Visualization of time series statistical data by shape analysis (GDP ratio changes among Asia countries)

Yukari Shirota¹, Takako Hashimoto² and Riri Fitri Sari³

¹Gakushuin University, Tokyo, Japan
²Chiba University of Commerce, Chiba, Japan
³Universitas Indonesia, Depok, Indonesia

yukari.shirota@gakushuin.ac.jp

Abstract. It has been very significant to visualize time series big data. In the paper we shall discuss a new analysis method called “statistical shape analysis” or “geometry driven statistics” on time series statistical data in economics. In the paper, we analyse the agriculture, value added and industry, value added (percentage of GDP) changes from 2000 to 2010 in Asia. We handle the data as a set of landmarks on a two-dimensional image to see the deformation using the principal components. The point of the analysis method is the principal components of the given formation which are eigenvectors of its bending energy matrix. The local deformation can be expressed as the set of non-Affine transformations. The transformations give us information about the local differences between in 2000 and in 2010. Because the non-Affine transformation can be decomposed into a set of partial warps, we present the partial warps visually. The statistical shape analysis is widely used in biology but, in economics, no application can be found. In the paper, we investigate its potential to analyse the economic data.

1. Introduction

We would like to analyse a deformation. For a long time, it has been of great importance to measure shapes of objects such as shapes in morphology. In shape analysis, it is important how to compare different shapes; they have different sizes and orientations as well as different shapes. The process of transforming different sets of data into one coordinate system is called image registration by register marks. That is a tough problem and the image registration problem had been discussed for a long time. However, the research conducted by University of Leeds members and others made remarkable progresses in shape analysis to solve the problem[1]. The research field is called “statistical shape analysis” or “geometry driven statistics.”[2, 3]. The field of statistical shape analysis involves methods for the study of the shapes of objects where location, rotation and scale information can be removed to compare the shapes[4]. To compare the two shapes, we use landmarks which illustrate the corresponding pairs. The landmarks are translated, rotated, and scaled so that they lined up and matched as closely as possible. The procedure is known as a generalized Procrustes analysis[3]. The shape analysis applications include medical imaging and morphometrics for biology [5-11]. Although there are many papers on biology and medical imaging fields, as far as we know, there are no economics analysis paper by this statistical shape analysis except [12]. We would like to apply the statistical shape analysis methods for the economics analysis[13, 14]. In the paper, we analyse the GDP ratio changes from 2000 to 2010 in Asia between the agriculture and industry, value added (percentage of GDP). We handle the data as a set of landmarks on a two dimensional image to see the deformation using the principal
components. In the next section, we shall explain the GDP data we use. In Section 3, we present the principal warps of the given formation. In Section 4, concerning the deformation, the resulting partial warp data and others are visually presented. In Section 5, we shall discuss the potential and possibilities of the statistical shape analysis to analyse the economic data. Finally we conclude the paper in Section 6.

2. GDP Data
In the section, we shall explain the GDP data we use. The data are the GDP ratio changes from 2000 to 2010 in Asia between the agriculture and industry, value added (percentage of GDP) cited from World Bank statistical data (http://data.worldbank.org/indicator/NV.IND.TOTL.ZS). The value added of an industry is the contribution of a private industry or government sector to overall GDP. Value added equals the difference between an industry’s gross output (consisting of sales or receipts and other operating income, commodity taxes, and inventory change) and the cost of its intermediate inputs (including energy, raw materials, semi-finished goods, and services that are purchased from all sources) (https://www.bea.gov/faq/index.cfm?faq_id=184). Agriculture, value added (percentage of GDP) includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs (http://data.worldbank.org/indicator/NV.IND.TOTL.ZS). The target countries are ten countries: they are China, India, Indonesia, Thailand, Philippines, Malaysia, Singapore, Japan, Korea, and Pakistan. The data is presented in Figure 1. The individual country is denoted as one landmark on the formation. We handle the data as a set of landmarks on a two dimensional image. In Singapore, Japan and Malaysia, only the industry percentage decrease are seen. On the other hand, China and Indonesia, and India show only an agriculture decline tendency. To compare the two formations, we have to make landmarks translated, and re-scaled. The formation after the translation and scaling is called a “pre-shape.” The pre-shapes of Figure 1 is shown in Figure 2. The scaling makes the axis have no dimension. Therefore, in all figures from Figure 2 to 6, the axis has no dimension. The scaling factor for a pre-shape is the centroid size which is the square root of the sum of squared Euclidean distances from each landmark to the centroid [1].

![Figure 1](image_url)

**Figure 1.** Asian countries’ Agriculture and Industry, value added (percentage of GDP) in 2000 and 2010. The arrow illustrates each country’s difference from the 2000 position to the 2010 position.
We would like to analyse the differences separately by the Affine and non-Affine transformations. By global differences we mean large scale trends, such as overall Affine transformation. Local differences are on a smaller scale, for example highlighting changes in a few nearby landmarks. Global differences are smooth changes between the figures, whereas local changes are the remainder of the components of a deformation and are less smooth [1]. When we consider the meaning of “a few nearby landmarks”, for example, in skull data of animals, the meaning is the physically connection on the skull.

![Figure 2. Pre-shapes of Asian countries’ agriculture and industry, value added (percentage of GDP) in 2000 and 2010. The arrow illustrates each country’s difference from the 2000 position to the 2010 position. The two axes have no dimension because we obtained pre-shapes by scaling. In the scaling, the original Euclidian distance is divided by the Centroid size.](image)

On the other hand, in the GDP data, there is no geographical relationships among nearby countries; the deformation in ten years depends only on the country’s characteristics. In the next section, we introduce a thin-plate interpolation among landmarks. In a skull data, there is a real physical surface. In biology, we can conduct reasoning of the cause and effect relationship of the deformation. However, in the GDP data, there is no real thin-plate. There, we can see only the resulting changes; the causes of the change are considered from each country’s characteristics. The paramount issue of the economical shape analysis is the interpretation of the thin-plate. Concerning this, we have to study continuously.

### 3. Principal Warps

In the section, we will present the principal warps of the given pre-shape of the 2000 data which we represent by T. The principal warp is defined just on one shape. For example, given the shape T, the set of principal warps is defined. For the shape T, a bending energy matrix can be calculated. Then when we can do an eigen-decomposition of the matrix, the obtained eigenvectors are called principal warps. After an eigen-decomposition of the bending energy matrix, we can get (k-3) non-zero eigenvalues and eigenvectors. In this example, as there are 10 landmarks (countries), the number of the principal warps is seven. Figure 3 illustrates the first and second principal warps of T among the 10 principal warps. The spline surface is determined so that the bending energy is minimized. We know theoretically that the i-th partial warp is defined as an inner product between the i-th principal warp eigenvector of T and the basis function vector \( s(t) = (\sigma(t - t_1), ..., \sigma(t - t_k))^T \) where \( \sigma(t - t_k) \) is the interpolation base function of which centre is \( t_k \), the k-th landmark position.
The first principal warp shows the Philippine’s difference relative to Indonesia and Thailand (See Figure 4). The second principal warp reflects the difference of Indonesia, Japan and Korea relative to Thailand, Philippines, and Singapore (See Figure 5). The eigenvalues are as follows: 109, 76, 38, 28, 18, 11 and 3. That means that the first and second principal warps are dominant ones. Then let us consider the meaning of the principal warp. In general, when a matrix is given and we may get its eigenvectors, each eigenvector shows the principal characteristic of the matrix, which is the PCA (Principal Component Analysis). Like that, given the formation T, the principal warp means the principal characteristic of the shape T which is interpolated with the thin-plate.

4. Partial Warps of the Deformation
In the section, we analyse the deformation from the 2000 pre-shape (T) to the 2010 pre-shape (Y). As shown in Figure 5, The deformation is decomposed into the Affine transformation and the non-Affine transformation. The non-Affine transformation is decomposed into the seven partial warps. The number of partial warps is equal to the number of principal warp eigenvectors, which is the theory. In other words, to express the deformation T to Y, we use the principal components of the given T where the partial warp is a projection mapping of the pre-shape Y to the principal warp eigenvectors of T. In the example, because the number of principal warps was seven, the number of partial warps is also seven. The original pre-shape T and the shape after the Affine transformation of T are shown in Figure 5. The Cartesian transformation grids are called transformation grids. The Affine transformation is expressed the skewed transformation grids. The Affine transformation shows a global change tendency. In the example, we can see from the Affine transformation a globally positive correlation between agriculture and industry and growth of both. An industrious growth is a bit bigger than an agricultural one. We suppose that a thin-plate exists according to the transformation grids. The total non-Affine transformation is shown in Figure 5. The transformation grids are changed as if the thin-plate is folded.

Next, we will divide the non-Affine transformation to seven partial warp eigenvectors. That is a principal component of the deformation T to Y. The first, second and third partial warps are presented in Figure 6. In each partial warp figure pair, the right figure shows the change of the transformation grids. The grid change shows the local changes and is not smooth. The difference/change on each country can be expresses by a vector. A set of the vectors are plotted from the origin (0, 0) (See the oval area in the right graph). The left figure is a magnification one of the oval area in the right figure. In the left figure, instead of the circle marks without country names, country names are plotted around the origin (0, 0). The country name in the left figure is used there as a point marker. Theoretically the vector points are listed on a straight line as shown in the right and left figures. Although we see the decline angles of the two corresponding ovals are not the same, this is because the axis scaling is not the same. If we use the same axis scaling, the decline angles become the same.

In the first partial warp, Indonesia’s change direction is opposite to the direction of Philippines’ one (See the left figure). As shown in the right figure of the first partial warp, the transformation grids show the local changes in the neighbourhood around Philippines and Indonesia; there the transformation grid lines are skewed. Then in the second partial warp, there are a local change on Korea and Japan and another local change on India and Pakistan. The two changes are the opposite directional changes as shown in Figure 5. This change is mainly caused by the same tendency in the second principal warp of T as shown in Figure 4. Because the partial warp is the mapping of Y to the principal warp of T, the same tendency can be seen as the base principal warp.

In the third partial warp, India shows the opposite directional change to ones of Singapore and Malaysia. In Figure 6, the transformation grids by the third partial warp shows the big wave-like change in the vicinity of India; on the other hand, the vicinity of Singapore shows the opposite directional grid change.
5. Discussion
In the section, we shall consider the effects of the shape analysis method in economic data analysis. In this GDP example, finally we found that in the first partial warp Indonesia and Philippines offer the opposite directional changes, concerning the growth of industry percentage, although Indonesia and Philippines are located in the neighbourhood in the 2000 pre-shape. Secondly, the second principal local change is that Korea and Japan and India and Pakistan show the opposite directional changes concerning the growth of agricultural percentage. Like this, the statistical shape analysis method can describe the local changes of each country. We discussed, in Section 2, that economics data has no real thin-plate among landmarks. Although the real thin-plate does not exist among countries, the partial warps are helpful to see the local change of each country. As the aim of this paper is to demonstrate the shape analysis method on economics data, we cannot say the reason of each local change in detail. We can, however, say that the statistical shape analysis method has the power of decomposition of the complicated local changes. In other words, we can extract principal components of the complicated local movements of each country on the deformation.

The statistical shape analysis extracts principal components of the non-Affine transformation part using the bending energy matrix. Concerning the principal component related analysis, in economics data analysis, we mainly use PCA (Principal Component Analysis) and SVD (Singular Value Decomposition) (See [15]). For example, time series data such as stock prices are analysed by using Random Matrix Theory which uses SVD [16-18]. The purpose of the existing studies by PCA and SVD is to extract principal components for groping of objects such as companies and countries. However, we use the shape analysis to extract principal components of their local movements. Then we focus on each country’s peculiar movement. The advantage of the shape analysis is that we can examine the non-Affine transformation and that we can divide the non-Affine transformation to several principal components. Almost all data deformation analysis in economics have mainly examined the Affine transformation. However the shape analysis method enabled us to conduct another approach of analysing time series data analysis. The shape analysis method should be used more widely for the economics data analysis.

![Figure 3. The pre-shape of the 2000 data and its first and second principal warps.](image-url)
Figure 4. In each principal warp eigenvector of the 2000 pre-shape, each country has its value in the bar chart. The value indicates the contribution of the country to the thin-plate interpolation surface. The y axis shows a value of each country in the principal warp eigenvector and the y axis has no dimension.

Figure 5. The deformation is divided to Affine and non Affine transformations.
Figure 6. The first, second and third partial warps of the deformation from 2000 to 2010. Only the difference vectors of each countries are plotted in the oval areas around the origin (0, 0).
6. Conclusion
In the paper we explained the way of applying the statistical shape analysis method on economics data. We used as an example the agriculture and industry value added percentage of GDP data in Asia between in 2000 and in 2010. We analysed the deformation of the growth ratio among the Asian 10 countries. There a country is expressed as a landmark. A set of 10 landmarks forms a shape. To analyse the deformation between the two shapes, the deformation is decomposed into the Affine transformation and the non-Affine transformation. When we are not interested in the total change tendency and we are interested in each country’s peculiar movement, we would like to extract only the non-Affine transformation. The non-Affine transformation can be decomposed into a set of partial warps. Each partial warp features the character of the local changes. For example, the first partial warp shows the first principal component of the local change. In the data case, the first partial warp shows that Indonesia and Philippines show the opposite directional changes concerning the growth of industry percentage, although they are located in the neighbourhood in 2000.

Almost all data deformation analysis in economics have mainly examined the Affine transformation part. However the shape analysis method enabled us to conduct another approach of analysing time series data analysis. The shape analysis method should be used more widely for the economics data analysis. We would like to apply the shape analysis method to many kinds of economic data.

References
[1] I. L. Dryden and K. V. Mardia, *Statistical shape analysis* vol. 4: J. Wiley Chichester, 1998.
[2] I. L. Dryden and K. V. Mardia, *Statistical shape analysis with Applications in R (Second Edition):* J. Wiley Chichester, 2016.
[3] I. L. Dryden and J. T. Kent, *Geometry Driven Statistics: *Wiley Online Library, 2015.
[4] K. Mardia, F. Bookstein, and J. Kent, "Alcohol, babies and the death penalty: Saving lives by analysing the shape of the brain," *Significance,* vol. 10, pp. 12-16, 2013.
[5] H. Bayanati, R. E. Thornhill, C. A. Souza, V. Sethi-Virmani, A. Gupta, D. Maziak, *et al.,* "Quantitative CT texture and shape analysis: Can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer?," *European radiology,* vol. 25, pp. 480-487, 2015.
[6] F. Nemmi, U. Sabatini, O. Rascol, and P. Péran, "Parkinson's disease and local atrophy in subcortical nuclei: insight from shape analysis," *Neurobiology of aging,* vol. 36, pp. 424-433, 2015.
[7] M. Burgaleta, A. Sanjuán, N. Ventura-Campos, N. Sebastian-Galles, and C. Ávila, "Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis," *NeuroImage,* vol. 125, pp. 437-445, 2016.
[8] S. Srivastava, S. B. Lal, D. Mishra, U. Angadi, K. Chaturvedi, S. N. Rai, *et al.,* "An efficient algorithm for protein structure comparison using elastic shape analysis," *Algorithms for Molecular Biology,* vol. 11, p. 27, 2016.
[9] F. L. Bookstein, *The measurement of biological shape and shape change* vol. 24: Springer Science & Business Media, 2013.
[10] A. McIntosh, F. Bookstein, J. V. Haxby, and C. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage,* vol. 3, pp. 143-157, 1996.
[11] F. L. Bookstein, "Integration, disintegration, and self-similarity: characterizing the scales of shape variation in landmark data," *Evolutionary biology,* vol. 42, pp. 395-426, 2015.
[12] A. K. Laha, "Big Data Analytics in Management: Some Statistical Challenges and Opportunities," in *The 33rd Leeds Annual Statistical Research Workshop,* pp. 107-108.
[13] Y. Shirota and T. Hashimoto, "Visual Explanation of Deformation Theories in Shape Analysis," *Gakushuin Economics Papers,* vol. 54, p. in printing, 2017.
[14] C. Apriono, R. F. Sari, Y. Yano, and Y. Shirota, "Economic Indicator Evaluation Based on Shape Deformation Analysis of Indonesian Provinces Statistics," *Gakushuin Economics Papers*, vol. 54, pp. 1-22, 2017.

[15] J. Friedman, R. Tibshirani, and T. Hastie, "The elements of statistical learning: data mining, inference, and prediction," Springer, 2013.

[16] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "A random matrix theory approach to financial cross-correlations," *Physica A: Statistical Mechanics and its Applications*, vol. 287, pp. 374-382, 12/1/2000.

[17] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Physical Review E*, vol. 65, p. 066126, 06/27/2002.

[18] Y. Yano and Y. Shirota, "SVD and Text Mining Integrated Approach to Measure Effects of Disasters on Japanese Economics: Effects of the Thai Flooding in 2011" in *Neural Information Processing (LNCS 9949)*, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, and D. Liu, Eds., ed: Springer, 2016, pp. 20--29.