Geometric Biometrics as a Robust Approach to Quantification of Livestock Phenotypes

Catherine McVey, Pablo Pinedo, Department of Animal Sciences, Colorado State University

Introduction

In most standard linear models, model specification includes the simplifying assumption that covariates are known without error (Kutner et al 2005). For some simple categorical variables, like gender or herd of origin, this is perhaps true, excluding errors in data entry (Meyer 1997). Measurements that seek to capture more complex and dynamic sources of information, however, can seldom be represented without error. In this case, a thorough understanding of the nature of the error associated with a measurement system is necessary to fully evaluate the appropriateness of such simplifying assumptions, and where necessary, make accommodations and adjustments in the development of a robust model. At the outset of a chapter focused entirely on metric validation and characterization of error, it therefore seems prudent to briefly reflect more abstractly on the process of measurement of complex features.

When approaching the task of extracting anatomical information from a digital image, it is essential that one not conceptualize a cow simply as a solid object of finite dimensions existing in three-dimensional space. Instead, conceive of an image of a cow as existing in a high dimensional space, sometimes colloquially referred to in the natural sciences as a hyperspace, built from a composite of information of numerous types (Hurlbert 1981; Ojiem et al 2006; Van Heel 1984). Part of the high dimensional space in which such an image lives will capture information about the physical attributes of an animal, but many other dimensions will capture extraneous information classified in this application as noise - age, coat length, coat color, cleanliness, emotional state, position, light exposure, shadow exposure, background, etc. In developing a novel measurement system, the goal is to extract the maximum amount of information of a desired type, the signal, from information deemed extraneous, the noise (Measurement Systems Analysis Work Group 2010). This is effectively done through a series of data compressions steps. Careful consideration must be given to the assumptions made at each stage of dimensionality reduction to account for potential sources of error introduced by the compression technique selected (Kirby 2001). Finally, full characterization of a measurement system requires a thorough characterization of the resulting metrics to ensure they demonstrate traits amendable to standard methods of statistical inference and predictive modeling (Kutner et al 2005).

Dimension Reduction

The first major compression of information utilized by this measurement system occurred in reducing the temporal dimension down to a single time point. It was assumed that boney structural features of a mature cow's head would not change significantly over a time window spanning only two weeks, excluding any obvious physical injuries. It should be stated that this is fully an assumption, as it does not appear that this issue has been addressed in the existing body of published research, and validation through an extended longitudinal study of the boney structures of the bovine face has been left to future research. In the process of obtaining facial photos, it was observed that this assumption may have been violated for select boney and cartilaginous structures obscured by a significant amount of soft tissue, as variations in facial expressions might obscure measurement of facial structure on a much finer time scale. This observation was explored indirectly as part of the larger metric validation procedure.

The second and perhaps most significant compression of information utilized by this measurement system was exclusion of pixel exposure information so that facial structures were represented only by the distances between landmark structural points. An image is typically represented by an $m \times n \times 3$ matrix, where each pixel index contains, depending on the format, a real number value reflecting the exposure level or hue intensity at that position in the captured scene. Even for a camera with moderate pixel resolution, this represents a massive amount of information contained in many thousand pixel values. Photos are often compressed to a grayscale $m \times n \times 3$ matrix for image analysis purposes, but even for modestly large image databases, standard pixel-based analysis techniques, like eigenface analysis, can quickly become computationally intractable (Kirby 2010). This would impose computational constraints that could limit the applicability of such a system, particularly if it were to be implemented at a breed-level.
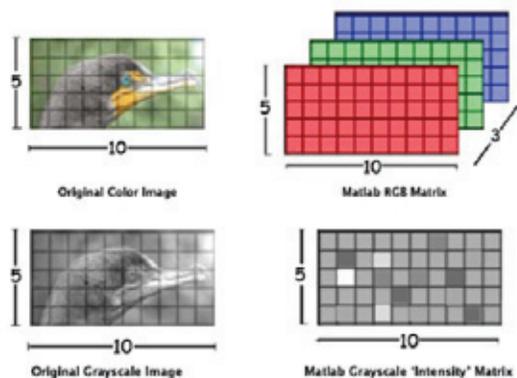
Figure 1: Basic Anatomy of a Digital Image (Singh 2012)

Perhaps of greater concern, however, is the susceptibility of pixel-based image analysis techniques to extraneous noise. At its core, eigen face analysis is just a form of principal components analysis, where in this case each pixel index represents a variable. Given a large number of images captured under very controlled conditions, eigen face analysis can be an effective means of dimension reduction, eliminating noise and redundancy in a set of training images to yield a much smaller set of basis images that concentrate the information of the signal. Like standard principal components analysis, however, eigen face algorithms are greedy algorithms, creating at each step a basis image which captures the maximum amount of variation found in the original training set possible. Subsequently, eigen face algorithms offer very little control with regards to how information is partitioned within basis images (Kirby 2001). This makes them a poor choice for isolating a specific class of features, such as facial anatomy, when conditions do not allow for tight control of other factors influencing image quality. Fluctuations in image exposure alone, with all other factors held constant, typically account for at least 10 dimensions in the resulting image space (Beveridge et al 2009). They also cannot discern between changes in the intended subject of the image (foreground) and random artifacts in the background, which can be difficult to control in a field setting. Additionally, in animals, eigen face algorithms seem more influenced by changes in coat pattern than overall facial morphology, a major concern for application to Holstein populations (Caiafa et al 2005). Neural network-based image analysis techniques, which can be thought of as a non-linear extension of principal components based eigen-techniques, are a newer approach, and subsequently not as well defined (Kirby 2001). However, deep learning algorithms, likely by virtue of their multiple differentiable layers, seem more adept at paring complex components of images down into their simpler components, making them perhaps a better algorithmic candidate for extraction of facial phenotypes from farm quality images (RSPI Vision 2017). Unfortunately, robust networks require large and diverse image data sets to train, making them difficult to implement for applications without existing databases.

Given these constraints, it was determined that a face mesh approach was more appropriate for this domain of application. With this approach, key anatomical landmarks of the face are determined a priori. All images in a database are subsequently annotated with these landmark points, and their coordinate location within the pixel matrix of each image recorded and used in subsequent analyses. Large image databases have been used to train fully automated algorithms for landmark point extraction for applications in humans, but such work has not been pursued for animal populations, which again imposes practical constraints on the scale at which this technique can be imposed. Here again, however, deep learning algorithms have shown promise in this area, and new research indicates that learning algorithms for landmark point extraction trained on larger human databases may be effectively adapted to livestock features with much smaller reference data bases when strategic constraints are applied (Rashid 2017). For the purposes of this largely exploratory study, it was deemed sufficient to simply extract landmark points manually. By using MatLab's GINPUT tool to interactively select a predetermined series of key anatomical reference points on the face, and storing their coordinate locations within the pixel matrix, extraneous information related coat pattern and features of the farm environment, like variable lighting exposure and changing background content, were effectively excluded. It should be noted, however, that in applying this compression, a significant amount of structural information was inevitably lost as well, with only structural points that had been identified as descriptive a priori being retained for further analysis. This could serve as a source of bias, if certain regions of the face or types of structural variations were not adequately described by the landmark points defined. Additionally, physical selection of these anatomical points within the image was not without error, requiring targeted analysis to determine the magnitude and systematic nature of this source of measurement error.

The third major compression of information came from reducing facial structures from 3 to 2 physical dimensions. It would be possible to represent structural features of the face via a 3-dimensional image, and thereby capture all dimensions of facial shape, up to the resolution of the camera and accuracy of the stitching algorithm (Aldridge et al 2011; Obafemi-Ajayi et al 2014). To do so, however, would be an expensive and time-consuming endeavor, requiring specialized equipment and greater restraint of the animal. Thus, this compression decision was driven predominantly by practical concerns, as it was deemed that loss of information was outweighed by gains in accessibility realized by developing this measurement system around

the specifications of any standard quality digital camera. In projecting a 3D object onto a 2D plane, however, several sources of error are introduced. The first and most important is angle of the object relative the plane of the camera. Significant variations in angle related to depth can effectively distort the resulting image as it is projected onto the plane, effectively warping the relative distances between facial structures. This is a major concern, as it not only distorts the perceptions of facial shape, but because of the underlying geometry, tends to do so in a systematic way. In other words, errors from this source are not necessarily random, and tend to be correlated, which breaks the assumptions of many statistical models (Kutner et al 2005). This source of error was addressed in two ways. The first approach was procedural, attempting to reduce variation in camera angle as much as is possible on a farm working with large and at times disagreeable animals. Side profile images were obtained parallel to the surface of the cheek. This was partially achieved by attempting to center the image on the eye, and then aligning as closely as possible the ridges of the eye orbitals on either side of the forehead. Front profile images where obtained parallel to plane of the forehead. This was achieved by attempting to equalize the distance between either eye and the center of the forehead on either side of the face, and then seeking to obtain an image where the nose appeared as long as possible.

Figure 2: Impact of Out-of-Plane Variations in Face Angle on Coordinate Locations

The second major source of error introduced by this compression came from variations in relative position of the camera to the cow. Varying distances between the camera and object changes the proportion of the frame dedicated to the cow's face. This in turn changes the number of pixels

dedicated to capturing structural features of the cow's face (i.e. image resolution). If raw pixel distances between anatomical points were used, changes in image resolution would become a major source of error due to differences in scaling. This issue is often addressed by scaling the image to a known reference length of an object with the frame of the image, but this solution was deemed impractical on a working farm environment. Attempting to place a reference object in the frame near the cow so that it would be in a plane equidistant to the camera with the cow would not only significantly increase the amount of time required to obtain an image, but also increase stress experienced by the animal and put the handlers in a more exposed position. Instead this issue was addressed by developing biometrics that either reported angles or distances as proportions. Computation of angles between traits are of course geometrically dependent only on their relative, not absolute, distances. Similarly, by using proportions to report relative distance measures, the scaling factor of the image was effectively "divided out". Thus, this measurement system should be inherently robust to changes in image resolution that result from variable distances between camera and cow, as well as any variations in specs of the camera used or degree of zoom applied. Practically speaking, this greatly simplified the process of acquiring images of the cows and allowed greater focused to be placed on reducing variations in image angle.
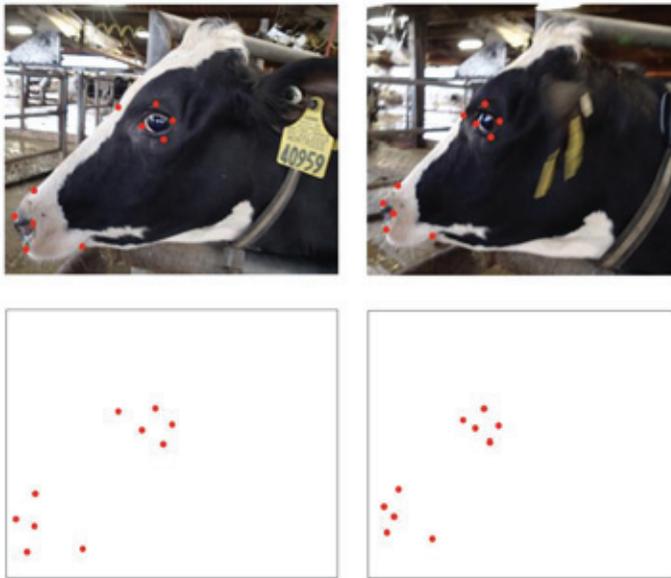

$$\frac{300}{30} = \frac{2(150)}{2(15)} = \frac{150}{15} = 10$$

Figure 3: Illustration of Elimination of Scale Effect by Division Op

One exception to this assumption of distance invariance was that, when the photograph was taken extremely close to the cow's face, as frequently happened when photos were acquired in the feed bunk, there seemed to be a significant interaction between position and angle. Put simply, when quite close to the cow, aligning the camera using the eye structural reference point created the correct

90-degree angle for the central part of the face, but still left a significant angle between the camera and distal parts of the face, namely the nose. Auxiliary image measurements were used in an attempt to correct for this potential source of error in such images.

The fourth and final source of compression comes from converting the 2D coordinate vectors representing the locations of key structural points into 1D descriptive measures that could be used as covariates in predictive models. Previous studies have frequently accomplished this by simply taking the Euclidean distance between all pairwise combinations of anatomical points, globally normalizing by the sum of all lengths to correct for differences in image resolution, and then reducing the number of candidate variables by using a multivariate compression technique such as principal components (Cole et al 2016; Aldrige et al 2001), or else using data clustering techniques designed for high dimensional input (Obafemi-Ajayi et al 2014). While this procedure is quite simple to apply, it has two key drawbacks. The first and most significant is that the resulting distance measures are directly geometrically related, resulting in complex correlations structures. A slight change in the relative position of one anatomical point would be reflected in slight changes in all pairwise distances of which that point is a member. When points change their relative positions due to underlying face shape, associated Euclidean distance terms will change as well, but so many of these points would change simultaneously that it becomes difficult, if not impossible, to discern the nature of this geometric shift just from direct appraisal of the data.

Algorithmic Solutions

Principal component analysis is a means to concentrate this redundant information, but in doing so assumptions of linearity are necessary. When a number of facial features shift simultaneously, their cumulative effects on individual pairwise distances may not necessarily be additive, which could potentially lead to inflation of the parameter space or misleading reparameterizations (Kirby 2001; Johnson & Richard 2007). When the relative position of points changes due to error in point selection, as opposed actual changes in facial shape, this error is also subject to geometric constraints between pairwise combinations of points, potentially leading to correlation in the error structures. Most correlation-based multivariate techniques, including principal component analysis, require the assumption that error terms are uncorrelated. When this is not in fact true, correlation in error is mathematically interpreted as correlation in the signal. As a result, application of these dimension reduction techniques lead to concentration of both signal and error simultaneously (Johnson & Richard 2007), which is at best inefficient but also a potential source of bias in downstream analysis.
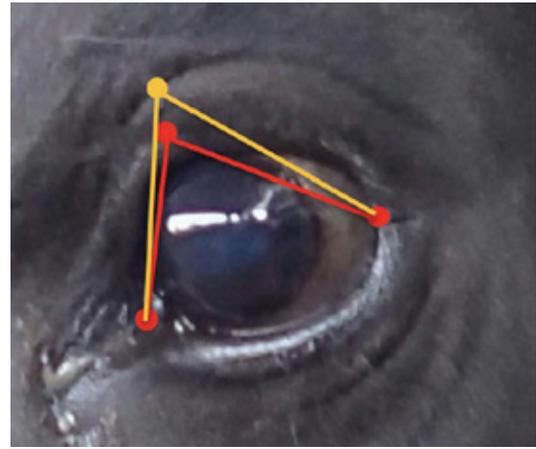

Figure 4: When horizontal eye landmarks are constant, error in point selection leads to triangular relationship between edges

The second drawback of this technique is that principal component analysis, while an effective means of dimension reduction, is limited in terms of its descriptive ability. For exceedingly high dimensional input, it is difficult to determine from the orthogonal bases vectors what information is captured in each new transformed variable. In other words, it might be possible to determine from the relative scale of orthogonal basis values that a given dimension is dedicated largely to describing variations in eye structure, but it would be difficult if not impossible to determine what this would relate to in terms of the underlying structural variability without a means of effectively regenerating the face (Nielson et al 2011). This makes any subsequent models built using reparameterized variables difficult to interpret. While this is perhaps sufficient for purely predictive models, it makes it difficult if not impossible to assess the biological appropriateness of such results. Further, as principal component analysis is not a model-based technique, it is not generally considered readily extrapolateable to novel data sets, which makes it more appropriate for descriptive studies as opposed to predictive modeling (Johnson & Richard 2007).

In an effort to overcome these drawbacks, a geometric approach to biometric extraction was developed. This approach had two key goals. The first was to minimize correlation between resulting biometrics, attempting to isolate specific changes in shape using targeted geometric relationships on the front end of the algorithm to create independence between measurements, as opposed to applying an indiscriminate orthogonalization technique like PCA on the back-end. This was done in two ways. The first was that, as opposed to normalizing pixel distances between points using the sum of all pairwise distances, it divided by distances between nearby points that were selected to produce more intuitive interpretations of shape. For example, instead of describing the height of the eye as

a proportion of overall face size using the sum of distances, which would in turn be influenced by many other unrelated anatomical factors like jowl depth or nose length, it was compared directly to the length of the eye, or to the depth. The second means of achieving this goal was to make use of projection lengths over simple Euclidean distances. The coordinate locations of many key anatomical points were frequently observed to be influenced by multiple independent variations in facial shape. By projecting such a point onto a number of carefully selected reference slopes from nearby facial features, the effects of these independent shape variants could be more effectively broken up into distinct distance measures to isolate their independent effects. For example, the location of the highest point of the eye is influenced by two variants in eye shape: how tall the eye is, and how angular the top of the eye is (i.e. how far forward is the highest point). Simple Euclidean distances would capture both effects at the same time. By instead relying on projections, the angularity of the eye is captured by projecting the highest point of the eye onto the horizontal plane of the eye (Eye Height Point Proportion - EHPP), and the height of the eye is captured by projecting the highest point of the eye onto the plane perpendicular to the horizontal reference plane of the eye (Eye Height Proportion - EHP).
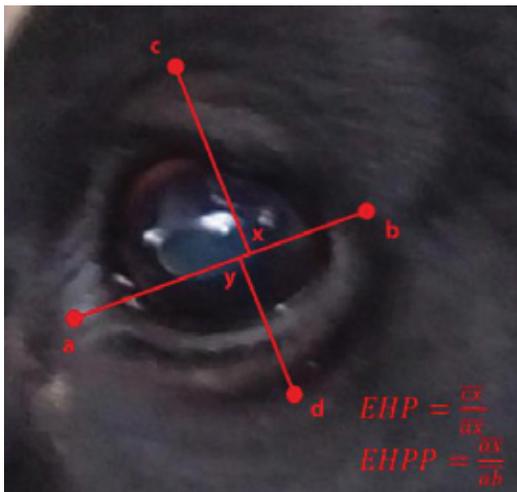

Figure 5: Example of Geometric Biometric using Orthogonal Projects

The second goal pursued with a geometric approach to biometric extractions was to reduce the impact of errors in point selection. This was done in several ways. First, by attempting to define facial shapes using targeted comparisons of length measures, as opposed to Euclidean distances between all pairwise combinations of anatomical reference points, and by relying on a local as opposed to global normalization scheme, any error incurred in the selection of the coordinate location of a given anatomical reference point was effectively isolated to only a targeted

handful of metrics, and not amplified across the broader set of metrics. Put more simply, if the coordinate location of the highest point of the eye was selected poorly in a given picture, that error was only seen in a subset of the eye biometrics and had no impact on biometrics extracted from the nose, topline, or forehead. This characteristic was enhanced by relying predominantly on projection lengths. Just as the projections were used to break down distances into the distinct influences of shape, they also effectively orthogonalized the components of error. This was particularly helpful for traits where coordinate selection was perhaps clear in one direction but less easy to distinguish for another. For example, take highest point of the eye. For very rounded eyes, multiple coordinate selections might return points with very similar vertical distances horizontal plane of the eye, but a great deal of variability in the horizontal distance. This error in point selection would in turn be isolated only to metrics that relied on the horizontal component of this point location and have virtually no influence on metrics that rely only on the vertical distance, whereas for simple Euclidean distance this error would influence any pairwise combination that involved this point.
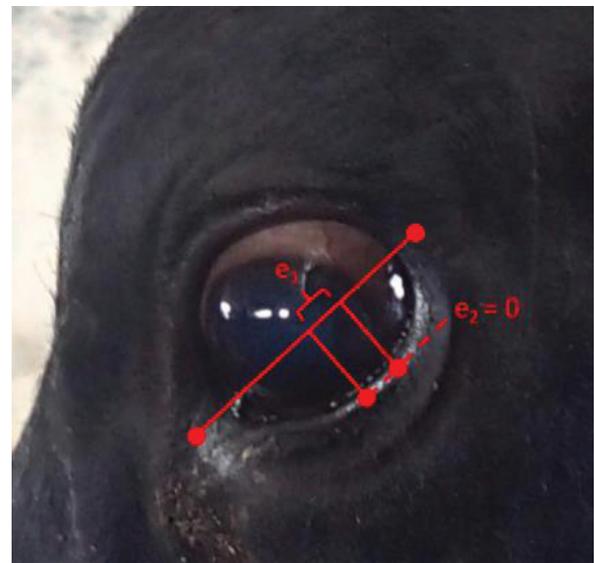

Figure 6: Orthogonalization of Error Component

Finally, measurement error due to error in point selection was also reduced with the strategic use of interpolated points, defined as the intersection point of two extrapolated lines formed by anatomical points from other regions of the face. Efficiently calculated using a solution to standard cross-product formula, such points were frequently used to infer the location of an anatomical structure that could not be reliably identified by eye. This frequently happened for traits obscured by a significant amount of flesh or muscle. The location of the back of the jaw is an example of an interpolated point. Often difficult to identify visually for cows with significant amounts of skin and fat deposits

around the jowl, this point was interpolated by projecting a line from the bottom of the chin along the jaw bone and then finding its perpendicular intersect with the back of the poll.

Figure 7: Example of an Interpolated Landmark Point (red)

A total of 104 candidate geometric biometrics was developed to fully describe the shape of the bovine face. To assess the efficacy of this novel approach to biometric extraction from digital images, geometric biometrics will be compared to standard normalized length measures within each region of the face to determine which strategy demonstrates more robustness to measurement error while minimizing correlations between metrics without use of dimension reduction techniques. Final estimates of repeatability will then be used to select which candidate biometrics demonstrate sufficient robustness to warrant farther study in predictive models of dairy productivity and longevity.

Citations

> Aldridge, K., George, I.D., Cole, K.K., Austin, J.R., Takahashi, N., Duan, Y., Miles, J.H. "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes." 2011. Molecular Autism. 2:15.

> Bates, D., Mächler, M., Bolker, B.M., Walker, S.C. "Fitting Linear Mixed-Effects Models using lme4" Journal of Statistical Software. 2015. 67:1.

> Beveridge, J.R. Draper, B.A., Chang, J.M., Kirby, M., Kley, H., Peterson, C. "Principal Angles Separate Subject Illumination Spaces in YDB and CMU-PIE." 2009. IEEE Transactions on Pattern Analysis and Machine Intelligenge. 31:2. 351:356.

> Caiafa, C.F., Proto, A.N., Vergani, D., Stanganelli, Z. "Development of individual recognition of female southern elephant seals, Mirounga leonina, from Punta Norte Peninsula Valdes, applying principal components analysis" 2005. Journal of Biogeography. 32: 1257-1266.

>Cole JB, Manyama M, Kimwaga E, Mathayo J, Larson JR, Liberton DK, et al. (2016) Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. PLoS Genet 12(8): e1006174. doi:10.1371/journal.pgen.1006174

> Hurlbert, Stuart. "A gentle depilation of the niche: Dicean resource sets in resource hyperspace." 1981. Evolutionary Theory 5:177-184.

> Johnson, R.A., Wichern, D.W. "Applied Multivariate Statistical Analysis" 6th Edition. 2007. Pearson Education Inc. USA.

> Kirby, Michael. "Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns." 2001. John Wiley & Sons Inc. USA.

> Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. "Applied Linear Statistical Models." 5th Edition. 2005. The McGraw-Hill Companies, Inc. New York, USA.

> Lukasz Komsta and Frederick Novomestky. "moments: Moments, cumulants, skewness, kurtosis and related tests". 2015 R package version 0.14. *https://CRAN.R-project.org/package=moments*

> Measurement Systems Analysis Work Group. "Measurement System Analysis Reference Manual" 4th Edition. 2010. Chrysler Group LLC, Ford Motor Company, and General Motors Corporation. USA.

> Meyer, Karin. "Estimates of genetic parameters for weaning weight of beef cattle accounting for direct-maternal environmental covariances" 1997. Livestock Production Science. 52: 187-199.

> Mrode, R., Thompson, R. "Linear Models for the Prediction of Animal Breeding Values." 3rd Edition. 2014. CAB International. UK

> Nielson, H.B. Jarmer, H.O. "Genome Wide Association Study Identifying Determinants of Facial Characteristics for Facial Image Generation." 2011. World Intellectual Property Organization. PCT/DK2010/050325.

> Obafemi-Ajayi, T., Miles, J.H., Takahashi, N., Qi, W., Aldridge, K. Zhang, M. Xin, S.Q., He, Y., Duan, Y. "Facial Structure Analysis Separates Autism Spectrum Disorders into Meaningful Clinical Subgroups" 2015. Journal of Autism and Developmental Disorders. 45:5:1302\[Dash]1317.

> Ojiem, J.O., de Ridder, N., Vanlauwe, B., Giller, K.E. "Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems." 2006. International Journal of Agricultural Sustainability. 4:1:79\[Dash]93.

> Rashid, M., Gu, X., Lee, Y.J. "Interspecies Knowledge Transfer for Facial Keypoint Detection." 2017. Preprint. *https://arxiv.org/abs/1704.04023*

> Revelle, W. (2017) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, *https://CRAN.R-project.org/package=psych Version = 1.7.3.*

> RSPI Vision. "Deep Learning and Convolutional Neural Networks." 2017. Online. *https://www.rsipvision.com/exploring-deep-learning/*

> R Core Team. "R: A language and environment for statistical computing". 2016 R Foundation for Statistical Computing, Vienna, Austria. URL *https://www.R-project.org/*

> Singh, Bhupendra Pratap. "Imaging Applications of Charge Coupled Devices (CCDs) for Cherenkov Telescope" 2012. Project Report Bhabha Atomic Research Centre. Online. *file:///Users/catie/Downloads/Bhupendra_report.pdf*

> Van Heel, Marin. "Multivariate Statistical Classification of Noisy Images (Randomly Oriented Biological Macromolecules)". 1984. Ultramicroscopy. 13: 165-184.

> Venables, W. N. & Ripley, B. D. "Modern Applied Statistics with S". 4th Edition. 2002. Springer, New York

NOTES