Pak. J. Statist.
2013 Vol. 29(1), 129-138

STATISTICAL SHAPE ANALYSIS FOR THE CLASSIFICATION OF RENAL TUMORS AFFECTING CHILDREN

Stefan Markus Giebel[1], Jang Schiltz[1] and **Jens-Peter Schenk[2]**
[1] University of Luxemburg, Luxemburg. Email: stefan.giebel@gmx.de
[2] University Hospital of Heidelberg, Div. of Pediatric Radiology, Germany

ABSTRACT

In this research project, we describe an application of statistical shape analysis. In order to differentiate the various kidney tumours appearing in childhood we use shape analysis on two-dimensional magnetic resonance images (MRI). We show that this mathematical procedure can be an interesting tool to assist the radiologist who is required to make a decision based on their intuition and their experience in lack of specific tumour characteristics. This study is the first one using MR images in oncology for statistical shape analysis. Our method is innovative in the way to find suitable landmarks and to test the differences, even if the sample size is small. In order to test the mean shape, the statistical test of Ziezold is used.

KEYWORDS

Statistical shape analysis; two-dimensional objects; renal tumors; mean shape; landmarks.

1. INTRODUCTION

In a wide variety of disciplines it is of great practical importance to measure, to describe and compare the shapes of objects. In general terms, the shape of an object, data set, or image can be defined as the total of all information that is invariant under translations, rotation and isotropic rescaling. The field of shape analysis involves hence methods for the study of the shape of objects where location, rotation and scale can be removed. The two- or more dimensional objects are summarized according to key points called landmarks. This approach provides an objective methodology for classification whereas even today in many applications the decision for classifying according to the appearance seems at most intuitive.

Statistical shape analysis is concerned with methodology for analyzing shapes in the presence of randomness. It is a mathematical procedure to get the information of two- or more dimensional objects with a possible correction of size and position of the object. So objects with different size and/or position can be compared with each other and classified. Procedures are used in some metric space to get the shape of an object without information about position and size, centralisation and standardisation.

Interest in shape analysis began in 1977. D.G. Kendall (1977) published a note in which he introduced a new representation of shapes as elements of complex projective

spaces. K.V. Mardia (1977) on the other hand investigated the distribution of the shapes of triangles generated by certain point processes, and in particular considered whether towns in a plain are spread regularly with equal distances between neighbouring towns. The full details of this elegant theory which contains interesting areas of research for both probabilists and statisticians where published by D. Kendall (1984) and F. Bookstein (1986). The details of the theory and further developments can be found in the textbooks by C.G. Small (1996) and I.L. Dryden & K.V. Mardia (1998). In this paper, we describe one interesting application of statistical shape analysis on renal tumors.

2. RENAL TUMORS IN EARLY CHILDHOOD

In the special case of oncology there is no theoretical medical reason to select a specific group of landmarks for differentiation. All landmarks in this research have thus to be selected by an explorative procedure.

Nephroblastoma (Wilms' tumor) (Wilms, 1889) is the typical tumor of the kidneys appearing in childhood. Therapy is organized in therapy-optimizing studies of the International Society of Paediatric Oncology and Haematology (SIOP) in Europe. Indication of preoperative chemotherapy is based on radiological findings. The preferred radiological method is sonography and MRI. Both methods avoid radiation exposure, which is of great importance in childhood. Preoperative chemotherapy is performed without prior biopsy (Schenk 2006, Furtwängler 2005).

Information of the images of magnetic resonance tomography, especially the renal origin of a tumor and the mass effect with displacement of other organs, is needed for diagnosis. Next to nephroblastomas other tumors of the retroperitoneum exist, which are difficult to differentiate (Schenk, 2008). Renal tumors in childhood are classified in three histological risk groups of malignancy (low, intermediate, high). Typical Wilms' tumors mostly belong to intermediate histological risk group, as in our study. In intermediate risk group different subtypes of nephroblastoma tissue exist (Graf 2003).

In our sample of tumors in childhood, there are four different types of tumors: Nephroblastoma, neuroblastoma, clear cell sarcoma and renal cell carcinoma. Renal cell carcinomas are very rare in childhood. They represent the typical tumors of adult patients. In nephroblastoma preoperative chemotherapy (SIOP-protocol) is indicated, in contrast this preoperative chemotherapy is not indicated in renal cell carcinoma. Clear cell sarcomas are very rare in childhood and are characterized by high malignancy. Neuroblastoma is the main differential diagnosis to nephroblastoma in the retroperitoneum. It is the typical tumor of the sympathetic nervous system and suprarenal glands. Infiltration of the kidney is possible. The tumor grows with encasement of vessels. Due to the high importance of radiological diagnosis for therapy, it is of great interest to find markers for a good differentiation of tumors.

For the application we have to consider, that nephroblastoma (Wilms' tumor) is the typical renal tumor in childhood and the number of cases of non-Wilms'-tumors are always small, except of neuroblastoma.

3. SAMPLE

The research sample consists of both frontal and transversal images of tumors. We studied 24 cases of tumors in frontal perspective (18 nephroblastoma, 3 neuroblastoma, 2 clear cell carcinoma and 1 renal cell carcinoma) and 15 tumors in transversal perspective (8 nephroblastoma, 3 neuroblastoma, 2 clear cell carcinoma and 2 renal cell carcinoma).

Because MRI has no radiation exposure it is the preferred radiological method. Using the images in frontal and transversal direction two three dimensional objects are constructed. Then the mass point of the three dimensional object is calculated for every tumor to make them comparable. The renal/retroperitoneal tumors distinguish themselves by their position in the three dimensional body. Since we can use only one two-dimensional image of all the existing images of a given patient we have to assure that a similar image can be found in the data of all the other patients.

The two dimensional image nearest to the mass point is used. Then 24 landmarks are chosen on the border of the tumor by keeping constant the angle between two following landmarks. Every landmark is thus a cut point between the surface of the tumor and the rest of the body.

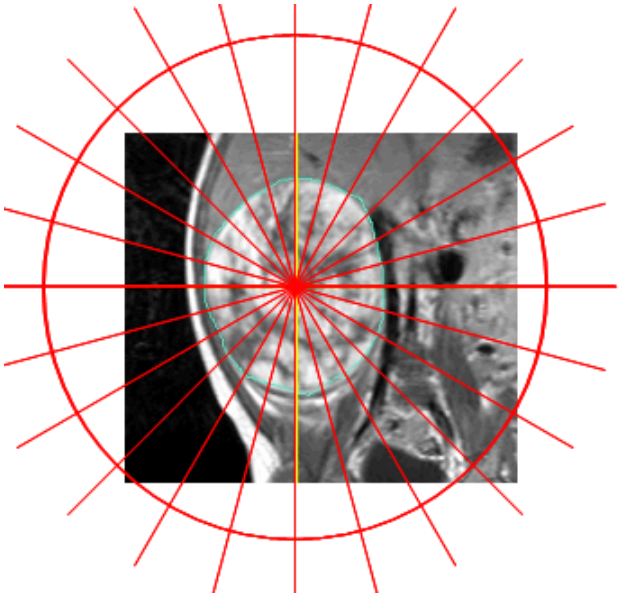

Fig. 1: Tumor of the right kidney and the 24 landmarks taken as cut points in coronal MRI scan.

By using the statistical shape analysis, the object is hence reduced to two dimensions, standardised and centered. Denote the number of landmarks by k. Every object o_i in a space V of dimension m is thus represented in a space of dimension $k \times m$ by a set of landmarks:

$$\forall i = 1..n, o_i = \{l_1..l_k\}, l_j \in \mathbb{R}^m$$

In our study the dimension of the space is 2 and the number of landmarks is 24.

First the Euclidean norm of the object is computed and the landmarks are standardised to allow comparability:

- For every $i, i = 1,...,k$ **,** the size of each object is determined by the computation of its euclidian norm.

$$\|o_i\| = \sqrt{\sum_{j=1}^{m} \left| l_j^i \right|^2} \tag{1}$$

- The landmarks are standardized by dividing them by the size of the object. j is the index of a landmark of i -th object:

$$l^0 = \frac{l_j^i}{\|o_i\|} \tag{2}$$

Then the landmarked are centered by the following procedure

- For every $i, i = 1,...,k,$ we compute the arithmetic mean z^i of the k landmarks each object:

$$z^i = \frac{1}{k} \sum_{j=1}^{k} l_j^i \tag{3}$$

- We center all the landmarks of each object by substracting this mean:

$$l^* = l_j^i - z^i \tag{4}$$

In our case we have only to deal with two dimensional centers. Because of the survey our original data are centered on the three dimensional mass point.

4. COMPUTING THE MEAN SHAPE

After the objects are centred and standardised the mean shape of nephroblastoma is calculated and also the distance of all nephroblastomas to their mean shape, a representative object of all objects. The "mean shape" should have the smallest distance to all other objects. We are using the Hermité inner product in the following algorithm. The "mean shape" is calculated by the algorithm of Ziezold (1994):

- For $i = 1,...,\ n$

$$\tilde{m} \mapsto w_i(\tilde{m}) = \begin{cases} \frac{\langle \tilde{m}, o_i \rangle}{|\langle \tilde{m}, o_i \rangle|} & \text{if } \langle \tilde{m}, o_i \rangle \neq 0 \\ 1 & \text{if } \langle \tilde{m}, o_i \rangle = 0 \end{cases} \tag{5}$$

- $\tilde{m} \mapsto T(\tilde{m}) = \frac{1}{n} \sum_{i=1}^{n} w_i(\tilde{m}) o_i$ (6)

- We then compute recursively,

$$\tilde{m} = T\left(\tilde{m}_{r-1}\right), r = 1,2,... \tag{7}$$

- till the following stopping rule is fulfilled:

$$\tilde{m} = T\left(\tilde{m}\right) \tag{8}$$

The advantage of the algorithm of Ziezold is, that it allows in contrast to Procrustes analysis to compute the mean shape very quickly.

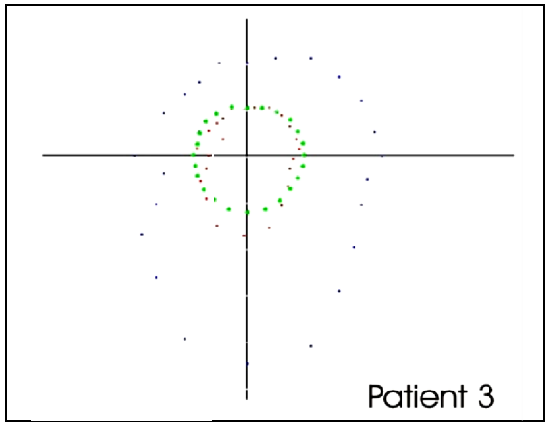

Fig. 2: Patient No. 3: green: the mean shape, red: the centered and normed data and blue: original data

As in figure 2 can be seen, patient No. 3 is quite far away from the "mean shape".

5. RESULTS OF DISTANCES

We rank all tumors according to their distance to the mean shape such that a high rank means a high distance to the mean shape.

Table 1: Distances d_f of the nephroblastomas (Wilms' tumors) to the mean shape in frontal perspective

Patient No.	$d_{f:}$	Rank
1	0.0849	3
2	0.1009	6
3	0.2260	18
4	0.0968	5
5	0.1567	13
6	0.1113	8
7	0.1940	17
8	0.1448	12
9	0.1854	16
10	0.1290	11
11	0.1834	15
12	0.0772	2
13	0.0916	4
14	0.1058	7
15	0.1126	9
16	0.0541	1
17	0.1178	10
18	0.1754	14

In Dryden & Mardia it is mentioned that the Euclidean distance between the fitted objects used here is the partial Procrustean distance. The arithmetic mean of the distance in the frontal perspective is $\overline{d} = 0.1304$

Table 2: Distances d_t of the nephroblastomas (Wilms' tumors) to the mean shape in transversal perspective

Patient No.	$d_{f:}$	Rank
8	0.0998	3
12	0.0966	2
15	0.0772	1
6	0.1164	4
5	0.1791	8
4	0.1752	7
2	0.1112	3
9	0.1414	5

The arithmetic mean of the distance in the transversal perspective is $\overline{d} = 0.1239$.

Table 3: Distances of the nephroblastoma to the mean shape in frontal perspective

Patient No.	$d_{all} = d_f + d_t$	Rank
8	0.2446	5
12	0.1734	1
15	0.1888	2
6	0.2277	4
5	0.3358	8
4	0.2734	6
2	0.2121.	3
9	0.3268	7

The arithmetic mean of the distance is $\bar{d} = 0.248$.

6. DIFFERENTIATION

For testing the mean shape used for the differentiation of the different types of tumors we use the test of Ziezold (1994).

Test of Ziezold (1994):

- **1. Step: Definition of the set of objects**

  There is one set $M = \{o_1, ..., o_N\}$ that can be divided into two subsets: objects with the characteristics A: $A^{sample} = \{o_1, ..., o_n\} = \{a_1, ..., a_n\}$ and objects with the characteristics B: $B^{sample} = \{o_{n+1}, ..., o_N\} = \{b_{n+1}, ..., b_N\}$. The subset A is a realisation of a distribution P and the subset B is an independent realisation of a distribution Q.

  Hypothesis: $H_0 : P - Q$
  Alternative: $H_1 : P \neq Q$

  Define the level of significance α. If the probability for H_0 is smaller, we neglect H_0 and assume H_1.

- **2. Step: Computing the mean shape**

  The mean shape is calculated by means of the algorithm of Ziezold (1994). Let m_0 denote the mean shape of the subset A.

- **3. Step: Computing the u -value**

$$u_0 = \sum_{j=1}^{n} card\left(b_k : d\left(b_k, m_0\right) < d\left(a_j, m_0\right)\right)$$

- **4. Step: Determination of all the possibilities of dividing the set into two subset with the same proportion**

- **5. Step: Comparing the u_0-value to all possible u-values. Computing the rank (small u-value mean a small rank).**
- **6. Step: Calculate the p-value for H_0:**

$$p_{r=i} = \frac{r}{\binom{N}{n}} \text{ for } i = 1,...,\binom{N}{n},$$

where r is the rank for which we assume a rectangular distribution.

The same results are thereafter obtained also for the subset B.

Table 4: Distances of Non-Wilms'tumors to Wilms 'tumors

sets		Types					
Nr.	kind of tumor	d_f	$rank_{19}$	d_t	$rank_9$	$d_f + d_t$	$rank_9$
No.19	clear cell sarcoma	0.2324	19	0.1182	6	0.3506	9
No.20	clear cell sarcoma	0.1605	15	0.1356	6	0.2961	7
No.21	neuroblastoma	0.2969	19	0.1275	6	0.4244	9
No.22	neuroblastoma	0.1775	16	0.4151	9	0.5926	9
No.23	renal cell carcinoma	0.1082	8	0.0834	2	0.1916	9
No.24	neuroblastoma	0.1227	11	0.1793	9	0.3020	7
No.25	renal cell carcinoma	n.a.	n.a.	0.0822	2	n.a.	n.a.

n.a.: no suitable data avaliable (low quality)
rank of one Non-Wilms' tumor in the group of Wilms' tumor

It can be shown that the renal cell carcinoma are very difficult to differentiate them from Wilms' tumors. They are very near to the mean shape of Wilms' tumors. Neuroblastoma and clear cell carcinoma on the other hand have a greater distance to the mean shape of Wilms-tumors than renal cell carcinoma. For proving the statement we use the test of Ziezold (1994).

Due to the number of suitable cases in transversal perspective in our sample is very small we use only the frontal perspective. For the frontal perspective, we get the following result:

Table 5: Result with / without renal cell carcinoma

sets		Differentiation						
kind of tumor 1	kind of tumor 2	u_0	$m_=$	$m_<$	$p - Intervall$	k	n	$\binom{n}{k}$
Wilms	"Non-Wilms"	28	1345	9375	[0.070, 0.080]	18	24	134596
"Non-Wilms"	Wilms	54	2345	105663	[0.785, 0.802]	6	24	134596
Wilms	N1+K	17	210	921	[0.027, 0.034]	23	18	33649
N1+ K	Wilms	43	704	26118	[0.776, 0.797]	23	5	33649

$N1$: neuroblastoma; K: clear cell sarcoma; $m_=$: number of permutations with the same u-value; $m_<$: number of cases with a smaller u-value

Thus, the kind of tumors can be only be differentiated in the direction of the mean shape of the Wilms' tumors. If we want check the usefulness of the differentiation

procedure after standardisation and centering, we have to look at different possible procedures.

Table 7: Differentiation of Wilms' tumors to Non-Wilms' tumors in frontal perspective

Characteristics		Differentiation of the kind of tumors						
Property	Type of data	u_0	$m_=$	$m_<$	$p-Intervall$	k	n	$\binom{n}{k}$
shape I	"2D centred" + "standardised"	28	1345	9375	$[0.070, 0.080]$	18	24	134596
shape II	"3D centred" + "standardised"	37	2575	27506	$[0.2044, 0.2235]$	18	24	134596
figure	"2D centred"	43	2971	37787	$[0.2808, 0.3028]$	18	24	134596
original	"3D centred"	45	3125	44495	$[0.3306, 0.3538]$	18	24	134596

Since the u_0-value and the p-value are smallest in the case of two dimensional centered and standardized images we can conclude that our procedure is useful for differentiation. It should be remarked, that the group of Non-Wilms' tumors is heterogeneous. Hence the differentiation for every kind of tumors is of further interest.

Table 8: Differentiation of all kind of tumors from each other

sets		Differentiation						
kind of tumor 1	kind of tumor 2	u_0	$m_=$	$m_<$	$p-Intervall$	k	n	$\binom{n}{k}$
Wilms	N1	12	47	122	$[0.0924, 0.1271]$	3	21	1330
N1	Wilms	15	36	834	$[0.6271, 0.6541]$	18	21	1330
Wilms	K	5	4	13	$[0.0737, 0.0895]$	2	20	190
K	Wilms	0	103	0	$[0.0053, 0.5421]$	18	20	190
Wilms	N2	11	3	11	$[0.6667, 0.7778]$	18	19	18
K	N1	0	7	0	$[0.1, 0.7]$	2	5	10
N1	K	1	2	5	$[0.6, 0.7]$	3	5	10
K	N2	0	3	0	$[0.3333, 1]$	2	3	3
N1	N2	1	2	1	$[0.5, 0.75]$	3	4	4

$N1$: neuroblastoma; $N2$: renal cell carcinoma; K: clear cell sarcoma; $m_=$: number of permutations with the same u-value; $m_<$: number of cases with a smaller u-value

According to table 8, clear cell sarcoma can be differentiated in both direction and Wilms' tumors can be differentiated from neuroblastoma.

CONCLUSION

We have shown that shape analysis of two dimensional data can be used for medical decision processes. It is an objective tool to differentiate the renal tumors appearing in early childhood. For more precise results, we need more data to analyze, especially in the case of renal cell carcinoma, neuroblastoma and clear cell sarcoma. In the view of application our first results have to be shown on more data. The sample used here is about 20% of the data for one year in Germany. Moreover, in further research we will analyze three dimensional landmarks of tumors to see, if we get the similar results for this type of sample. Shape analysis can be a solution in decision processes on two

dimensional data. Especially in medicine shape analysis is an objective tool to differentiate objects not in an intuitive way.

In further research three dimensional landmarks of tumours are necessary. The three dimensional landmarks are taken as a cut point between the surface of the tumour and the line between the edge of the platonic object and the mass point.

REFERENCES

1. Anderson, C.R. (1997). Object recognition using statistical shape analysis, PhD thesis, University of Leeds.
2. Bookstein, F.L. (1986). Size and shape spaces for landmark data in two dimensions. *Statistic Sciences*, 1, 181-242.
3. Bookstein, F.L. (1996). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*. 58, 313-365.
4. Coppes, M.J., Campbell C.E. and Williams, B.R.G. (1995). Wilms Tumor: Clinical and Molecular Characterization, Austin Texas USA: RG Landes Company, 1-55.
5. Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.
6. Fischer M., Oberthür A., Von Schweinitz D. and Simon T. (2005). Das Neuroblastom. *Der Onkologe*, 11, 1054-1064.
7. Furtwängler, R., Schenk, J.P., Reinhard, H., Leuschner, I., Rübe, C., von Schweinitz, D. and Graf, N. (2005). Nephroblastom- Wilms- Tumor. *Der Onkologe*, 11, 1077-1089.
8. Graf, N. and Reinhard, H. (2003). Wilms-Tumoren. Diagnostik und Therapie. *Der Urologe* [A], 42, 391-409.
9. Kendall, D.G. (1977). The diffusion of shape. *Adv. Appl. Probab.*, 9, 428-430.
10. Kendall, D.G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16, 81-121.
11. Kendall D.G. and Kendall W.S. (1980). Alignment in two dimensional random sets of points. *Advances in Applied probability*, 12, 380-424.
12. Mardia, K.V. (1977). Mahalanobis distance and angles. In: Krishnaiah, P.R. (ed) *Multivariate Analysis IV*, 495-511, Amsterdam: North Holland
13. Schenk, J.P., Schrader, C., Zieger, B., Furtwängler, R., Leuschner, I., Ley, S., Graf, N. and Troeger, J. (2006). Reference radiology in nephroblastoma: accuracy and relevance for preoperative chemotherapy. *Fortschr. Röntgenstr*. 178, 38-45.
14. Schenk, J.P., Graf, N., Günther, P., Ley, S., Göppl, M., Kulozik, A., Rohrschneider, W.K. and Tröger, J. (2008). Role of MRI in the management of patients with nephroblastoma. Eur Radiol., 18, 683-91.
15. Small, C.G. (1889). *The Statistical Theory of Shape*. Springer-Verlag, New York.
16. Wilms, M. (1889). Die Mischgeschwülste der Niere, Verlag von Arthur Georgi, Leipzig; 1-90.
17. Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces, *Trans. 7th Prague Conference Inf. Th., Statist.*, Dec. Funct., Random Processes, Vol. A. Reidel, Dordrecht, Prag, 591-602.
18. Ziezold, H. (1994). Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane. *Biometrical Journal*, 36, 491-510.