Probable and Improbable Faces

Lewis, John P.
Victoria University

出版情報：MI lecture note series. 50, pp.43-49, 2013-10-21. 九州大学マス・フォア・インダストリ研究所
バージョン：
権利関係：

Probable and improbable faces

J.P. LEWIS
Victoria University, Wellington, New Zealand
(joint work with Ken Anjyo, Zhenyao Mo, Taehyun Rhee)

Problem. In computer vision and graphics research, facial expression and identity are commonly modeled as a high-dimensional vector space, often with a multidimensional Gaussian density. This choice of representation has associated algorithmic approaches such as linear interpolation and maximum a posteriori (MAP) solution of inverse problems.

In this paper we argue several things: 1) the linear and Gaussian assumptions are not strictly correct. 2) existing research that starts from these assumptions has implicitly assumed a low dimensional setting. In high dimensions, common algorithmic approaches such as MAP may not be justified. 3) most importantly, we show that the problems resulting from these assumptions are not just hypothetical, but are visible in a practical computation.

Linear models. The faces of realistic computer characters in movies are most often generated using the "blendshape" representation [LA10, ATL12, SILN11, LYYB13]. This is a linear representation of the form $\mathbf{f} = \mathbf{Bw}$, where $\mathbf{B}$ is a linear but non-orthogonal basis having semantic meaning. Bilinear (tensor) face models have also been proposed [VBPP05]. In computer vision, approaches such as active appearance models (AAM) [CET98] and morphable models [BV99] use an orthogonal basis generated by principal component analysis (PCA), and assume the multidimensional Gaussian prior. Psychological research has also employed such linear models with a multivariate Gaussian prior [Val12].

PCA assumes that the data is jointly Gaussian, in that the PCA basis vectors are the eigenvectors of a covariance matrix that does not capture any non-Gaussian statistics.

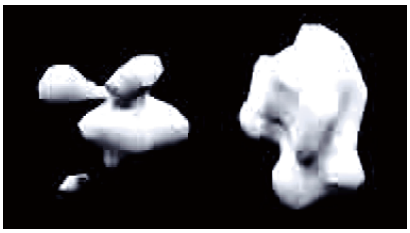

Figure 1. Face proportions are not Gaussian: rendering of the kernel density of several face proportions from a database of 400 faces.

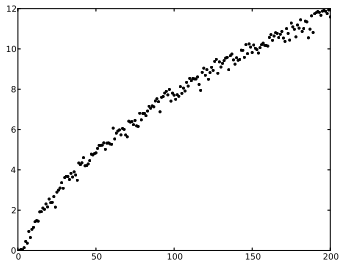

FIGURE 2. The closest distance among 1000 unit-variance multidimensional Gaussian random variables to the mean (vertical axis) as a function of the dimension (horizontal axis). In 100 dimensions every point in this simulation is more than six standard deviations from the mean.

The Gaussian assumption leads to a frequently employed prior or regularizer of the form $\mathbf{c}^T\mathbf{\Lambda}^{-1}\mathbf{c}$ where $\mathbf{c}$ is the vector of PCA coefficients and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues (variances). The Gaussian assumption also naturally leads to the MAP approach to regularising inverse problems. This approach selects model parameters M as the mode of the posterior $P(D|M)P(M)$ given data D. With a Gaussian model the posterior also has a Gaussian form.

The appropriate number of dimensions for a linear facial model of expression or identity has been variously estimated to be in the range 40–100 [MS07, PS00, MXB06]. High quality blendshape facial models used in movie visual effects typically have on the order of 100 dimensions [LA10].

In figure 1 we show that the common multidimensional Gaussian assumption is not strictly accurate. This figure shows a kernel density plot of several simple measurements of facial proportions measured from 400 selected photographs from the facial database [PWHR98]. It is also somewhat obvious that a linear model is not entirely appropriate for facial *expression*. For example, the motion of the jaw has a clear rotational component. On the other hand, the widespread use of the blendshape representation in movies (albeit sometimes with nonlinear correction terms [SILN11]) is an argument that linear models suffice even if they are not strictly accurate. It is less clear whether a vector space model of facial *identity* is appropriate, or if a (nonlinear) manifold assumption would be more accurate. While these comments call into question the linear and Gaussian assumptions, existing research does not indicate whether these objections are important in practical computations.

High-dimensional phenomena. High dimensional data is generally subject to a collection of nonintuitive phenomena collectively known as the "curse of dimensionality" [Wan11]. Examples of such phenomena are that a) in high dimensions, "all data is far away" with high probability (Figure 2), b) randomly chosen vectors are nearly

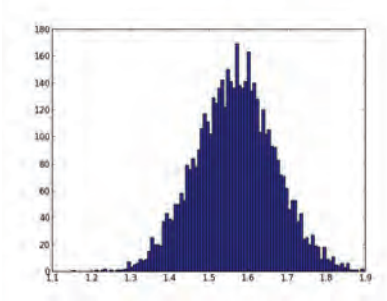

FIGURE 3. In high dimensions, most data are nearly orthogonal. Histogram of the angles between all pairs of 100 randomly chosen isotropic Gaussian random variables in 100 dimensions. The angles cluster around $\pi/2$.

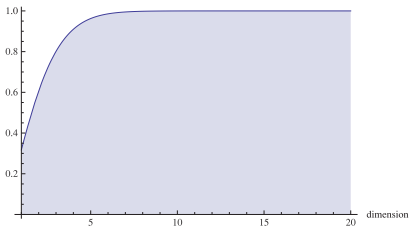

FIGURE 4. Probablility that a sample from a unit variance Gaussian is outside the unit hypersphere for various dimensions.

orthogonal (Figure 3), and c) the probability mass of the data is overwhelmingly located near the surface of the hypervolume, with the interior of the volume essentially empty.

Current face computation approaches generally overlook these phenomena. A notable exception is [PS09], who described the following apparent paradox: the squared Mahalanobis distance $\mathbf{c}^T\mathbf{\Lambda}^{-1}\mathbf{c}$ follows a χ^2 distribution with n degrees of freedom, since it is the sum of i.i.d. squared Gaussian variables of variance $\frac{c_i^2}{\lambda_i}$. The expectation of this distribution for d dimensions is d, thus we expect the length of the standardized squared coefficient vector of a typical face to be d. However under the multidimensional Gaussian model, the face at the origin (the mean face) is the most probable, and the length of its squared coefficient vector is zero.

[PS09] also state a hypothesis that faces should lie on the shell of a hyperellipsoid dictated by the squared coefficient length. While this hypothesis is correct, the noted difference between the mean and the expected Mahalanobis distance does not in itself justify this statement. The resolution to the apparent paradox is simply that it is the

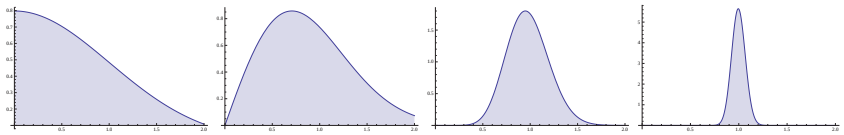

FIGURE 5. The radially integrated Gaussian $N(0, \mathbf{I}_n)$ in various dimensions. Each subfigure shows the radially integrated Gaussian profile $S_{d-1}(r)G(r)$ (vertical axis) plotted in units of $\sqrt{d}$ (horizontal axis). From left to right: 1, 2, 10, and 100 dimensions. In high dimensions the probability concentrates in a shell centered at radius $\sqrt{d}$.

difference between the variance and mean. A zero-mean random variable can (and typically does!) have a nonzero variance. Randomly sampling from a multidimensional Gaussian will generate a sequence of samples that have *both* the expected mean and variance of course.

Discussion. Next we will establish the statement that high dimensional data is concentrated overwhelmingly near the surface of the hypervolume. In the case of a uniformly distributed random variable in a hypercube, this is easy to see. Consider a unit hypercube in d dimensions, that encloses a smaller hypercube of side $1 - \epsilon$. As $d \to \infty$, the volume of the enclosed hypercube is $(1 - \epsilon)^d \to 0$.

The fact that the multivariate Gaussian is a heavy tailed distribution in high dimensions is less obvious. For example, [Val12] states, "even for a face space of high dimensionality, the assumption of a multivariate normal distribution means that... There will be many typical faces that will be located relatively close to the center".

Discussion of the multivariate Gaussian is simplified by a "whitening" transformation $c_i \to c_i/\sqrt{\lambda_i}$ from the original hyperellipsoidal density to an isotropic density. We can also consider a unit-variance density without loss of generality. In this case the probability that a point is within a hypersphere of radius r is proportional to

$$\int_0^r S_{d-1}(r)G(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^r r^{d-1}G(r)dr$$

where d is the dimension, $G(r) = \frac{1}{\sqrt{(2\pi)^d}} \exp^{-r^2/2}$ is the isotropic unit variance Gaussian density function, $S_{d-1}(r) = \frac{2\pi^{d/2}r^{d-1}}{\Gamma(d/2)}$ is the "surface area" of the d-hypersphere, and Γ is the Gamma function. This can be used to plot the tail probability that a point lies outside the unit hypersphere in various dimensions (Figure 4). While in one dimension the majority of the probability mass is within the unit interval, in 100 dimensions the probability that a point is outside the unit hypersphere is 1. to within machine precision! It may be worth contrasting the mode of the high-dimensional Gaussian with the Dirac delta generalised function. The delta function has zero width but unit

volume when integrated over. In contrast, the high-dimensional Gaussian has nonzero width near the origin, but negligible volume.

High dimensional data can also be tightly concentrated in a shell of relatively narrow thickness. In the case of the multi-dimensional Gaussian, the majority of its mass is concentrated within a shell centered at radius $\sqrt{d}$. Figure 5 plots the radially integrated unit variance Gaussian profile $S_{d-1}(r)G(r)$ relative to the distance $\sqrt{d}$ (i.e. with a change of variable $r \to r\sqrt{d}$). The data is concentrated increasingly around $\sqrt{d}$ (relative to the distance $\sqrt{d}$ itself) in high dimensions.

The observations collected above lead to the remarkable conclusion that algorithms such as MAP may be nonsensical in high dimensions! While this conclusion is not widely known in the computer vision and graphics community (MAP is commonly used for face computations with models having 10-100 dimensions), it has been noted elsewhere. Mackay states [Mac96], "probability density maxima often have very little associated probability mass even though the value of the probability density there may be immense, because they have so little associated volume... the locations of probability density maxima in many dimensions are generally misleading and irrelevant. Probability densities should only be maximized if there is good reason to believe that the location of the maximum conveys useful information about the whole distribution."

Example computation: interpolating in face space. Figure 6 contrasts two approaches to interpolating facial identity. The images are not photographs but are synthesized with an AAM [MLN04]. The face on the far left is generated from a coefficient vector $\mathbf{c}_l$ sampled from a multivariate Gaussian with the appropriate variances (eigenvalues). The face on the far right is also randomly chosen, but its coefficient vector $\mathbf{c}_r$ is modified to constrain it to having a specified inner product $\langle \mathbf{c}_l, \mathbf{c}_r \rangle_{\mathbf{\Lambda}^{-1}} = -0.8$ so as to place it on the opposite side of the coefficient volume. The inner product uses the inverse eigenvalue-weighted norm $\langle \mathbf{c}_l, \mathbf{c}_r \rangle_{\mathbf{\Lambda}^{-1}} = \mathbf{c}_l^T \mathbf{\Lambda}^{-1} \mathbf{c}_r$. The dimensionality of the space (length of the coefficient vector) is 181.

The top rows in figure 6 shows linear interpolation through the Gaussian coefficient space. The midpoint of this interpolation passes closer to the center (mean) face than either end. This results in a somewhat "ghostly" face that lacks detail. The linear interpolation also has the undesired result that (for example) interpolating from a person of age 40 to a person of age 45 might pass through an intermediate face of apparent age 25, if that is the mean age of the database underlying the AAM.

In the lower panels of figure 6 we interpolate "around" a hyperellipsoidal shell in the coefficient space rather than across the volume. Given initial and final coefficient vectors $\mathbf{c}_l, \mathbf{c}_r$, at each step a coefficient vector is generated that interpolates the norm of these vectors (although in fact the difference in norm is expected to be small due to phenomena mentioned above). This interpolation remains inside the high probability shell of the hyperGaussian and generates distinctive faces throughout the interpolation.

Conclusion. This paper describes known high-dimensional phenomena that call into question common assumptions underlying much computer vision, graphics, and

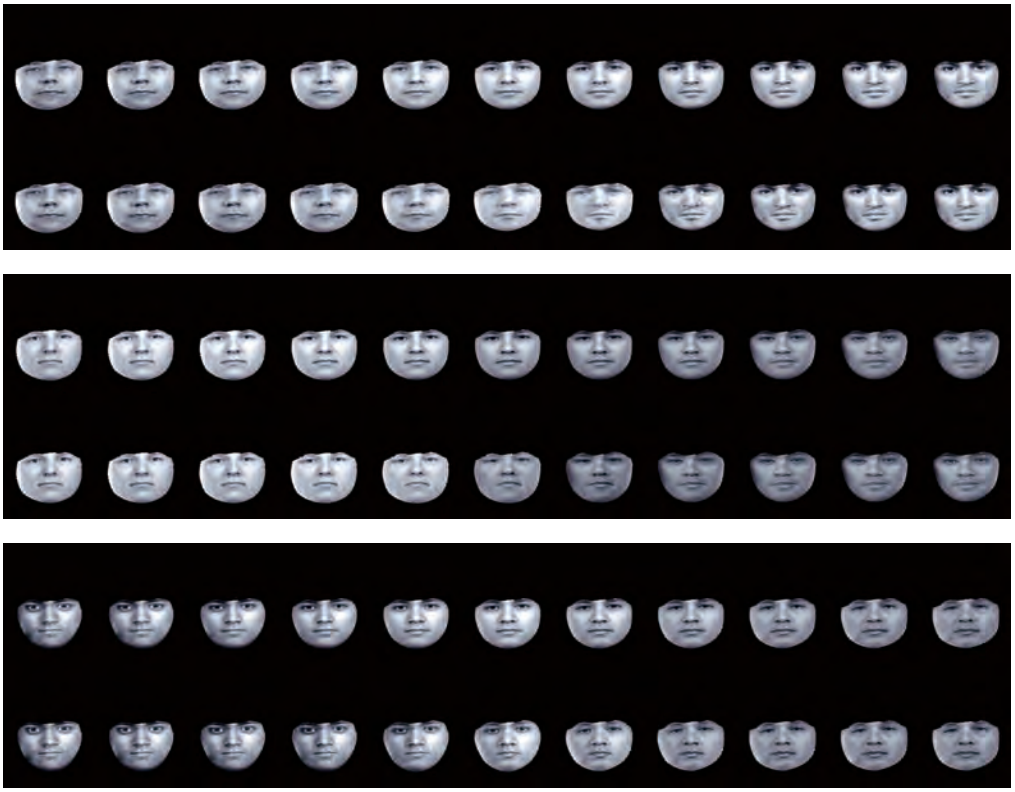

FIGURE 6. Interpolating between a randomly chosen face (left column) and a second face (right column) nearly on the opposite side of the hyperellipse of coefficients. Top row of each image: linear interpolation of coefficients. The middle images lack distinctiveness. Bottom row of each image: interpolating "around the hyperellipse". Detail is preserved throughout the interpolation. Please enlarge to see details.

psychological research on face computation. In particular, we question approaches that assume that typical faces lie in the interior of a high-dimensional Gaussian density. These objections are not merely hypothetical, but are visible in a simple face computation. Our conclusion highlights the need to develop new algorithms that address the intrinsically high-dimensional nature of facial identity and expression.

Acknowledgements. This research is partially supported by the Japan Science and Technology Agency, CREST project. JPL acknowledges a helpful discussion with Marcus Frean.

References

[ATL12] Anjyo K., Todo H., Lewis J.: A practical approach to direct manipulation blendshapes. *J. Graphics Tools 16*, 3 (2012), 160–176.

[BV99] Blanz T., Vetter T.: A morphable model for the synthesis of 3d faces. In *Proceedings of ACM SIGGRAPH* (Aug. 1999), ACM SIGGRAPH, pp. 187–194.

[CET98] Cootes T. F., Edwards G. J., Taylor C. J.: *Active Appearance Models*, vol. 1407 of *Lecture Notes in Computer Science*. 1998.

[LA10] Lewis J., Anjyo K.: Direct manipulation blendshapes. *Computer Graphics and Applications (special issue: Digital Human Faces) 30*, 4 (2010), 42–50.

[LYYB13] Li H., Yu J., Ye Y., Bregler C.: Realtime facial animation with on-the-fly correctives. *ACM Transactions on Graphics 32*, 4 (July 2013).

[Mac96] MacKay D. J.: Hyperparameters: Optimize, or integrate out? In *Maximum entropy and Bayesian methods*. Springer, 1996, pp. 43–59.

[MLN04] Mo Z., Lewis J., Neumann U.: Face inpainting with local linear representations. In *BMVC* (2004), BMVA, pp. 347–356.

[MS07] Meytlis M., Sirovich L.: On the dimensionality of face space. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 7 (2007), 1262–1267.

[MXB06] Matthews I., Xiao J., Baker S.: On the dimensionality of deformable face models. CMU-RI-TR-06-12, 2006.

[PS00] Penev P. S., Sirovich L.: The global dimensionality of face space. In *Proc. 4th Int'l Conf. Automatic Face and. Gesture Recognition* (2000), pp. 264–270.

[PS09] Patel A., Smith W.: 3D morphable face models revisited. *Computer Vision and Pattern Recognition* (2009).

[PWHR98] Phillips P. J., Wechsler H., Huang J., Rauss P.: The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing J. 16*, 5 (1998), 295–306.

[SILN11] Seo J., Irving G., Lewis J. P., Noh J.: Compression and direct manipulation of complex blendshape models. *ACM Trans. Graph. 30*, 6 (Dec. 2011), 164:1–164:10.

[Val12] Valentine T.: *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges*. Scientific Psychology Series. Taylor & Francis, 2012, ch. Face-Space Models of Face Recognition.

[VBPP05] Vlasic D., Brand M., Pfister H., Popovic J.: Face transfer with multilinear models. *ACM Trans. Graph. 24*, 3 (2005), 426–433.

[Wan11] Wang J.: *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*. Springer, 2011.