The use of image data in the assessment of equine conformation – limitations and solutions

Druml Thomas[1], Gabdulkhakova Aysylu[2], Artner Nicole[2], Brem Gottfried[1], Kropatsch Walter[2]

[1] Veterinary University Vienna, Institute of Animal Breeding and Genetics, Veterinärplatz 1, 1220 Vienna / Austria

[2] Vienna University of Technology, PRIP, Favoritenstrasse 9 / 186-3, A-1040 Vienna / Austria

I. Introduction

In this small scale experiment we want to examine which strategies and procedures of image processing techniques in horses may lead to a successful interpretation of the traits. Moreover, we test the applicability and the possibilities of image analysis for the purpose of animal breeding where the body shape – here called *conformation* – is a central part of breeding and selection concepts. Based on 40 two-dimensional anatomical and somatometric landmarks digitized from standardized images, variation of shape within a set of Lipizzan horses should be analyzed using Generalized orthogonal least-squares Procrustes (GPA) procedures and be correlated to the results from linear type trait classifications. The combination of "objective" measures from the horse shapes and the evaluations from the classifiers could be a tool for standardizing and fine calibrating selection procedures and breeding objectives. The objective of the "new phenotyping" is to increase the accuracy, objectivity and throughput of phenotypic estimation while reducing costs and minimizing labor through automation. Images and new statistical tools have the power to enhance experimental design and to extract biologically meaningful signals from environmental noise.

II. State of art in horse phenotyping

Concise measurement of a large number of lengths, circumferences and angles related to parts of the body is considered very important in breeding of many livestock populations, notably in cattle and horses. These measurements are used in selection and for the comparison of populations. Selection goals are often related to maintenance of a certain "medium" body size, a certain format, a characteristic type and several breed specific morphological features. Several publications [10, 4, 1] were focusing on the phenotypic variation of farm animal populations using morphological body measurements. In order to validate the measuring accuracy, the authors compared repeatabilities of measured traits and came to similar conclusions, even across species. Within a set of up to 37 different anatomical defined linear, circular and angular body measurements, the repeatabilities ranged from 0.1 to 0.95. This high variability within and between traits is the result of the different postures animals can take during the measuring process – we can call it "within animal variability". Some traits like height at withers, at back and at croup, body length and cannon bone circumference show a high consistency between repeated measurements ($r = 0.95 – 0.82$). On the other hand, ankles or measuring traits, which are hidden below a significant layer of muscles, are not easy to reproduce ($r < 0.70$) [10, 4]. In breeding programs of horses only four body measurements are being considered – these are: height at withers (repeatability (r)= 0.95), height at tape (r = 0.91-0.95), circumference of chest (r = 0.45-0.95) and cannon bone circumference (r = 0.82 - 0.95). All other traits, like length or posture of shoulder, form of neck, rearquarter aso., get evaluated within a scoring or classification procedure.

Subjective visual assessment of animals by classifiers is undertaken for several purposes in farm livestock. Linear type traits, body condition score, or carcass conformation are just a few examples where these "phenotyping" methods are used for gathering numerical data for further analysis and selection procedures. Linear type classification in horses is routinely performed in many countries, and records are used for the prediction of breeding values, longevity, sportive capability, feet and leg problems, locomotion and much more. One of the difficulties in assessment is the effect of individual classifiers. For validation purposes of scoring data and to ensure, that classifiers rank animals consistently, repeatability measures or intraclass correlations may be used to validate scoring of individuals. Classifiers may differ in their mean score, in the range of scale that they use, or in their attitudes. For breeding value estimation based on BLUP (Best Linear Unbiased Prediction) methods within breeds of big population sizes may be adjusted using complex statistical models. Within small populations, where only a few dozen animals are registered per year, such statistical models are limited due to lack of data. Therefore the fact that the classifiers rank animals consistently – i.e. they use a consistent trait definition at all times – becomes crucial, as their records are the basis for selection or culling decisions. Usually classifiers are trained regularly, in order to improve their rating consistency. As opposed to a small numbered horse breeds with limited offspring per year, the training of classifiers becomes difficult, as they observe and score a few samples out of the breed specific natural variation.

Due to these problems - low repeatability of body measurements and poor reliability of assessments in phenotyping procedures - within the context of horse conservation breeding programs, we aim to develop a standardized image based evaluation procedure, which may

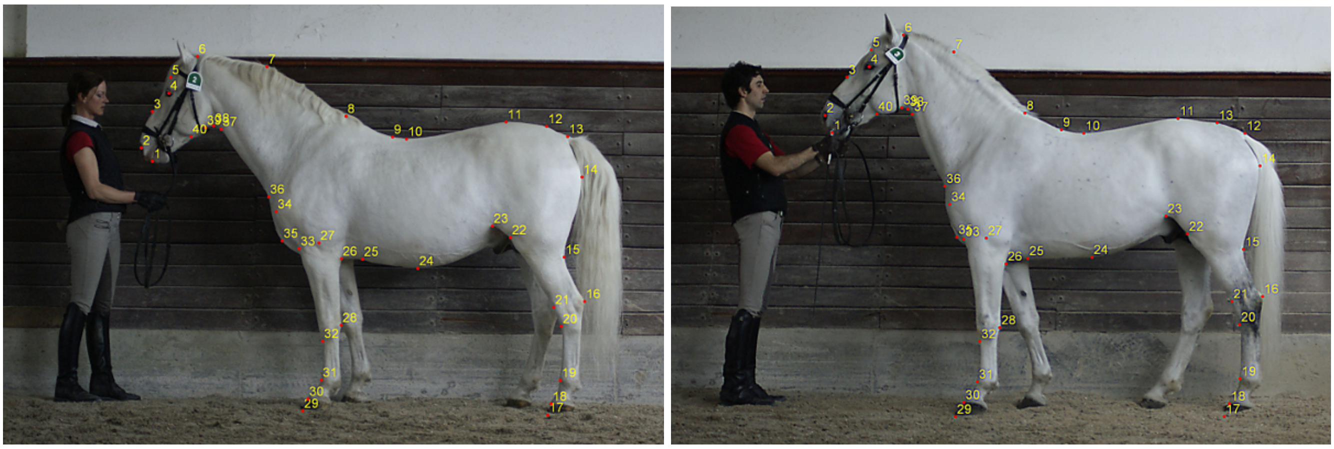

Fig. 1. Position of horses and landmarks used for this study. The horses are presented in a standardized way, where the left foreleg and the hock-fetlock connection of the left hind leg are nearly vertical, the right fore and hind legs also can be seen from the left side, the neck is kept in a natural way according to the conformation of the horse.

assist in ranking of animals within breeds with limited population size. Using the image data and machine learning concepts enables to achieve more objective and stable assessments of linear type traits over time, and, therefore, aims to preserve the phenotypic and genetic characteristics of breeds with limited population sizes.

III. Interrater Reliability and Shape Analysis of Lipizzan Horses

In a small scale experiment data was recorded during a teaching workshop by 16 Lipizzan horse breeders and three Lipizzan stud farm directors (hereinafter named students and instructors) in the Lipizzan stud farm Lipica. The linear type trait scoring protocol was defined by the instructors and it contained 10 different traits which were scored at a scale of 10 points by 0.5 units increase. Additionally standardized images of the 11 inspected horses were taken according to the following principles: distance between a horse and a camera = 18 meters; camera lens (100 mm) focused onto the centre of gravity of the horse; horses were presented in the same postures. In order to assess the agreement among classifiers, the Kappa statistics according to [3] were calculated using the SAS macro mkappa [6] which is stable to multiple raters and missing data. The resulting Kappa coefficients are presented in Tab.1 and reflect the poor rater-reliability within the students group (mean kappa = 0.03) and the instructors group (mean kappa = 0.25). In comparison to this results [12] reported weighted Kappa coefficients for body condition scores within three cattle herds that ranged between 0.17 for students and 0.78 for instructors. In evaluations of classification results generally measurements like repeatability, or inter- and intraclass correlation coefficients are used. This approach does not consider the differences in scale and mean of the single evaluations, and, therefore, results in higher values of concordance than the Kappa test for exact agreement.

TABLE I. Classifier's uniformity across ten traits per animal represented by Kappa coefficients (n.e. = not estimable). The Kappa test accounts for exactness of agreement between scores given by different classifiers. The poor values of the students group reflect the difficulty in achieving consistent results in the evaluation of horse conformation.

Horse	Instructors				Students				Together			
	N Judges	Kappa	SE	P value	N Judges	Kappa	SE	P value	N Judges	Kappa	SE	P value
1	3	0.25	0.18	0.0900	15	0.02	0.02	0.1162	18	0.01	0.02	0.2182
2	3	0.20	0.14	0.0809	16	0.02	0.004	0.0001	19	0.008	0.01	0.3683
3	2	0.49	0.14	0.0003	15	0.03	0.01	0.0012	17	0.02	0.01	0.0087
4	3	0.19	0.14	0.0862	15	0.04	n.e.	n.e.	18	0.05	n.e.	n.e.
5	3	0.29	0.15	0.0220	15	0.01	0.01	0.3655	18	0.02	0.01	0.1402
6	1	n.e.	n.e.	n.e.	8	0.09	0.05	0.0368	8	0.09	0.05	0.0368
14	3	0.19	0.17	0.1317	13	0.03	0.02	0.0249	13	0.03	0.02	0.0156
15	3	0.21	0.16	0.0898	13	0.03	0.01	0.0156	13	0.04	0.01	0.0014
16	3	0.19	0.15	0.0967	14	0.02	0.02	0.0769	14	0.02	0.01	0.1054
17	1	n.e.	n.e.	n.e.	15	0.02	n.e.	n.e.	15	0.02	n.e.	n.e.
20	3	0.26	0.10	0.0036	12	0. 04	0.02	0.0684	12	0.04	0.02	0.0069
All		0.11	0.04	0.0029		0.07	n.e.	n.e.		0.08	n.e.	n.e.
Mean		0.25				0.03				0.03		

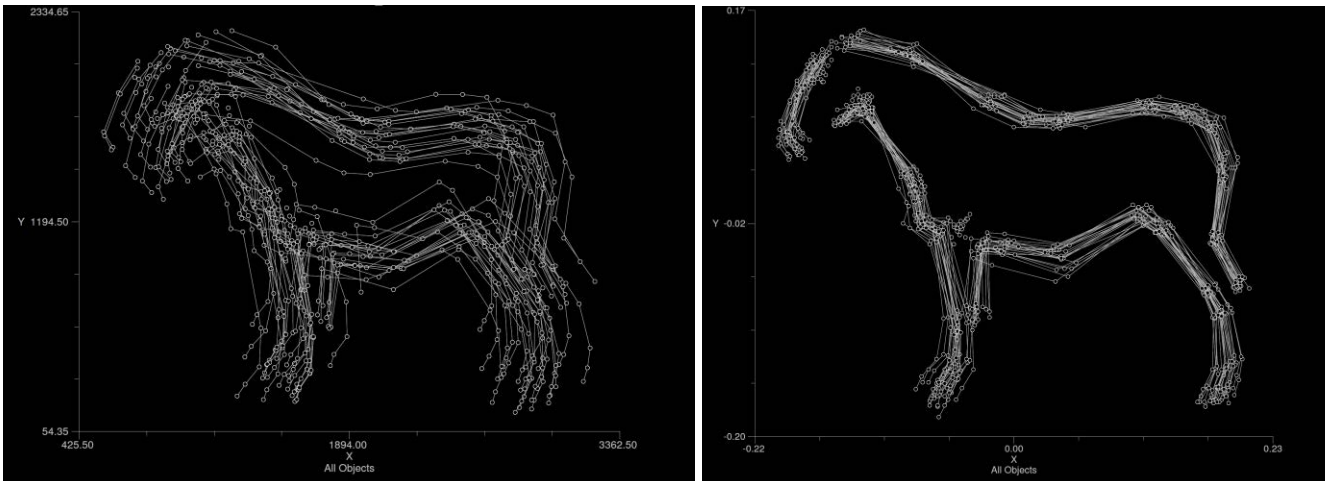

Fig. 2. Digitized dataset before applying the Procrustes fit (on the left), i.e. before x and y configurations being centered, scaled to unit size and rotated against to each other achieving minimum squared distance, and optimal superimposed specimens of unit size and minimum distance to the sample mean (on the right).

The images taken during this workshop will be analyzed applying Generalized orthogonal least-squares Procrustes (GPA) procedures [11], where 40 anatomical and somatometric landmarks are extracted from the pictures (see Fig.1.), their x and y configurations being centered, scaled to unit size and rotated against to each other achieving minimum squared distance [2]. The resulting shape variables – uniform components, partial warp scores and relative warp scores (PCA scores of the partial warp scores) – will be used for further analysis of the shape variation from the Lipizzan horses that were classified during the workshop.

The relative warps summarize the variation among the specimens with respect to their partial warp scores in as few dimensions as possible. The partial warps are eigenvectors derived from the bending energy matrix from a thin plate spline (TPS) as described in [2] and represent non-affine shape deformation. Partial warp scores are calculated by projection of differences between aligned forms to the reference onto the partial warps. They can be interpreted as a bent surface of the thin plate spline. An example of shape variation along Principal Component Axes (relative warp scores) is shown in Fig.3. The first four PCA axes account for 84% of the total variance in this horse sample. Each axis shows specific shape changes from its minimum to the maximum value.

The next step in the analysis is to perform a regression analysis of the shape coordinates on the rating data, in order to determine the shape variation which is associated with the classifying results. This work was still in progress when submitting this abstract.

In Fig.3. it is evident that PCA axes 1 and 2 mostly account for variation between different postures, which is the main problem when working with image data from living animals. The possible solution to this problem is provided in the next section.

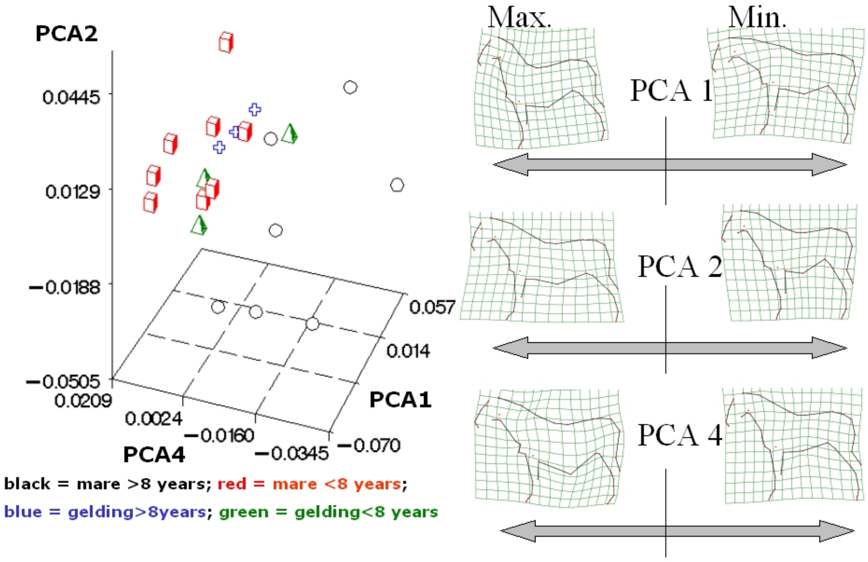

Fig. 3. PCA Analysis of partial warp scores derived from the bending energy matrix of thin plate splines (TPS) based on the data pool of 21 Lipizzan horses. Shape changes along the PCA axes 1, 2 and 4, which account for 70 % of shape variation within a sample of 21 Lipizzan horses from the Austrian stud farm of Piber. Although PCA axes 1 and 2 account mostly for effects of posture, PCA scores separate the single horses according to their age class and sex.

IV. Possible Solutions for the “within animal variance” problem

We discovered that phenotypic traits provide insufficient information for measuring the “within animal variance” and modeling the shape of the horse while obtained from the anatomical landmarks in a 2D image. The factor is that the linear distances between the anatomical landmarks are an interpolation of the corresponding horse body curve. Illumination conditions as well as the pose of the horse impact the perception of its shape and, thus, the position of the landmarks. In area of animal breeding there is a need in quantitative trait's measurements which are effective, reliable, objective and comparably cheap. 3D scanning techniques can become a major step in development and analysis of the horse traits. Professional 3D surface-imaging systems capture texture and depth information within a few milliseconds and with high accuracy (less than 2 mm root mean square (RMS) error) [9]. In this manner, in project CAESAR[1] the usage of 3D technologies provided more research capabilities for the study, as well as fast and reliable ways to measure anthropometric characteristics. Unlike human, it is very challenging to keep a horse standing still. The existing 3D modeling approaches achieve the pose invariance by applying geodesic distance [5], segmentation into meaningful entities [7] and shape-diameter-function [8]. However, they are not applicable for the traits which cannot be measured on the surface (shoulder joint, most caudal point of the scapula, thurls, etc.), since they lie beneath layers of tissue (muscle, fat) on the bones of a horse.

Therefore, it is mandatory for the horse model to also have the bone and body structure (skeleton). A proposed extension of the classical medial axis transform would be to replace the spherical structuring elements of the skeleton by ellipsoids fitted into the 3D shape of the horse. The obvious advantage is that the ellipsoid has an orientation and two ellipsoids can be coupled at the natural articulation point. Each ellipsoid induces an internal coordinate system (ellipsoidal coordinates) allowing mapping the 3D data of the horse onto the elliptical reference systems of each rigid part. This makes the representation invariant to the natural (articulated) movements of a horse. In addition, it enables to locally estimate the bone structure inside the shape by comparing it to an anatomical model. Having mapped the acquired data to a reference coordinate system, different views can be fused to increase the reliability of the data and to estimate the variability for a single horse and for horses of a similar genotype. In this 3D horse model, the shape-adapted reference systems take the role of normalization that is of crucial importance for applying, e.g. principle component analysis to the different classes of data. The development of a novel 3D horse model enables to enhance the normalization and standardization problem. This leads to further reducing the “within animal variance” and improving the repeatability.

V. Conclusion

Conformation plays a crucial role in the area of animal breeding. The performed small scale experiment revealed the impediment to interpretation of phenotypic traits from 2D images since the acquired PCA axes represent the deviation in postures rather than in breeding characteristics. In our view the solution to this problem lies in utilization of 3D scanning technologies in conjunction with structural representation (bone structure) of the horse shape.

The existing dataset[2] contains 348 images randomly selected from the Web. They represent horses of multiple breeds positioned in various postures. Such dataset can be utilized for the purposes of verifying the phenotypic trait measurements. In contrast, our collection of standardized side images provides a groundtruth for evaluation of phenotypic as well as genotypic characteristics of Lipizzaner horses.

Acknowledgment

Aysylu Gabdulkhakova acknowledges financial support by the Austrian Agency for International Cooperation in Education & Research (OeAD) within the OeAD Sonderstipendien program, financed by the Vienna PhD School of Informatics[3].

References

[1] Averdunk, G., Flatnitzer, F., Gottschalk, A. und Schwarz, E. Über die Genauigkeit der Messung einiger Körpermaße bei Kühen. Bayer. Lw. Jb. 47, 599-604, 1970.

[2] Bookstein, F. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge, 1991.

[3] Fleiss, J. Measuring nominal scale agreement among many raters. Psych. Bull., 76, 378-382, 1971.

[4] Grundler, C. und Pirchner, F. Wiederholbarkeit der Beurteilung von Exterieurmerkmalen und Reiteigenschaften. Züchtungskunde 63, 273-281, 1991.

[5] Katz, S., Leifman, G. und Tal, A. Mesh segmentation using feature point and core extraction. *The Visual Computer* 21.8-10 (2005): 649-658, 2005.

[6] SAS ves. 9.1, SAS Institute Inc. SAS

[7] Serino, L., Sanniti di Baja, G. und Arcelli, C. Using the skeleton for 3D object decomposition. In: Image Analysis, pp. 447-456. Springer, 2011.

[8] Shapira, L., Shamir, A. und Cohen-Or, D. Consistent partitioning of meshes. Technical report, The interdisciplinary center, 2005.

[9] Tzou, J., Artner, N., Pona, I., Hold, A., Placheta, E., Kropatsch, W. und Frey, M. Comparison of three-dimensional surface-imaging systems. Journal of Plastic, Reconstructive & Aesthetic Surgery, 67(4):489–497, 2014.

[10] Zechner, P., Zohman, F., Sölkner, J., Bodo, I., Habe, F., Marti, E. und Brem, G. Morphological description of the Lipizzan horse population. *Livest. Prod. Sci.*, 69, 163-177, 2001.

[11] Rohlf, F. J. and D. E. Slice. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zool., 39:40-59, 1990.

[12] Kristensen, E., Dueholm, L., Vink, D., Andersen, J.E., Jakobsen, E.B., Illum-Nielsen, S., Petersen, F.A., Enevoldsen, C. Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. J. Dairy Sci., 89:3721-3728, 2006.

[1] http://store.sae.org/caesar/

[2] http://www.msri.org/m/people/members/eranb/

[3] http://www.informatik.tuwien.ac.at/teaching/phdschool