THE ROLE OF GEOMETRY FOR AGE ESTIMATION

Pavan Turaga, Soma Biswas and Rama Chellappa

Center for Automation Research, University of Maryland, College Park
{pturaga,soma,rama}@umiacs.umd.edu

ABSTRACT

Understanding and modeling of aging in human faces is an important problem in many real-world applications such as biometrics, authentication and synthesis. In this paper, we consider the role of geometric attributes of faces, as described by a set of landmark points on the face, in the perception of age. Towards this end, we show that the space of landmarks can be interpreted as a Grassmann manifold. Then the problem of age estimation is posed as a problem of function estimation on the manifold. The warping of an average face to a given face is quantified as a velocity vector that transforms the average to a given face along a smooth geodesic in unit-time. This deformation is then shown to contain important information about the age of the face. We show in experiments that exploiting geometric cues in a principled manner provides comparable performance to several systems that utilize both geometric and textural cues. We show results on age estimation using the standard FG-Net dataset and a passport dataset which illustrate the effectiveness of the approach.

Index Terms— Face Geometry, Age Estimation, Grassmann manifold, Regression

1. INTRODUCTION

The modeling of the appearance of human faces is an important component in several applications such as biometrics, animation, and picture annotation. Aging is a source of variation which has only recently been gaining attention. Understanding the appearance variations induced by aging is important for applications where the claimed identity and the enrolled face may show a large difference in apparent age. Studies in neuroscience have shown that facial geometry is a strong factor that influences age perception [1]. In [1], it is shown that shape-averaged faces are perceived to be younger. Further, the 'distance' from the average is a strong indicator of the apparent age of the person. The regions where a given face shows a large difference in shape from a shape-averaged face when further exaggerated, results in a caricature [2]. Young faces exhibit distinct growth-related anthropometric trends. Anthropometric variations in adults are distinctive to a lesser degree than in children, but nevertheless they do exhibit drifts in facial features surrounding the mouth, eyebrows etc [3]. This is illustrated in figure 1 where distinct geometric changes can be observed as a person ages.

By facial geometry we refer to the location of 2D facial landmarks on images. Several models have been proposed that explain geometric variations with age. Examples include the works of [4, 5] who proposed various physics based models to explain craniofacial growth. More recently, Narayanan and Chellappa [6] applied these models in conjunction with anthropometric data to identify different growth parameters for different parts of the face. Physics-based approaches such as these have mostly found use in synthesis applications such as age progression and regression, where it is important to synthesize realistic younger or older looking faces.

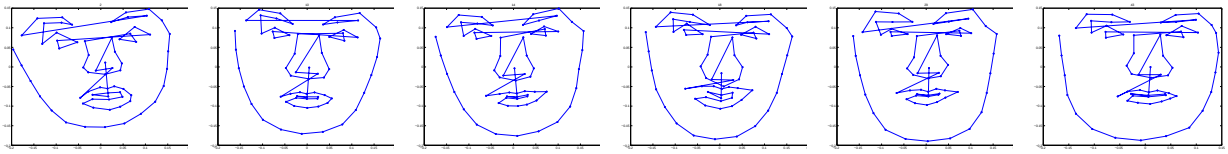
In this paper, we consider the problem of 'age-estimation'. Several regression-based methods have been proposed to estimate the age of a face from images. Lanitis et al. [7] constructed an aging function based on a parametric model for human faces and performed automatic age progression, age estimation, face recognition across aging. Fu et al. [8] combined dimensionality reduction methods such as PCA, LLE, LPP, OLPP etc with regression. Guo et. al. [9] proposed robust regression followed by local adjustments for age estimation and showed that local adjustments improve performance. All these approaches mainly differ in the features used and variations in the choice of regression methods. However, the relative roles of geometry and texture have not been studied in detail.

In this paper, we show that a principled approach to model facial geometry can play a significant role in modeling facial aging. Specifically, we show how to characterize the 'space' of these facial landmarks. We demonstrate that even without relying on texture, the results are comparable to several complex and optimized state-of-the-art systems, and even outperform many of them. Thus, this can form the basis of a more principled approach to facial geometric modeling that can be optimized to reach even higher performance levels in several applications.

Organization of Paper: In section 2, we discuss an affine-invariant approach to modeling facial geometry as a Grassmann manifold. In section 3, we review in brief the differential geometric tools needed to study the aging problem on the Grassmann manifold. In section 4 we present experiments, and conclusions are given in section 5.

2. MODELING THE GEOMETRY OF THE FACE

In this paper, we are interested in the 2D geometry of facial landmarks. The shape observed in an image of a face is a perspective projection of the 3D locations of the landmarks. Standard approaches to describe shapes involve extracting features such as shape context [10] etc. These approaches extract coarse features which correspond to the average properties of the shape. These approaches are particularly useful when landmarks on shapes cannot be reliably located across different images or do not necessarily correspond to physically meaningful parts of the object. However, in the case of faces, there exist physically meaningful locations such as eyes, mouth, nose etc which can be reliably located on most faces [11]. This suggests the use of a representation that exploits the entire information offered by the location of landmarks instead of relying on coarse features. There exist several automatic methods to locate facial landmarks which work well on constrained images such as passport photos (c. f. [12]). It is in constrained scenarios such as these that the methods proposed here are applicable.



(a) Age 2 (b) Age 10 (c) Age 14 (d) Age 18 (e) Age 29 (f) Age 43

Fig. 1. Facial geometric variation across ages. Samples shown correspond to individual 2 from the FG-net dataset.

The drawback of using the locations of landmarks is that they are sensitive to transformations such as affine transforms, view changes etc. In order to account for this, shape theory studies the equivalent class of all configurations that can be obtained by a specific transformation (e.g. linear, affine, projective) from a given base shape. A shape is represented by a set of landmark points, given by a $m \times 2$ matrix $L = [(x_1, y_1); (x_2, y_2); \ldots; (x_m, y_m)]$, of the set of m landmarks of the centered shape. The *shape space* of this base shape is the set of equivalent configurations that are obtained by transforming the base shape by an appropriate spatial transformation. For example, the set of all affine transformations forms the *affine shape space* of that base shape.

The *affine shape space* [13] is very important because small changes in camera location or change in the pose of the subject can be approximated well as affine transformations on the original base shape. The affine transforms of the shape can be derived from the base shape simply by multiplying the shape matrix L by a 2×2 full rank matrix on the right. For example, let A be a 2×2 affine transformation matrix i.e. $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Then, all affine transforms of the base shape L_{base} can be expressed as $L_{affine}(A) = L_{base} * A^T$. Note that, multiplication by a full-rank matrix on the right preserves the column-space of the matrix L_{base}. Thus, the 2D subspace of $\mathbb{R}^m$ spanned by the columns of the matrix L_{base} is an *affine-invariant* representation of the shape. i.e. $span(L_{base})$ is invariant to affine transforms of the shape. Subspaces such as these can be identified as points on a Grassmann manifold. We now define the Grassmann manifold.

The Grassmann manifold $G_{k,m}$ is the space whose points are *k-planes* or *k*-dimensional hyperplanes (containing the origin) in $\mathbb{R}^m$ [14]. To each k-plane ν in $\mathbb{R}^m$, we can associate an $m \times k$ orthonormal matrix Y such that the columns of Y form an orthonormal basis for the plane. However, since the choice of basis is not unique, we need to define an equivalence class of orthonormal basis vectors which span the same subspace. Hence, to each k-plane ν in $G_{k,m}$ is associated an equivalence class of $m \times k$ matrices YR in $\mathbb{R}^{m\times k}$, for $R \in SO(k)$, where Y is an orthonormal basis for the k-plane.

The tall-thin Procrustes representation is most widely used due to its computational efficiency as in [15]. In the current scenario, $k = 2$ and m is the number of facial landmarks. The basic premise of our work is that the perceived age will show a functional dependence on the geometry of the face. Given several faces X_i, along with their respective ages y_i, the goal is to estimate a function $y = f(X)$ that can explain the aging patterns. This can be formulated as a regression problem. Regression problems are mostly studied in Euclidean vector spaces and there exist a wealth of methods for robust regression. Regression has been applied to age-estimation tasks before by assuming that faces, or features extracted from faces lie in a Euclidean space such as in [7, 8, 9]. However, for geometric features considered here, we need to solve the regression problem on the Grassmann manifold. The Grassmann manifold is not a vector space, thus precluding the use of classical techniques. To solve this problem we use the differential geometry of the manifold. All points on the manifold are projected onto the tangent plane at a mean-point and standard vector-space methods are applied on the tangent plane, which is a Euclidean vector space.

3. GEOMETRY OF THE GRASSMANN MANIFOLD

The Grassmann manifold $G_{k,m}$ can be viewed as a quotient group of the orthogonal group $SO(m)$. An in depth treatment of this subject can be found in [14]. Briefly, geodesic paths on $SO(m)$ are given by one-parameter exponential flows $t \rightarrow exp(tB)$, where $B \in \mathbb{R}^{m\times m}$ is a skew-symmetric matrix. The quotient geometry of the Grassmann manifold implies that geodesics in $G_{k,m}$ are given by one-parameter exponential flows $t \rightarrow exp(tB)$ where B has a more specific structure given by $B = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}$, where $A \in \mathbb{R}^{(m-k)\times k}$. The matrix A parameterizes the direction and speed of geodesic flow. Given a point on the Grassmann manifold S_0 represented by orthonormal basis Y_0, and a direction matrix A, the one-parameter geodesic path emanating from Y_0 in direction B is given by $Y(t) = Q\, exp(tB)\, J$, where, $Q \in SO(m)$ and $Q^T Y_0 = J$ and $J = [I_k; 0_{m-k,k}]$.

Given an 'average-face' or a shape-normalized face, we would like to quantify the deformation that can warp the average to any given face. We can conveniently model these deformations via geodesics on the Grassmann manifold. To represent the average face, we may either choose a generic face as the average, or more generally compute it from a database of faces. In this case, we need to compute a mean point from a collection of points on the Grassmann manifold. The can be done by computing the 'Karcher' mean [16]. Given a set of subspaces $\{S_i\}$, the mean subspace is one that minimizes

$$\mu = \min_{S\in G_{k,m}} \sum_{i=1}^{n} d^2(S_i, S), \tag{1}$$

where $d(S_i, S)$ is the geodesic length between the points. This problem has a unique solution if the points S_i are clustered close together on the manifold. This is a reasonable assumption in our case since all faces share some global similarities. However, there is no closed form solution to the problem. An iterative algorithm to compute it is as follows. Start with an initial guess $\mu^{(0)}$. Choose an updating parameter $\tau \in [0, 1]$, stopping criterion $\delta > 0$, and maximum iterations N_{max}. Then,

1. Compute the velocity parameter A_i that takes $\mu^{(k)}$ to S_i for each i. Compute the average velocity matrix $\overline{A} = \frac{1}{n}\sum_i A_i$.
2. Update the mean by moving $\mu^{(k)}$ in the direction of $\overline{A}$ for time $t = \tau$ along the manifold. Let the point thus obtained be denoted $Y(\tau)$.
3. Set $\mu^{(k+1)} = Y(\tau)$ as the new estimate of the mean. Go back to step 1 and repeat till convergence i.e. $d(\mu^{(k)}, \mu^{(k+1)}) \leq \delta$ or till the maximum iterations are exceeded.

Once the average face is computed, we parametrize all faces by the velocity vector that transforms the average face to a given face in unit-time. For all these computations, there exist numerically efficient algorithms and we refer the reader to [17] for details. Here, we adopt the algorithms proposed in [17]. We shall use these velocity parameters as aging signatures. Once these velocity parameters are computed, we can *flatten* them to a vectorial form. Once this is done, we can apply standard Euclidean space regression methods on the velocity parameters.

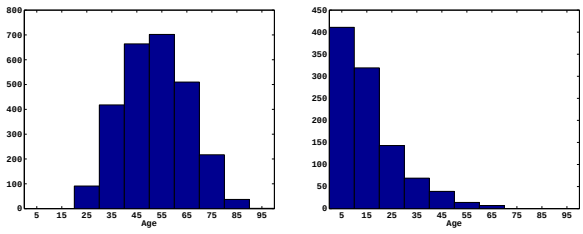
4. EXPERIMENTS

We evaluate the strength of the geometric attributes on age-estimation tasks on two datasets. The first dataset is the Passport dataset [6] which contains mostly adult faces. The age distribution of the faces is shown figure 2(a). In this dataset, we used 47 fiducial points marked manually. The second is the publicly available FG-Net dataset [18], which contains both adult and young faces. The distribution of ages is shown in figure 2(b). Some sample images from this dataset are shown in figure 3. For this dataset, 68 fiducial points are available with each face.

Given a face and its landmarks, we extract the tall-thin orthonormal representation using SVD. Given the matrix of landmarks $L = [(x_1, y_1); (x_2, y_2); \ldots; (x_m, y_m)]$, we compute its rank-2 SVD $L = U\Sigma V^T$. The affine-invariant Grassmann representation of L is then given by $Y_L = U$. Now given several examples Y_i with corresponding ages y_i, we want to estimate the aging-function $y = f(Y)$ in a robust manner. Given a training set, we compute the shape-normalized face μ. For each face in the training set Y_i, we compute the velocity vector A_i as described in section 3. Then, we estimate the aging function $y_i = f(A_i)$ using standard regression methods.

For performing regression using the velocity vectors, we use ϵ-SVMs, RVMs, and ridge regression (regularized linear least-squares). We use the ϵ-SVM with $\epsilon = 0.02$, the cost parameter $C = 1000$, and regularization parameter $\lambda = 10^{-6}$. For RVMs, there are no parameters to tune except the number of iterations for the RVM optimization routine. We set this to 50 iterations. For ridge regression, the regularization parameter λ is chosen to be $\lambda = 10^{-6}$. We use the polynomial kernel of degree 2 i.e. $K(A_1, A_2) = (1 + A_1^T A_2)^2$, where A_1 and A_2 are the vectorial forms of the velocity matrices.

Two metrics have been proposed in literature for quantifying the performance of age-estimation algorithms. The first criterion measures the mean absolute error (MAE) in age-estimation across the entire dataset. i.e. $MAE = \frac{1}{N}\sum_i |l_i - \hat{l}_i|$, where N is the size of the dataset, l_i is the true age of the i^{th} person being tested, and $\hat{l}_i$ is the assigned age. The second metric is the cumulative match score. The cumulative score is defined as $CS(j) = N_{e\leq j}/N \times 100\%$, where $N_{e\leq j}$ is the number of test-images on which the absolute error in age-estimation is within j years.



(a) (b)

Fig. 2. Distribution of ages in (a) Passport, (b) FG-Net dataset

Fig. 3. Sample images of the same subject from the FG-Net dataset.

Passport dataset: In the passport dataset, we performed a leave-one-out testing in which the regression algorithms are trained on the entire dataset except one sample on which the testing is done. The MAE results using various algorithms is summarized in table 1. The SVM and RVM based regression are seen to perform better than the simpler ridge-regression. We see that the lowest MAE was achieved by using velocity vectors with RVM regression and it is 8.84 years. Considering that the average age in this dataset is 42 years, the obtained MAE is quite encouraging. Figure 4 shows the cumulative score curves as a function of the error-level. We see that about 85% of the faces are classified within 15 years of the true age.

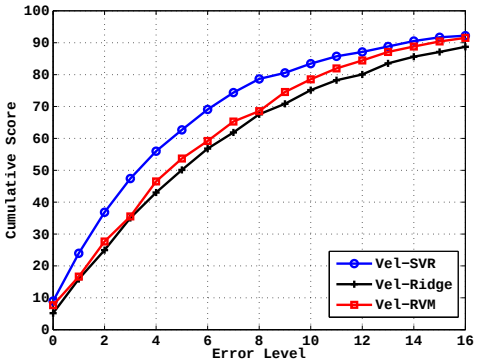


Fig. 4. FG-Net data Cumulative scores using velocity parameters with polynomial kernel.

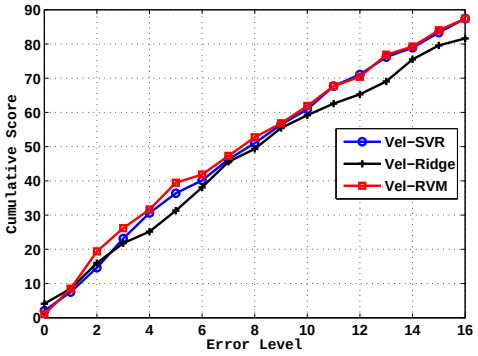


Fig. 5. Passport data Cumulative scores using velocity parameters with polynomial kernel.

FG-Net dataset: For the FG-Net dataset, we performed a leave-one-person-out testing as has recently been suggested [19]. In this mode, all images corresponding to the same person are used for testing and the remaining images are used for training. The results of the proposed framework on the FG-Net dataset is shown in table 2. The lowest MAE was obtained by using SVM + polynomial kernel on velocity vectors. MAE in this case was 5.89 years. The table also shows a comparison with other recently published methods. The cumulative scores of the proposed methods is shown in figure 4. We see that more than 90% of the faces are classified within 15 years of their true age.

The methods compared in table 2 rely on joint structure and tex-

ture information such as using the entire images, or using Active Appearance models. Inspite of not relying on textural cues, it is worth noting that accurate characterization of geometry provides better results in many cases. This however does not downplay the role of texture in age-perception, and the proposed methods may be further combined with textural features. We see that the proposed approach is comparable to the state-of-the-art methods and even outperform most of them except RUN1 [20] (MAE = 5.78) and LARR [9] (MAE = 5.07). Also, it is worth noting that since geometric variations are more pronounced in children than adults, it might explain why the age-estimation error in the passport dataset is larger than in the FG-Net dataset.

Method	**Ridge Regression**	**SVM**	**RVM**
Warping Velocities	15.72	9.78	8.84

Table 1. Mean-Absolute Errors using different regression methods using the warping velocities on the Passport dataset.

	Method	**MAE**
Warping Velocities	Ridge	7.57
	SVM	5.89
	RVM	6.69
Other Algorithms	AAS [21]	14.83
	WAS [19]	8.06
	Ages [19]	6.77
	$Ages_{lda}$ [19]	6.22
	QM [7]	6.55
	MLP [7]	6.98
	RUN1 [20]	5.78
	LARR [9]	5.07

Table 2. Comparison of Mean-Absolute Errors using proposed methods with state-of-the-art on the FG-Net dataset.

5. DISCUSSION AND CONCLUSION

We have shown how accurate characterization of the space of landmarks leads to mathematically sound methods for facial geometric modeling. We have shown the strength of the framework on age-estimation tasks on two datasets. The proposed methods are comparable to the best available methods. Thus, this can form the basis of a more principled framework for modeling facial geometry.

6. REFERENCES

[1] A. J. O'Toole, T. Price, T. Vetter, J. C. Bartlett, and V. Blanz, "Three-dimensional shape and two-dimensional surface textures of human faces: The role of "averages" in attractiveness and age," *Image and Vision Computing Journal*, vol. 18, no. 1, pp. 9–19, 1999.

[2] S. E. Brennan, "The caricature generator," *Leonardo*, vol. 18, no. 3, pp. 170–178, 1985.

[3] A. M. Alberta, K. Ricanek, and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Science International*, vol. 172, no. 1, pp. 1–9, 2007.

[4] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events : Implications for a theory of nonrigid shape perception," *Journal of Experimental Psychology : Human Perception and Performance*, vol. 1, no. 4, pp. 374–382, 1975.

[5] J. T. Todd, L. S. Mark, R. E. Shaw, and J. B. Pittenger, "The perception of human growth," *Scientific American*, vol. 242, no. 2, pp. 132–144, 1980.

[6] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 387–394, 2006.

[7] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.

[8] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," *International Conference on Multimedia and Expo*, pp. 1383–1386, 2007.

[9] G. D. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1178–1188, July 2008.

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *PAMI*, vol. 24, no. 4, pp. 509–522, 2002.

[11] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?," *Computer Vision and Image Understanding*, vol. 102, no. 2, pp. 117–133, 2006.

[12] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[13] C. R. Goodall and K. V. Mardia, "Projective shape analysis," *Journal of Computational and Graphical Statistics*, vol. 8, no. 2, 1999.

[14] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal Matrix Analysis and Application*, vol. 20, no. 2, pp. 303–353, 1999.

[15] P. Turaga, A. Veeraraghavan, and R. Chellappa, "Statistical Analysis on Stiefel and Grassmann Manifolds with Applications in Computer Vision," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[16] Karcher, H., "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, pp. 509–541, 1977.

[17] K. Gallivan, A. Srivastava, X. Liu, and P. VanDooren, "Efficient algorithms for inferences on grassmann manifolds," *12th IEEE Workshop Statistical Signal Processing*, 2003.

[18] "The fg-net aging database," *Available: http://www.fgnet.rsunit.com/*.

[19] X. Geng, Z. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

[20] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," *IEEE International Conference on Computer Vision*, , no. 7, pp. 1–8, 2007.

[21] A. Lanitis, C. Draganova, and C. Christodoulo, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Systems, Man and Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.