Bulletin of the *Transilvania* University of Braşov • Vol 2(51) - 2009
Series III: Mathematics, Informatics, Physics, 59-74

APPLYING SHAPE ANALYSIS IN MEDICINE AND ENGINEERING

Stefan Markus GIEBEL [1] **, Jang SCHILTZ** [1]
F.-B. FRECHEN [2] **, Wolfram FRANKE** [2]
and Jens-Peter SCHENK[3]

Communicated to:
9-ème Colloque franco-roumain de math. appl., 28 août-2 sept. 2008, Braşov, Romania

Abstract

We describe two applications of statistical shape analysis. For the differentiation of different types of retroperitoneal tumours appearing in early childhood we use shape analysis of two-dimensional MRI images. For the analysis of odours in biogas plants, we use shape analysis to interpret the data provided by electronic noses.

1 Introduction

In a wide variety of disciplines it is of great practical importance to measure, 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 translation, rotation and isotropic rescaling. The two- or more dimensional objects are characterized by 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 extract the information of two- or more dimensional objects after a standardization of their size and position. So objects with different size and/or position can be compared with each other and classified. To get the shape of an object without information about position and size, centralization and standardization procedures are used in some metric space.

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

[1]University of Luxembourg
[2]University of Kassel
[3]Pediatric Radiology, University of Heidelberg

of triangles generated by certain point processes, and in particular considered whether towns in a plain are spread regurlary with equal distances between neighbouring towns. The full details of this elegant theory which contains interesting areas of research for both probabilists and statisticains 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). We present two applications of statistical shape analysis, the classification of retroperitoeneal tumours of small children and the detection of odour.

2 The mean shape

Lets suppose that we have defined a statistical model for our population in some metrical space. The mean shape is then the shape with the smallest variance of all shapes in a group of objects.

If there are only two objects the "mean shape" is the average of the objects (Ziezold 1974).

In all other cases we use the algorithm of Ziezold (1994) to determine the mean shape. The advantage of that algorithm is that it allows to find the mean shape very fast.

Suppose that we search the mean shape m of n objects o_i, $i = 1, \ldots, n$. The norm is the Euclidean norm. The algorithm of Ziezold(1994) for determining the "mean shape" can then be described as follows:

Using the following algorithm

$$\dot{m} \mapsto w_i(\dot{m}) = \begin{cases} \frac{(\dot{m},o_i)}{|(\dot{m},o_i)|} & \text{if } (\dot{m}, o_i) \neq 0 \\ 1 & \text{if } (\dot{m}, o_i) = 0 \end{cases}$$

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

recursively there is a sequence

$$\dot{m}_r = T(\dot{m}_{r-1}),\ r = 1, 2, \ldots \text{ iterations}$$

criterion to stop

$$\dot{m} = t(\dot{m})$$

If there is no change in the algorithm for computing the mean shape, the algorithm stops. There are changes, if the mean shape is not in the position of the smallest variance to all shapes.

One example for determining the "mean shape" is shown in figure 1. The green triangle is the mean shape of the group of three triangles (yellow, red and blue).

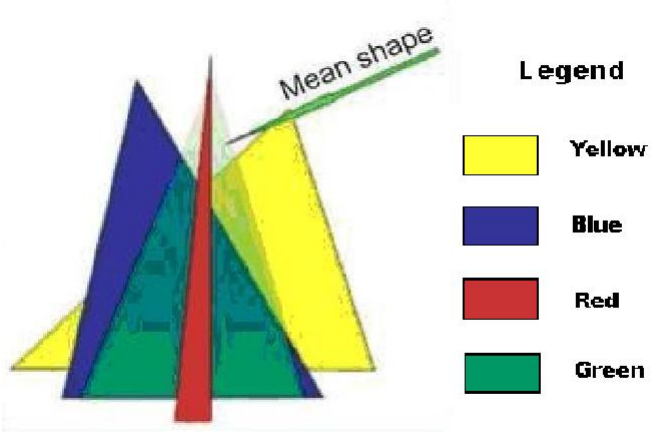

Figure 1: Determining the mean shape (green triangle) in a group of three triangles (yellow, red and blue)

The green one is in the optimal position to all three shapes. So it can be said, that the green triangle represented the group of the three objects.

3 The test of Ziezold (1994)

A mean shape is only of interest for applications in a discipline, if there is a possibility to prove, that it is characteristic for the objects in this group. For the distance of all objects to the mean shape we get two distributions: One distribution for the objects belonging to the group represented by the mean shape and a second distribution for all other objects. Ziezold proposes a test for the mean shape:

Suppose that one group of m objects is an independent realisation of the distribution P and the other group of k objects an independent realisation of the distribution Q.

We compute a p-value according to the test

H_0: $P = Q$

"There is no difference in the two distributions"

H_1: $P <> Q$

"There is a difference between the distributions P and Q"

1. **step**: Determine the mean shape
2. **step**: Compute distances to the mean shape and the u_0 according to the Mann Whitney-U-Test
3. **step**: Find all possible u-values (N possibilities) separating the group $(k + m)$ in two groups with m and k objects
4. **step**: Determine the rank r of u_0 in the group of all u-values
5. **step**: p-value $= r/N$
6. **step**: Compute the p-values in the other direction. Determine mean shape in the group of m objects. Reference is the mean shape of the second group with m objects.

4 Selection of the landmarks

Another point of view is to take only the landmarks that are relevant for differentiation. In many disciplines there is no theoretical background, why a landmark should be more relevant than another. Giebel (2007) proposes the following method for the selection of relevant landmarks:

Suppose that there are two groups of objects and that every object is described by n landmarks. We want to use only k landmarks to differentiate the two groups. We have to compute all possibilities for choosing k landmarks among the existing n.

1. **step**: Select k landmarks
2. **step**: Use the test of Ziezold
3. **step**: Choose the k-configuration with the smallest u_0 or with the smallest p-value
4. **step**: Compare the results with a random process

5 Application of shape analysis in medicine: Classification of renal tumours in 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 tumour) is the typical tumour of the kidneys appearing in childhood. Therapy is organized in therapy-optimizing studies of the Society of Pediatric Oncology and Hematology (SIOP). Indication of preoperative chemotherapy is based on radiological findings. The preferred radiologic 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).

Information of the images of magnetic resonance tomography, especially the renal origin of a tumour and the mass effect with displacement of other organs, is needed for diagnosis. Next to nephroblastomas other tumours of the retroperitoneum exist, which are difficult to differentiate (Schenk, 2008).

Renal tumours in childhood are classified in three stages of malignancy (I, II, III). Typical Wilms tumours mostly belong in stage II. In stage II different subtypes of nephroblastoma tissue exist (Graf 2003).

In our sample of tumours in childhood, there are four different types of tumours: nephroblastoma, neuroblastoma, clear cell carcinoma, and renal cell carcinoma. Renal cell carcinomas are very rare in childhood. They represent the typical tumours of adult patients. They have no high sensitivity for chemotherapy. Clear cell sarcomas are very rare in childhood and are characterized by high malignancy. Neuroblastoma is the main differential diagnosis to nephroblastoma. It is the typical tumour of the sympathetic nervous system and suprarenal glands. Infiltration of the kidney is possible. The tumour grows with encasement of vessels. Because of the high importance of radiological diagnosis for therapy, it is of great interest to find markers for a good differentiation of these tumours.

5.1 Sample

The research sample consists of 24 cases of tumours: 18 nephroblastoma, 3 neuroblastoma, 2 clear cell carcinoma and 1 renal cell carcinoma. Main diagnostic tools are sonography, CT and MRI.

Because MRI has no radiation exposure it is the preferred radiological method.

Before therapy, MRI images are sent to the radiological reference center for renal tumours in childhhood in Heidelberg.

These MRI examinations documented as DICOM files are used for shape analysis in our study.

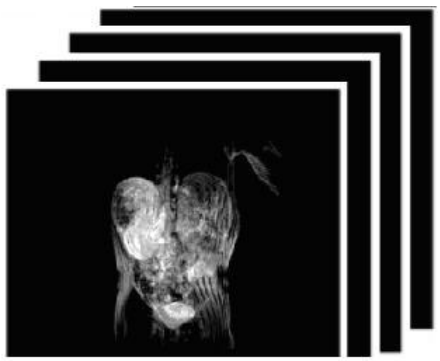

Figure 2: Images from MRI

Using the images in frontal direction a three dimensional object is constructed. One of the objects is shown in figure 3.

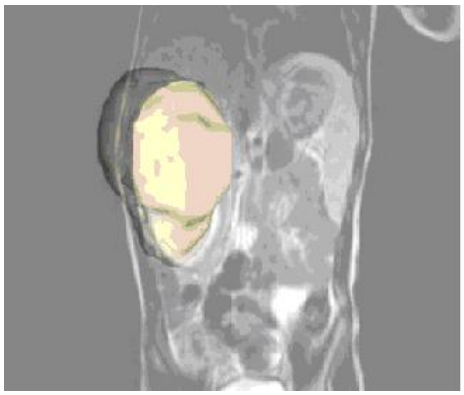

Figure 3:

Then the mass point of the three dimensional object is calculated for every tumour to make them comparable. The real tumours distinguish itself by their position in the three dimensional body. Since we can use only one two-dimensional image of all the existing images of a patient we have to assure that a similar image can be found in the data of the other patients.

In Figure 4 the mass point is in the centre. Around the centre 24 landmarks are taken.

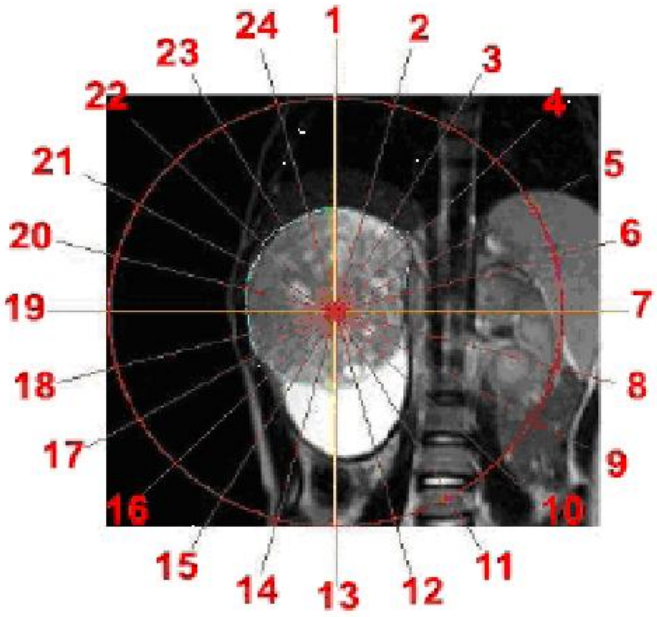

Figure 4: A similar image for every patient: The image as close as possible to the mass point and around the mass point 24 landmarks.

By the use of statistical shape analysis, the object is hence reduced to two dimensions, standardised and centred.

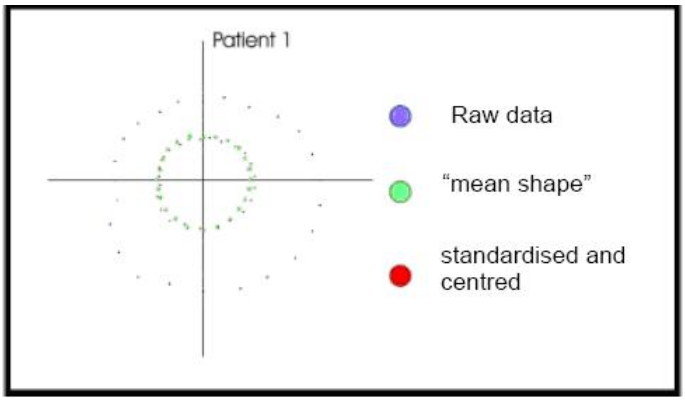

Figure 5: The raw data, the standardised and centred shape of ”Patient No. 1” and the ”mean shape” of ”Wilms-tumours” as a representative shape for all ”Wilms-tumours”

After all tumours are standardised and centred the tumours different in size and location can be compared to each other.

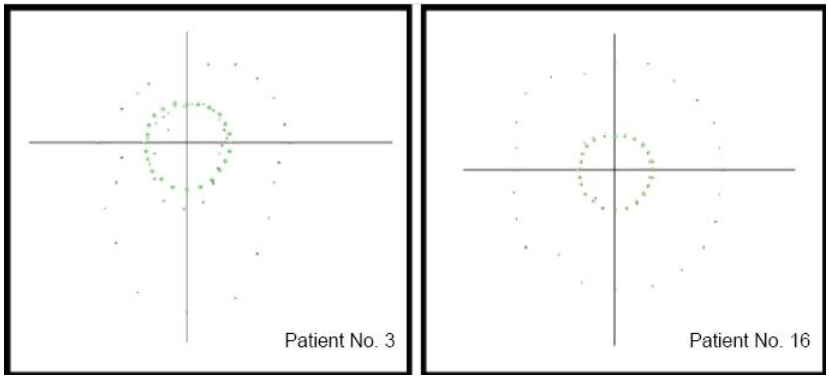

Figure 6: The original data of a patient with nephroblastoma, the shape and the mean shape of nephroblastoma

Optically the Patient No. 3 is very far away from the mean shape of nephroblastoma and the Patient No.16 is as close as possible to the mean shape. The overview of all 18 nephroblastoma is presented in figure 7:

patient		distance	
Nr.	Diagnose	d_f	$rangw$
Nr.1	n.b.	0.0849	3
Nr.2	IId	0.1009	6
Nr.3	IIc	0.2260	18
Nr.4	IIIa	0.0968	5
Nr.5	IIa	0.1567	13
Nr.6	IIb	0.1113	8
Nr.7	IId	0.1940	17
Nr.8	IId	0.1448	12
Nr.9	IId	0.1854	16
Nr.10	IIc	0.1290	11
Nr.11	IIb	0.1834	15
Nr.12	IIa	0.0772	2
Nr.13	IIc	0.0916	4
Nr.14	IIc	0.1058	7
Nr.15	IIc	0.1126	9
Nr.16	n.b.	0.0541	1
Nr.17	IIa	0.1178	10
Nr.18	IIc	0.1754	14

Figure 7:

5.2 Results of the test of Ziezold

For the 18 nephroblastomas we had information about the histological subtype in 16 cases at the time of statistical evaluation:

Subsets		Differentiation					
Tumortyp 1	Tumortyp 2	u_0	$m_=$	$m_<$	$p-Intervall$	k	$\binom{15}{k}$
$Typ\ a$	$\overline{Typ\ a}$	0	57	0	$[0.002, 0.125]$	3	455
$\overline{Typ\ a}$	$Typ\ a$	21	14	338	$[0.745, 0.774]$	12	455
$Typ\ b$	$\overline{Typ\ b}$	2	22	64	$[0.619, 0.819]$	2	105
$\overline{Typ\ b}$	$Typ\ b$	9	5	37	$[0.362, 0.409]$	13	105
$Typ\ c$	$\overline{Typ\ c}$	6	17	431	$[0.086, 0.090]$	6	5005
$\overline{Typ\ c}$	$Typ\ c$	14	155	780	$[0.156, 0.187]$	9	5005
$Typ\ d$	$\overline{Typ\ d}$	17	52	970	$[0.711, 0.749]$	4	1365
$\overline{Typ\ d}$	$Typ\ d$	10	40	153	$[0.113, 0.141]$	11	1365

m= ...: Number of cases with the same u-value
m<...: Number of cases with a lower u-value
The interval is a result of the smallest and the highest rank of u_0

Figure 8: Result of the test of Ziezold for differentiation the tissues of "Wilms-tumour

Legend: The Typ a, b, c, d are different tissues of "Wilms-tumours" The line over "Typ" means "Non-Typ"

The mean shape of Typ c and of Non-Typ c can be used for differentiation of each other. Although there are only 15 cases at all, the result for the p-value of H_0 is under 20

In the next figure the test is used for differentiation of the 24 tumours in their diagnosis.

Subsets		Differentiation						
Tumortyp 1	Tumortyp 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

m= ...: Number of cases with the same u-value
m<...: Number of cases with a lower u-value
The interval is a result of the smallest and the highest rank of u_o

Figure 9: Differentiation of renal tumours according to the test

Legend: Wilms: "Wilms-tumours"; K: clear cell sarcoma; N1: neuroblastoma; N2: renal cell carcinoma

The mean shape of Wilms-tumours can be used for differentiating the Wilms-tumours or the nephroblastoma from the group of the neuroblastoma and the clear cell sarcoma, because the p-value for H_0 is smaller than 10%. Only the mean shape of clear cell sarcoma can also be used for differentiating them from Wilms-tumours or the nephroblastoma.

In general the mean shape of Wilms-tumours can be used for differentiating from Non-Wilms-tumours with an u_0-value of 28 and a p-value for H_0 between 7% and 8%. The mean shape of Non-Wilms-tumours has an u_0-value of 54 and a p-value for H_0 between 78.5% and 80.5%.

The centralisation and standardisation is useful for differentiation, because without standardisation and two-dimensional centralisation - the original data centred on the three dimensional mass point - the mean shape of Wilms-tumours has for differentiating an u_0-value of 45 and a p-value for H_0 between 33.1% and 35.4%.

5.3 Result: Selecting relevant landmarks

To test the best five landmarks (landmark no. 3, 14, 15, 19 22) for differentiating the neuroblastoma from the nephroblastoma found above, the simulation from Giebel (2007) is used. In consequence of the heterogeneous group of non-nephroblastoma only three cases

with the diagnosis of neuroblastoma are compared to 18 nephroblastoma in the sample.

For deciding, if the selection is only a result of random processes, a random sample for comparison is produced.

To get a random landmark a normal distribution is used shown in figure 11. Because of the real centred and standardised data the median of the random distribution is 1 and the standard deviation is 0.5. 24 landmarks are produced for 21 cases.

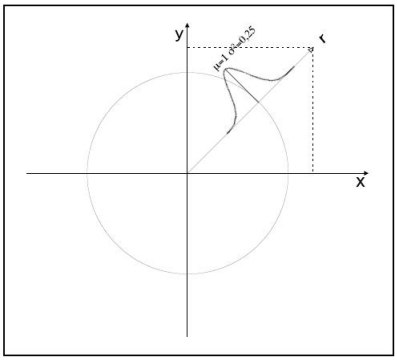

Figure 10: Producing the random data by using the normal distribution with median 1 and standard deviation 0.5

The original data produces more u-vales under five than the data produced by the random process upper described.

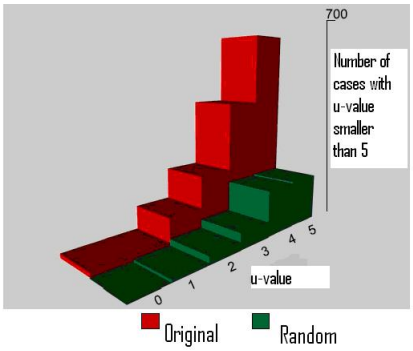

Figure 11: Comparing the original data with the random data: distribution of u-values in the original and random data

So the selection of landmarks in the real data should be better than the selection in the random process. The configuration of landmarks with u_0-values of 0 can be used and have to be tested in further samples for differentiating.

6 Detection of "odour"

Electronic noses are multi-sensor arrays. Every sensor measures a specific gas or substance. The data collected by electronic noses can be useful to avoid smell in canalization and to control biochemical processes in biogas plants.

Figure 12: "Electronic noses" produced by the company Airsense

The most suitable mathematical procedure is the statistical shape analysis. The shape of odour can be optically shown and tested for decision. The operator can see the shape or profile of odour and decides, if an intervention is necessary to avoid smell or to optimize the biochemical process in biogas plants.

The application of shape analysis to odour detection is shown through a second example.

The profiles of odour in biogas plants and their neighborhoods have to be differentiated from the profiles of odour in canalization. For the engineers it is a possibility to get a statement about the quality of odour and not only about the quantity. In contrast to the small sample size in medicine, here it is not possible any more to compute all possibilities.

So only a random selection is used for the test.

The selection of landmarks is not as easy as in the previous example, because every sensor represents a special chemical group or substance. In this example the quality of odour in the neighbourhood of biogas plants should be differentiated from canalization.

In Figure 15 you can see one biogas plants in Northern part of Hesse:

The most suitable mathematical procedure is the statistical shape analysis. The shape of odour can be optically shown and tested for decision. The operator can see the shape or profile of odour and decides, if an intervention is necessary to avoid smell or to optimize the biochemical process in biogas plants.

The application of shape analysis for odour is shown in one example. In this example the profiles of odour in biogas plants and their neighborhoods have to be differentiated from the profiles of odour in canalization. For engineering it is a possibility to get a statement about the quality of odour and not only about the quantity. In contrast to

upper research in medicine with a small sample it is not possible anymore to compute all possibilities. So only a random selection is computed for the test.

The selection of landmarks is not so easy as in the previous example, because every sensor represents a special chemical group or substance. In this example the quality of odour in the near of biogas plants should be differentiated from canalization.

In Figure 13 you can see one biogas plants in Northern part of Hesse:

Figure 13: Biogas plants in Northern part of Hesse: Location of the measurement of odour

The biogas plants are in a rural area. So there are many different qualities of odour in consequence of farming. Figure 16 shows part of a canalization.

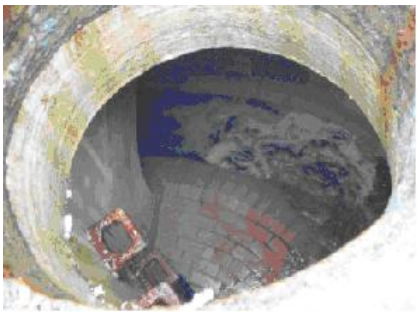

Figure 14: Canalization as one location of measurement

6.1 Sample

The odour is measured by the electronic nose. In every measurement we get a two-dimensional image of odour, the odour profile. This depends on the gas matrix which is in most cases individual for every kind of source.

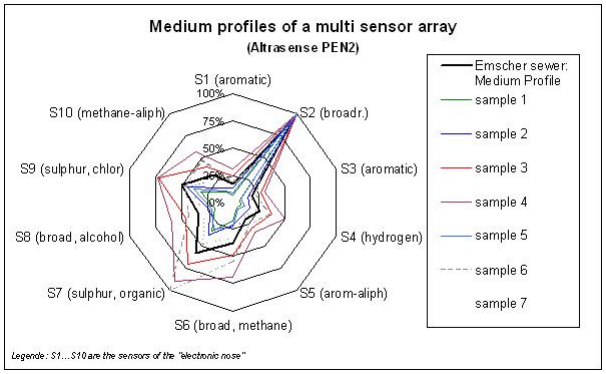

Figure 15: Odour profiles from different samples of same sewer

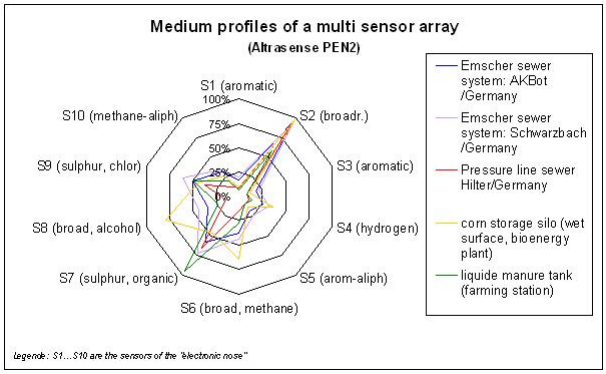

Figure 16: Odour profiles of different sources

These odour profiles from sources at the agricultural area (manure tank, storage silo) differ from the odour profiles of sewer air. Hence a separation between those sources can be done.

6.2 Result

We get an u_0-value of 530 for the mean shape of biogas plants. Comparing this value to 99 random calculated u-values we get a rank of 12 in a group of 100 random u-values and so the p-value is 0.12. In another other direction the p-value for the mean shape of canalization is 0.03. Hence, canalization can be differentiated from the biogas plants and the odour of canalization is more homogeneous than the odour of biogas plants. Selecting landmarks according to their relevance for differentiation is not so easy as in medicine: Every landmark represents a special chemical structure.

7 Conclusions and forecast

Shape analysis can be used in medicine and engineering to interpret two and three dimensional data. The results show that shape analysis gives interesting results and is worth further research.

7.1 Medicine

For the application of shape analysis on the classification of renal tumours we will use three dimensional landmarks in further research. A platonic object including the tumour will be used. The three dimensional landmarks are taken as cut points between the surface of the tumour and the line between mass point and edge of the platonic object.

Only real points on the surface of the tumour are used. If the cut point on the surface is not a real point, the real point as close as possible to the cut point is taken. The main advantage of this approach is that it minimizes the mistake due to approximation.

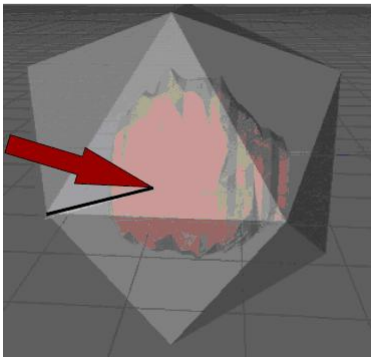

Figure 17: Further research: The three dimensional landmarks are taken as cut points between the surface of the tumour and the line between the mass point and the edge of the platonic object.

One problem that cannot be solved is that we do not have the same number of images (different slice thickness and spatial resolution) for every patient and not always the same imaging quality. It depends a lot on the used MRI-sequence-protocol and MRI system in regional imaging centres.

7.2 Engineering

We can use the interpretation of the application of shape analysis to odour differentiation also for the detection of explosive materials. In cooperation with the police office of the northern part of Hesse the odour profile of different explosive materials are measured. More than 100 times track hounds are asked in the northern part of Hesse for detection of explosive materials.

Electronic noses in this stage are not able to substitute completely track hounds, but they can indicate where track hounds should be used.

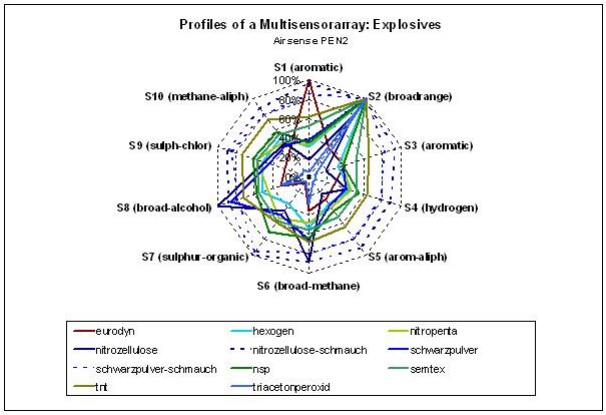

Figure 18: First measurement of electronic noses on explosive materials

Shape analysis can also be useful to detect drugs. Currently track hounds are mainly asked for detection of drugs, but they are not able to detect them without pause. In many cases like in prisons for instance, drug detection is needed 24 hours a day.

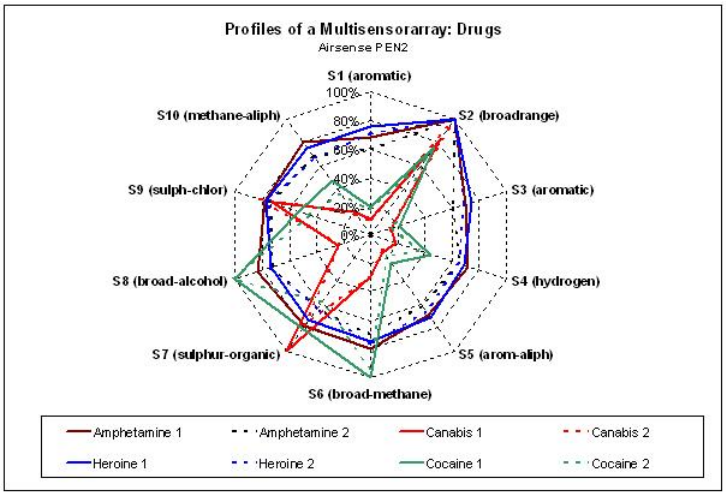

Figure 19: First measurement of electronic noses on drugs

It is necessary to use shape analysis or another mathematical procedures to interpret the data of electronic noses and to get a probability for the different substances. Electronic noses are only useful for detection of drugs and explosive materials if they are interpreted by means of a suitable mathematical procedure.

References

[1] Bookstein, F.L., *Size and shape spaces for landmark data in two dimensions (with discussion)*, Statist. Sci. **1** (1986), 181-242.

[2] Dryden, I.L., Mardia K.V., *Statistical Shape Analysis*, John Wiley, Chichester, 1998.

[3] Giebel, S.M., *Statistische Analyse der Form von Nierentunoren bei Kleinkindern*, master thesis of the University of Kassel, 2007.

[4] Graf, N., Semler, O., Reinhard, H., *Die Prognose des Wilms-Tumors im Verlauf der SIOP-Studien*, Der Urologe A, **43** (2003), 421-428.

[5] Kendall, D.G., *The diffusion of shape*, Adv. Appl. Probab. **9** (1977), 428-430.

[6] Kendall, D.G., *Shape manifolds, Procrustean metrics and complex projective spaces*, Bulletin of the London Mathematical Society **16** (1984), 81-121.

[7] Mardia, K.V., *Mahalanobis distance and angles*, Multivariate Analysis IV (Ed. P. R. Krishnaiah), North Holland, Amsterdam, 495-511, 1977.

[8] Schenk J.P., et. al. *Reference radiology in nephroblastoma: accuracy and relevance for preoperative chemotherapy*, Fortschr Rntgenstr, **178** (2006), 38-45.

[9] Schenk J.P., et. al. *Role of MRI in the management of patients with nephroblastoma*, Eur Radiol **18** (2008), 683-691.

[10] Small, C.G., *The Statistical Theory of Shape*, Springer Verlag, New York, 1996.

[11] Ziezold, H., *On expected figures and a strong law of large numbers for random elements in quasi-metric spaces*, Trans. 7th Prague Conference Inf. Th. Statistic. Dec. Funct., Random Processes, Prague, 1974, Reidel Dordrecht, Prague, Vol A. 591-602, 1977.

[12] Ziezold, H., *Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane*, Biometrical Journal **36** (1994), 491-510.