclusion of geometric considerations. For the analysis of shape, these are clearly indispensable too. As far as I can see, most people in the business do consider both.*

I modified the conclusions slightly to move closer on the first point so as not to appear unreasonably dismissive. I will move closer to the position in the final sentence when I see a presentation which links the 'metric' to a measurement process (see Tina memo 2010-009). Currently I have no evidence that this proposition is accepted. On the contrary, the persisttant use of differential geometric terminology for the description of their work would seem to imply to me that they are unaware of the possibility of a statistical link. If someone were to tell me that they accept this, then I would immediately ask; How can the issue of varying information content as a function of scale be addressed in a scale invariant metric space? If I recieved a satisfactory answer then I would take the position suggested by Chris Rose (above), that Riemann geometry can be applied as an approximation to a principled statistical approach.

If this is the case and if this is at all obvious, I am very surprised that this argument has not been made very publicly. It is not as if there hadn't been anyone out there in the past 20 years.

I believe that in many research fields the audience is self selecting in such a way that many of the fundamental assumptions regarding how to approach problems are already agreed upon by those who attend the conferences and write papers. I have no idea of what those working in this area would think of my views, but perhaps they should be aware that these opinions might reflect the way others view this literature.