

Parametric Manifold of an Object under Different Viewing Directions

Xiaozheng Zhang^{1,2}, Yongsheng Gao^{1,2}, and Terry Caelli³

¹ Biosecurity Group, Queensland Research Laboratory, National ICT Australia
{paul.zhang,yongsheng.gao}@nicta.com.au

² Computer Vision and Image Processing Lab, Griffith University, Brisbane, Australia
{x.zhang,yongsheng.gao}@griffith.edu.au
<http://maxwell.ict.griffith.edu.au/cvipl/>

³ Victoria Research Laboratory, National ICT Australia
terry.calli@nicta.com.au

Abstract. The appearance of a 3D object depends on both the viewing directions and illumination conditions. It is proven that all n -pixel images of a convex object with Lambertian surface under variable lighting from infinity form a convex polyhedral cone (called illumination cone) in n -dimensional space. This paper tries to answer the other half of the question: What is the set of images of an object under all viewing directions? A novel image representation is proposed, which transforms any n -pixel image of a 3D object to a vector in a $2n$ -dimensional pose space. In such a pose space, we prove that the transformed images of a 3D object under all viewing directions form a parametric manifold in a 6-dimensional linear subspace. With in-depth rotations along a single axis in particular, this manifold is an ellipse. Furthermore, we show that this parametric pose manifold of a convex object can be estimated from a few images in different poses and used to predict object's appearances under unseen viewing directions. These results immediately suggest a number of approaches to object recognition, scene detection, and 3D modelling. Experiments on both synthetic data and real images were reported, which demonstrates the validity of the proposed representation.

Keywords: pose manifold, 3D object, in-depth rotations, viewing directions, appearance prediction, object rendering.

1 Introduction

One of the major tasks of computer vision is to recognize objects through their appearances. Because the objects are usually 3 dimensional and their images are 2 dimensional, their appearances in images vary due to the changes in viewpoints and illumination conditions.

The image changes due to illumination variations have been extensively studied in the past two decades. Shashua [13] proved the images of a Lambertian object under all possible light source directions form a 3D linear subspace in n -dimensional image space. Belhumeur and Kriegman [2] pointed out that the image set is actually

m -dimensional *illumination cone* where m is the number of distinct surface normals. Basri and Jacobs [1] approximated the illumination cone with a 9-dimensional linear subspace with spherical harmonics approximation on light sources. With these encouraging results, the images of an object under all possible illumination conditions can be predicted and represented by a few base images. Consequently, the recognition tasks with illumination variations have been greatly facilitated.

Similarly, researchers searched for viewpoint (or pose) linear subspaces to simplify the pattern recognition problems under viewpoint variations. Because viewpoint and pose have the same effect in images of an object, this paper uses them interchangeably. Prince et al. [11] approximated pose variations in image space as non-linear transformations. Active appearance models [7] and eigen light fields [9] predicted novel appearances of a human face from exemplar appearances based on cross-identity similarity in human faces. These techniques rely heavily on within-class similarity which are not generalizable and may reduce between-class separability desired in recognition tasks.

Explicit 3D models have been used in the 2D pose-related recognition. Generally, a synthesis-and-analysis strategy is applied, which estimates 3D models from 2D input image(s). Then pattern recognition is conducted either in 3D space [3] or in 2D image space with rendered views [4]. In structure-from-motion (SfM), rigid pose variation is modeled as an affine transform in 2D images [15]. Shashua studied the inter-point relationships in pose variant images [12]. Ullman and Basri [15] proved that an image of a generally rotated 3D point set can be linearly represented by three other views with 6 coefficients and 3 constraints. In [8], level sets of motion field on images have been studied under perspective projection.

This paper addresses the fundamental issues of representing objects under varying viewing directions. Different from [2, 11] of treating images in n -dimensional image space, 2D images are considered as projected 3D points with textures and are transformed into vectors in a $2n$ -dimensional pose space. We attempt to answer the question: What is the set of all images under all poses in \mathbb{R}^{2n} ? Because pose variations at most involve a few rotation, translation, and scaling parameters, this set is compact and should be able to represent parametrically. Is this set able to be determined by a small number of exemplar images? If so, how many images are needed?

Considering the projected point sets of an object with n points as a $2n$ -D point, each pair of axes corresponds to a point in the image and the values are its coordinates (x - and y -coordinates). This paper proves: 1) Under all in-depth rotations along a fixed axis (e.g., y -axis), the images of an object in pose space form an elliptic pose manifold of a 2D pose linear subspace. 2) 2 images of a rotated object are enough to determine this pose ellipse, given the rotation angle between the two images. 3) Under all possible rotations, the images form a parametric pose manifold in a 6D pose linear subspace, whose two 3D projections are ellipsoids. We show experimental investigations on both synthetic and real data to demonstrate the validity of the proposed manifold representations.

Different from existing investigations in the fields of structure-from-motion [12, 15] and 3D model indexing [6, 10], this paper limits the transformations of an 3D object within pure rotations. Though allowance of general affine transformations makes the problem a linear problem, it introduces non-linear stretching and shearing which are not valid for transforming rigid 3D objects. Pure rigid rotations are a subset of affine transformations and are seemingly simpler. However, considering pure rotations only involves non-linear transformations due to the fact that 3D rotations form a special

rotational group and do not span the full 3×3 transformation linear space. We further show that under pure rotations, the points on the parametric manifold and the projected 2D images have one-to-one correspondences. Furthermore, instead of using sparse feature points as in [6, 10, 12, 15], this paper considers any 2D image of a 3D object as a dense set of points (or pixels) in the parametric manifold representation. Consequently, reconstructing an existing 2D image and rendering a virtual image under a novel viewing direction are feasible.

2 Parametric Pose Manifold

This section describes the proposed parametric pose manifold representations. We make the following assumptions: First, the 3D object has a rigid body and the projection from 3D to 2D space is orthographical. Second, no point is occluded in any images under different viewing directions. Third, only pure rotations about a single rotation center are considered and no translation/scaling is involved. All these three assumptions will be further discussed in Section 4 with possible solutions and future extensions of this paper.

2.1 3D-2D Pose Linear Subspace

Let $A \in \mathbb{R}^{3 \times n}$ be a matrix where each column of A is 3 coordinates in x , y , and z axes, respectively, and n is the number of points in a 3D object. Let $R \in \mathbb{R}^{3 \times 3}$ be a rotation matrix and $P \in \mathbb{R}^{2 \times 3}$ be a projection matrix which projects 3D data onto a 2D plane. Without loss of generality, we let the image plane be xy plane. Because the projection is orthographical, the projection matrix is

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (1)$$

Then the rotation and projection operation can be characterized by a “pose” matrix $R' = PR \in \mathbb{R}^{2 \times 3}$. Let $B \in \mathbb{R}^{2 \times n}$ be the matrix of pose R' mapped points on the 2D image plane, where each column contains x and y coordinates. Under orthographical projection, B is given by $B = R' A$. It's shown in [15] that x (or y) coordinates of a 3D object in such a case lie in a 3D linear subspace. Combining the two coordinates in the $2n$ -dimensional pose space, we have the following proposition.

Proposition 1. *The set $\mathcal{B} = \{B | B = R' A, \forall R' \in \mathbb{R}^{2 \times 3}\}$ forms a 6-dimensional linear subspace in the $2n$ -dimensional space.* \square

This 6-dimensional linear subspace is named as *3D to 2D pose linear subspace* of set A and we use $\mathcal{L}_3^2(A)$ to denote it. With Proposition 1, we are able to linearly represent any image under an arbitrary viewing direction using at most 6 basis “images”. These 6 basis “images” can be learned from 3 images of the rotated objects which are linearly independent. Pose matrices caused by pure rotations and orthographical projections do not span the entire 6-dimensional space $\mathbb{R}^{2 \times 3}$. Therefore, the basis “images” are not necessarily valid images because the 3D object is assumed rigid.

It was argued in [13] that the explicit representations are not necessary, due to the existences and allowances of non-rigid transformations which together with pure

rotations span the entire 6D space. For rigid objects, however, these non-rigid transformations are not valid and introduce false object representations (images). This paper derives this explicit parametric manifold representation and shows the necessity and advantages of the representation in image representations and appearance prediction.

2.2 2D-1D Elliptic Pose Manifold

Considering a rotation in 3D space, we have $R \in SO(3)$ so that $R^T R = I$ and $|R| = 1$. It has 3 degrees of freedom, e.g., the Euler angles (ϕ, θ, ψ) . Thus the 6 basis matrices R'_1, \dots, R'_6 above are not all independent and the dimension of pose subspace could be further reduced to as low as 3. This subsection starts with 2D rotations and 2D-to-1D projection. 3D cases will be discussed in Section 2.4 and 2.5.

From Proposition 1, we have the following proposition.

Proposition 2. *Given an n -point 2D coordinate matrix $A \in \mathbb{R}^{2 \times n}$, the pose projected set $\mathcal{B} = \{B | B = R' A, \forall R' \in \mathbb{R}^2\}$ forms a 2-dimensional linear subspace in n -dimensional space. \square*

Lemma 1. *If the two rows of A are orthogonal to each other, \mathcal{B} forms an ellipse in the pose linear subspace.*

Proof. Let $\mathbf{x} = [x_1, \dots, x_n]$ and $\mathbf{y} = [y_1, \dots, y_n]$ be the first and second rows of A , respectively, and we have $\mathbf{x}\mathbf{y}^T = 0$.

Let $R'_1 = [1, 0]$ and $R'_2 = [0, 1]$ be two bases for the pose matrices $\{R' | R' = PR, R \in SO(2), P = [1, 0]\}$. Then we have $B_1 = R'_1 A = \mathbf{x}$, $B_2 = R'_2 A = \mathbf{y}$, and $B_1 B_2^T = \mathbf{x}\mathbf{y}^T = 0$. Normalizing B_1 and B_2 , two orthogonal bases are obtained as $\hat{B}_1 = B_1/|B_1|$ and $\hat{B}_2 = B_2/|B_2|$ in $\mathcal{L}_2^1(A)$.

$\forall \theta$, B is calculated as $B = B_1 \cos \theta + B_2 \sin \theta = |B_1| \cos \theta \hat{B}_1 + |B_2| \sin \theta \hat{B}_2$. Projecting B onto $\mathcal{L}_2^1(A)$, we get the two coordinates $v_1 = |B_1| \cos \theta$ and $v_2 = |B_2| \sin \theta$. Let $a = |B_1|$, $b = |B_2|$, so that

$$\frac{v_1^2}{a^2} + \frac{v_2^2}{b^2} = 1. \quad \square \quad (2)$$

Next, we extend it to more general case where the two rows of A are not orthogonal.

Theorem 1. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, point set $A = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$, $\mathcal{B}(A)$ forms an ellipse in $\mathcal{L}_2^1(A)$.

Proof. Let $\alpha = \frac{1}{2} \tan^{-1} \frac{2\mathbf{x}\mathbf{y}^T}{\mathbf{x}\mathbf{x}^T - \mathbf{y}\mathbf{y}^T}$, $\mathbf{u}_1(\alpha) = \mathbf{x} \cos \alpha + \mathbf{y} \sin \alpha$, $\mathbf{u}_2\left(\alpha + \frac{\pi}{2}\right) = -\mathbf{x} \sin \alpha + \mathbf{y} \cos \alpha$. We have $\mathbf{u}_1 \mathbf{u}_2^T = 0$. Then, we can select $\hat{\mathbf{u}}_1 = \frac{\mathbf{u}_1}{|\mathbf{u}_1|}$ and $\hat{\mathbf{u}}_2 = \frac{\mathbf{u}_2}{|\mathbf{u}_2|}$ to be a set of orthogonal bases in \mathcal{B} . Projecting an arbitrary point $B(\theta) = (\mathbf{x} \cos \theta + \mathbf{y} \sin \theta) \in \mathcal{B}$ onto these bases results in

$$\mathbf{v}_1 = \frac{\hat{\mathbf{u}}_1}{|\mathbf{u}_1|} [(\mathbf{x}\mathbf{x}^T - \mathbf{x}\mathbf{y}^T \tan \alpha) \cos \theta \cos \alpha + (\mathbf{y}\mathbf{y}^T + \mathbf{x}\mathbf{y}^T \cot \alpha) \sin \theta \sin \alpha]. \quad (3)$$

Because $\tan 2\alpha = \frac{2}{\cot \alpha - \tan \alpha} = \frac{2\mathbf{x}\mathbf{y}^T}{\mathbf{x}\mathbf{x}^T - \mathbf{y}\mathbf{y}^T}$, we have the following equality.

$$\mathbf{x}\mathbf{x}^T - \mathbf{x}\mathbf{y}^T \tan \alpha = \mathbf{y}\mathbf{y}^T + \mathbf{x}\mathbf{y}^T \cot \alpha \doteq a|\mathbf{u}_1|. \quad (4)$$

Then $\mathbf{v}_1 = a \cos(\theta - \alpha) \hat{\mathbf{u}}_1$. Similarly, we have $\mathbf{v}_2 = b \sin(\theta - \alpha) \hat{\mathbf{u}}_2$, where $b = \frac{1}{|\mathbf{u}_2|} (\mathbf{x}\mathbf{x}^T - \mathbf{x}\mathbf{y} \cot \alpha)$. On $\mathcal{L}_2^1(A)$, the coordinates (v_1, v_2) satisfies Eq. 2. \square

We name this elliptic manifold *2D-1D pose manifold* and denote it as $\mathcal{M}_2^1(A)$. $\mathcal{M}_2^1(A) \subset \mathcal{L}_2^1(A)$ and it's a planar ellipse.

2.3 3D-1D Pose Manifold of a 3D Object Rotated along Y-Axis

Theorem 1 can be extended to a special case of 3D in-depth rotations along a single axis. Without loss of generality, it is assumed in this subsection the rotation axis is y-axis. Due to orthographical projection, 2D images of such rotated points have the same y-coordinates. This implies that adding n dimensions of y-coordinates to \mathcal{L}_2^1 and forming \mathcal{L}_3^2 does not change pose manifold \mathcal{M} . Therefore, we have the following proposition.

Proposition 3. *All images of a 3D object in 2n-dimensional pose space under all in-depth rotations along a single axis form an elliptic pose manifold.* \square

Proposition 3 provides a clear answer to the question, what the set of images of an object under all in-depth rotations along a single axis is. The set is an elliptic manifold in 2D pose linear subspace, which can be expressed using a single parameter, rotation angle θ .

2.4 Pose Manifold of a 3D Object under Arbitrary Rotations

In this subsection, we develop the ultimate pose manifold which describes the general 3D rotations of an object. We begin with 3D-1D case. The rotation group $\mathcal{R} = SO(3)$ is generated by Euler angles (ϕ, θ, ψ) so that $\forall R \in \mathcal{R}, \exists \phi, \theta, \psi \in \mathbb{R}$, such that $R = R_z(\psi)R_y(\theta)R_x(\phi)$ where $R_z(\psi), R_y(\theta), R_x(\phi)$ are rotation matrices in terms of x, y, z axes, respectively.

Lemma 2. *If the rows of A are orthogonal to each other, \mathcal{B} forms an ellipsoid pose manifold in the pose linear subspace.*

Proof. Let $\mathbf{x} = [x_1, \dots, x_n]$, $\mathbf{y} = [y_1, \dots, y_n]$, and $\mathbf{z} = [z_1, \dots, z_n]$ be the three rows of A , respectively. We have $\mathbf{x}\mathbf{y}^T = 0$, $\mathbf{x}\mathbf{z}^T = 0$, and $\mathbf{y}\mathbf{z}^T = 0$.

Let $R'_1 = [1, 0, 0]$, $R'_2 = [0, 1, 0]$, and $R'_3 = [0, 0, 1]$ be three bases for the pose matrices $\{R' \mid R' = PR, R \in SO(3), P = [1, 0, 0]\}$. R'_1 , R'_2 , and R'_3 correspond to Euler angles $(0, 0, 0)$, $(\frac{\pi}{2}, \frac{\pi}{2}, 0)$ and $(0, \frac{\pi}{2}, 0)$, respectively. Then we have $B_1 = R'_1 A = \mathbf{x}$, $B_2 = R'_2 A = \mathbf{y}$, $B_3 = R'_3 A = \mathbf{z}$ and $B_1 B_2^T = B_1 B_3^T = B_2 B_3^T = 0$.

Normalizing B_1 , B_2 , and B_3 , orthogonal bases are obtained as $\hat{B}_1 = B_1/|B_1|$, $\hat{B}_2 = B_2/|B_2|$, and $\hat{B}_3 = B_3/|B_3|$ in $\mathcal{L}_3^1(A)$. $\forall (\phi, \theta, \psi) \in \mathbb{R}^3$

$$\begin{aligned} B = & \hat{B}_1 |B_1| \cos \psi \cos \theta + \hat{B}_2 |B_2| (-\sin \psi \cos \phi + \cos \psi \sin \theta \sin \phi) \\ & + \hat{B}_3 |B_3| (\sin \psi \sin \phi + \cos \psi \sin \theta \cos \phi) \end{aligned} \quad (5)$$

Projecting B onto $\mathcal{L}_3^1(A)$, we get three coordinates as follows.

$$\begin{cases} v_1 = |B_1| \cos \psi \cos \theta \\ v_2 = |B_2|(-\sin \psi \cos \phi + \cos \psi \sin \theta \sin \phi) \\ v_3 = |B_3|(\sin \psi \sin \phi + \cos \psi \sin \theta \cos \phi) \end{cases} \quad (6)$$

Let $a = |B_1|$, $b = |B_2|$, and $c = |B_3|$, so that

$$\frac{v_1^2}{a^2} + \frac{v_2^2}{b^2} + \frac{v_3^2}{c^2} = 1. \quad \square \quad (7)$$

Fig. 1 shows an ellipsoid pose manifold of a 3D sphere under all 3D rotations including rolls, yaws, and tilts. We randomly sampled the points on the sphere, which makes $\mathbf{x}\mathbf{x}^T \neq \mathbf{y}\mathbf{y}^T \neq \mathbf{z}\mathbf{z}^T$. According to Lemma 2, the three semi-axes have different lengths, i.e. $a \neq b \neq c$. Therefore, this manifold is not spherical.

Now we extend this knowledge to 3D rotation and 2D projection. Because each image is represented by a $2n$ -dimensional vector, a set of bases can be chosen as

$$\hat{B}_1 = \frac{(x,0)}{|x|}, \hat{B}_2 = \frac{(y,0)}{|y|}, \hat{B}_3 = \frac{(z,0)}{|z|}, \hat{B}_4 = \frac{(0,x)}{|x|}, \hat{B}_5 = \frac{(0,y)}{|y|}, \hat{B}_6 = \frac{(0,z)}{|z|}. \quad (8)$$

Apparently, $\hat{B}_i \perp \hat{B}_j, \forall i \neq j$ and they are bases of $\mathcal{L}_3^2(A)$. Projecting $B(\phi, \theta, \psi)$ onto each basis respectively, we get six coordinates in the pose linear subspace. Similar to the 3D-1D case, we can get two ellipsoid expressions as

$$\begin{cases} \frac{v_1^2}{a^2} + \frac{v_2^2}{b^2} + \frac{v_3^2}{c^2} = 1 \\ \frac{v_4^2}{a^2} + \frac{v_5^2}{b^2} + \frac{v_6^2}{c^2} = 1 \end{cases}. \quad (9)$$

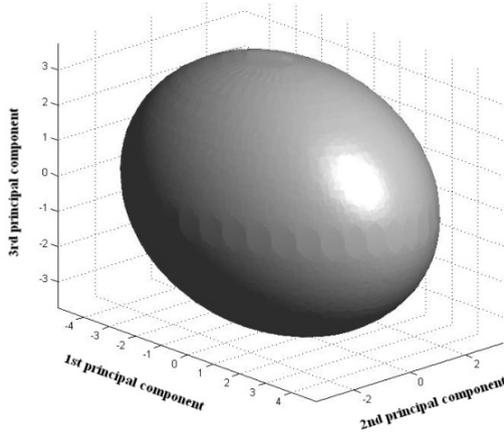


Fig. 1. 3D-1D pose manifold of a sphere with all rotations $R \in SO(3)$ and a projection $P = [1,0,0]$. The manifold is a 3D ellipsoid (not a sphere), because the sampling method on the sphere shape is not even.

Theorem 2. *If the rows of A are orthogonal to each other, all images of a 3D object in $2n$ -dimensional pose space under all rotations form a pose manifold in a $6D$ linear pose subspace described by Eq. 9. \square*

If the point set doesn't have the orthogonal property, we use a strategy similar to the proof of Theorem 2 and rotate the point set a specific angle so that it satisfies the orthogonal requirement. It's not difficult to obtain the rotation matrix $R_0 = R_z(\psi_0)R_y(\theta_0)R_x(\phi_0)$, where

$$\phi_0 = \frac{1}{2} \operatorname{atan} \frac{2yz^T}{yy^T - zz^T}, \theta_0 = \frac{1}{2} \operatorname{atan} \frac{2xz^T}{xx^T - zz^T}, \psi_0 = \frac{1}{2} \operatorname{atan} \frac{2xy^T}{xx^T - yy^T}. \quad (10)$$

Denote the rotated point set as $\{x', y', z'\}^T$, such that $\{x', y', z'\}^T = R_0\{x, y, z\}^T$. According to Theorem 2, all images of rotated $\{x', y', z'\}^T$ form the pose manifold described by Eq. 9. The original point set $\{x, y, z\}^T$ can be seen as one image of rotated $\{x', y', z'\}^T$. Therefore, the general case of non-orthogonal point set is proven and can be stated as follows.

Proposition 4. *All images of a 3D object in 2n-dimensional pose space under all rotations form a parametric pose manifold described by Eq. 9.* □

Because this manifold is a geometric structure existed in a 6D pose linear subspace, it is difficult to visualize it in 3D space. It is also difficult to use 3D senses to imagine this 6D geometry. To help readers to imagine this pose manifold, we show in Fig. 2 a few 3D projections of it from different projecting directions. Also, it has a similar but higher dimensional geometry of the following structure. Consider two identical cylinders along x - and y -axes, respectively, which are expressed as $\frac{y^2}{a^2} + \frac{z^2}{b^2} = 1$ and $\frac{x^2}{a^2} + \frac{z^2}{b^2} = 1$. The intersection of them is two ellipses in 3D space, which has a similar but low-dimensional structure to the pose manifold described by Eq. 9.

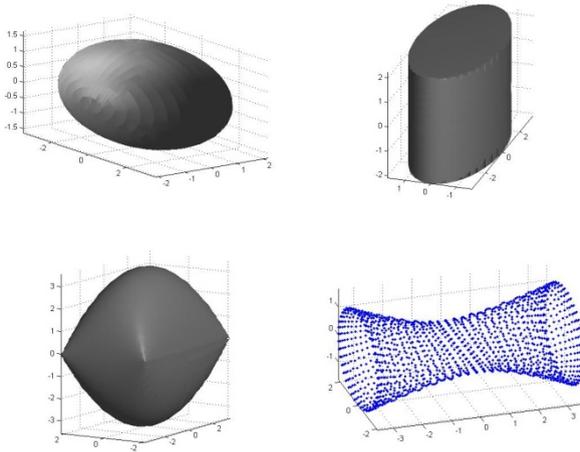


Fig. 2. 4 different 3D projections of the 6D pose manifold described by Eq. 9. Similar to a 3D object (a cylinder) forming different 2D projections (circles, rectangles, etc.) along different projecting directions, it forms different 3D projections in different projecting directions.

3 Experiments

We demonstrate the proposed pose manifold concepts using experiments on both synthetic and real images. In the first experiment, a 3D face scan from the USF Human ID database was used to synthesize the two input point sets with texture information tagged on them. The 3D rotation angle along y -axis is 10° between the two inputs. Fig. 3 shows these two rendered projections with textures for learning the pose manifold. Note that they are not 3D rendering, but 2D textured point clouds. From now on, no 3D face models are involved. We denote the two textured point clouds as

$$\mathbf{p}_1 = \begin{Bmatrix} x_{11}, \dots, x_{1n} \\ y_{11}, \dots, y_{1n} \end{Bmatrix}, \quad \mathbf{t}_1 = \begin{Bmatrix} r_{11}, \dots, r_{1n} \\ b_{11}, \dots, b_{1n} \\ g_{11}, \dots, g_{1n} \end{Bmatrix}, \quad \text{and} \quad \mathbf{p}_2 = \begin{Bmatrix} x_{21}, \dots, x_{2n} \\ y_{21}, \dots, y_{2n} \end{Bmatrix}, \quad \mathbf{t}_2 = \mathbf{t}_1, \quad \text{where } \mathbf{p}_1$$

and \mathbf{p}_2 are 2D point locations and \mathbf{t}_1 and \mathbf{t}_2 are RGB texture information.

An orthogonal bases are established as $\mathbf{s}_1 = \frac{\mathbf{x}_1}{|\mathbf{x}_1|}$ and $\mathbf{s}_2 = \frac{\mathbf{x}_2 - \mathbf{x}_2^T \mathbf{s}_1}{|\mathbf{x}_2 - \mathbf{x}_2^T \mathbf{s}_1|}$ which spans \mathcal{L}_3^2 . Projecting \mathbf{x}_1 and \mathbf{x}_2 onto \mathbf{s}_1 and \mathbf{s}_2 , we get $\mathbf{x}_1 = u_1 \mathbf{s}_1$ and $\mathbf{x}_2 = u_2 \mathbf{s}_1 + v_2 \mathbf{s}_2$. Due to Theorem 2, $\exists \alpha \in [0, \frac{\pi}{2}]$, $a > 0, b > 0$, such that

$$\begin{cases} u_1 \cos \alpha = a \cos \theta \\ u_1 \sin \alpha = b \sin \theta \\ u_2 \cos \alpha - v_2 \sin \alpha = a \cos(\theta + \Delta\theta) \\ u_1 \sin \alpha + v_2 \cos \alpha = b \sin(\theta + \Delta\theta) \end{cases}, \quad (11)$$

where $\Delta\theta$ is the rotation angle between the two images ($\Delta\theta = \frac{\pi}{18}$). Solving these equations, we get $\alpha = \text{atan}\left(\frac{u_1 \cos \theta}{v_2 \cos(\theta + \Delta\theta)} - \frac{u_2}{v_2}\right)$.

The textured point sets are shown in Fig. 3(a) and the two basis images ($a\mathbf{s}_1$ and $b\mathbf{s}_2$) are shown in Fig. 3(b). The semi-axes are multiplied to make the aspect ratios of basis images ‘‘correct’’. Then the manifold on 2D plane \mathcal{L}_3^2 is plotted in Fig. 4, with the two input point sets marked ‘‘■’’. Then any point on \mathcal{M}_3^2 forms a rotated image and a few examples are synthesized in Fig. 5.



Fig. 3. Pose manifold concepts on synthetic face data. (a) 2D projections of 3D rotated texture points along y -axis as inputs. (b) The two basis images ($\mathbf{s}_1, \mathbf{s}_2$) obtained from the data of (a) which spans \mathcal{L}_3^2 .

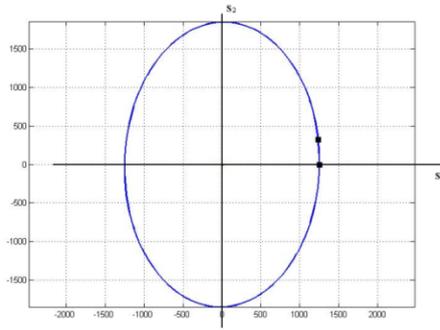


Fig. 4. The elliptic pose manifold \mathcal{M}_3^2 on \mathcal{L}_3^2 spanned by s_1 and s_2 as shown in Fig. 3(b). The two inputs are marked as ■.

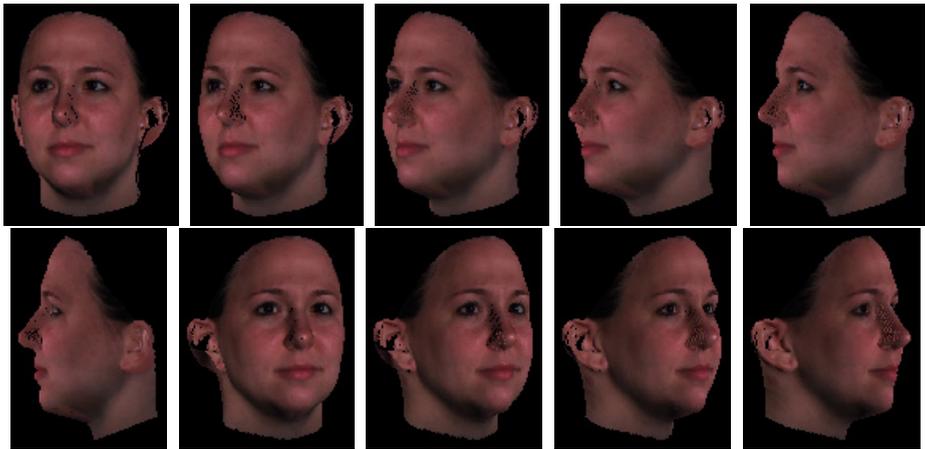


Fig. 5. A few synthesized example images from the pose manifold shown in Fig. 4

A 3D-2D pose manifold was built in Experiment 2 from three views of a fish model as shown in Fig. 6. Fig. 6(a) is an exact side view along z -axis, while Fig. 6(b) and 6(c) are the side views slightly rotated along y -axis and x -axis, respectively. From Proposition 4, two ellipsoids are estimated using the first 3 principal components and the 4-6 principal components, respectively. They are shown in Fig. 7. The input images and the synthesized images are marked as red and blue spheres, respectively.

Because the 3D-2D pose manifold is a 3D geometry in a 6D linear subspace, neither of the two projections (Fig. 7a and 7b) uniquely represents all possible rotations (or viewing directions). Some rotations might merge into one point on one of the projected ellipsoids. For instance, input images shown in Fig. 6(a) and 6(c) merge into the lower blue sphere in Fig. 7(a), and input images in Fig. 6(a) and 6(b) merge into the upper blue sphere in Fig. 7(b).

Taking a point on the manifold (red sphere on Fig. 7), a virtual rotated image of the fish model can be synthesized. A few examples are shown in Fig. 8. Note that the synthesized views are all from unseen angles and no actual 3D model is involved.

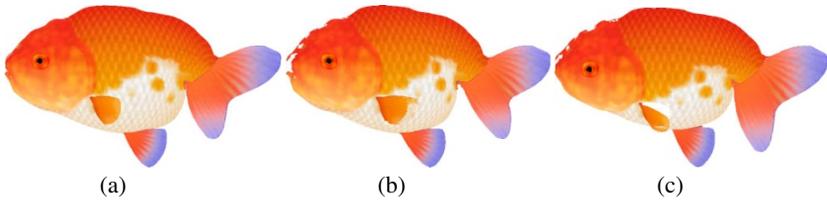


Fig. 6. Original fish images for estimating the pose manifold. (a) Exact side view, (b) side view rotated along y -axis by 20° , and (c) side view rotated along x -axis by 20° .

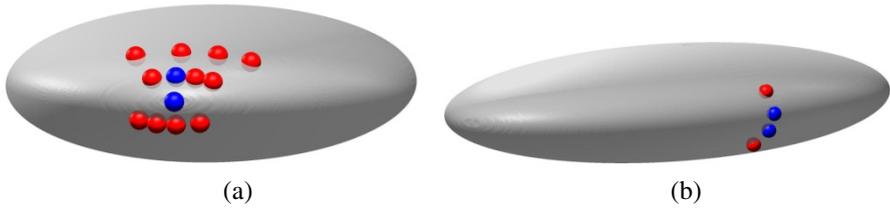


Fig. 7. Pose manifold estimated from the 3 images (●) shown in Fig. 6. The pose manifold is a 3D surface in a 6D linear subspace. (a) is its projection in the first 3 dimensions, and (b) is its projection in the 4th-6th dimensions. Combining (a) and (b) forms the entire parametric manifold. The rendered images shown in Fig. 8 are marked as ●.



Fig. 8. Synthesized images with simulated 3D rotations from the pose manifold shown in Fig. 7. The rotations are 20° apart in tilt and yaw.

In Experiment 3, we estimated the pose manifold using three real images of a human face under near-frontal rotations as shown in Fig. 9(a). The images were from CMU-PIE face database [14] and were resized/cropped to 80×80 pixels. The rotation angles of the three images along y -axis are $\theta_1 = -17^\circ$, $\theta_2 = 0^\circ$, and $\theta_3 = 16^\circ$. 67 corresponding points were manually labeled on each of these three images as shown in Fig. 9(b).

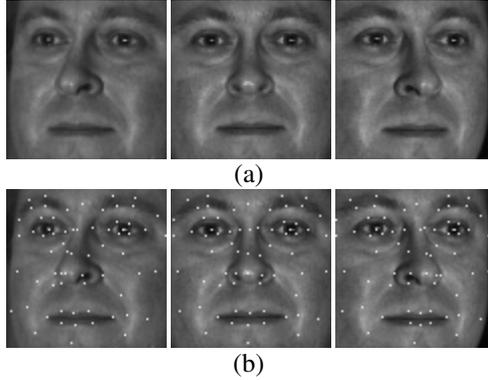


Fig. 9. Three real face images under rotations for constructing the pose manifold. (a) Original input images. (b) 67 labels on each image.

The following steps were performed to estimate the pose manifold and to use the pose manifold to predict images under different rotations.

1. The human face was sampled based on all pixels in image 2 and denoted the coordinates as $\mathbf{p}_2 = \{x_{21}, \dots, x_{2n}, y_{21}, \dots, y_{2n}\}$ and the textures as $\mathbf{t} = \{I_1, \dots, I_n\}$, where I_i is image intensity of pixel i .
2. Thin-plate spline (TPS) warping [5] was performed to establish a point-wise correspondence. The corresponding pixel locations in image 1 and image 3 are denoted as $\mathbf{p}_1 = \{x_{11}, \dots, x_{1n}, y_{11}, \dots, y_{1n}\}$ and $\mathbf{p}_3 = \{x_{31}, \dots, x_{3n}, y_{31}, \dots, y_{3n}\}$.
3. Principal component analysis was performed on C^T , where $C = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}^T$ and the first two principal components $\{\mathbf{s}_1, \mathbf{s}_2\}$ were used as bases for the pose linear subspace \mathcal{L}_3^2 .
4. The 3 input images were projected onto \mathcal{L}_3^2 and 3 2D vectors $(\{u_1, v_1\}, \{u_2, v_2\}, \{u_3, v_3\})$ were obtained. Fig. 10 shows the two basis images which represent two orthogonal in-depth rotations.
5. With known $\{\theta_1, \theta_2, \theta_3\}$, an ellipse $(\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1)$ was fitted as $a = \frac{1}{3} \sum \frac{u_i}{\cos \theta_i}$ and $b = \frac{1}{2} (\frac{v_1}{\sin \theta_1} + \frac{v_3}{\sin \theta_3})$.
6. Given an arbitrary rotation angle θ' , the image can be estimated as $\mathbf{p}' = a \cos \theta' \mathbf{s}_1 + b \sin \theta' \mathbf{s}_2$. The image can be rendered with textures \mathbf{t} .

Fig. 11 shows the estimated elliptic pose manifold (Step 5), 3 input images projected onto the pose linear subspace, the images synthesized with rotation angles $\Theta = \{\frac{k\pi}{12} | k = -4, -3, \dots, 3, 4\}$. These two experiments demonstrated that these parametric

pose manifold representations are valid in both view analysis and novel view synthesis. With known rotation angles for input images, minimal 2 images are required to estimate the elliptic pose manifold. In reality, more than 2 images (3 in the second experiment) will help to make the estimation more stable.

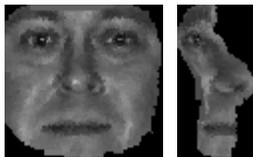


Fig. 10. The two basis images generated in Step 4 from Fig.8(a)

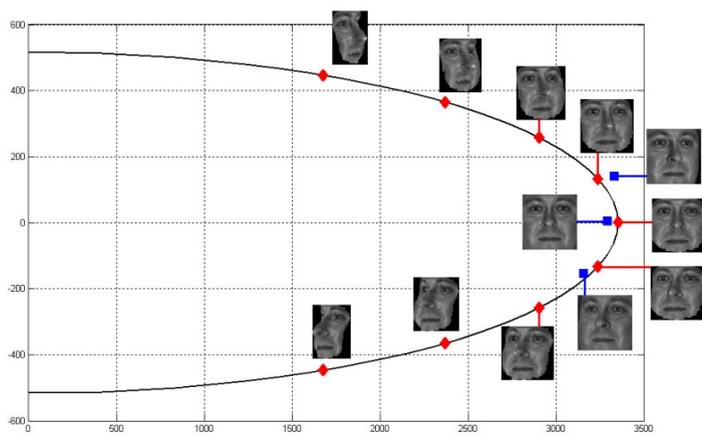


Fig. 11. 3 input images (■) and synthesized images (◆) marked on the estimated elliptic pose manifold in the 2D pose linear subspace spanned by $(\mathbf{s}_1, \mathbf{s}_2)$. This complete and continuous representation can extrapolate appearances of a face under unseen rotations.

4 Conclusions and Discussions

In this paper, we have demonstrated that 1) the set of images of an object under in-depth rotations along a single axis is an elliptic manifold in 2D pose linear subspace, and 2) the set of images of an object under all rotations is a parametric manifold described by Eq. 9 in a 6D pose linear subspace. Furthermore, the validity of the proposed pose manifold was illustrated with experimental investigations on both synthetic and real image data. These findings provide bases for computer vision systems which deal with pose/viewpoint variations.

There are a number of possible extensions and open issues that we would discuss as follows.

4.1 Translation and Scaling

This paper deals primarily with rotations, while translation and scaling also constitute viewpoint changes (e.g., panning and zooming). Similar to rotation, a small number of parameters are required to describe translation (3 parameters) and scaling (1 parameter) in 3D space. Therefore, images under such transformations also form lower dimensional parametric manifold. It is one of our future works in developing a rotation, translation and scaling manifold theory. An alternative is to normalize the images under translation and scaling and they are easier to “correct” in 2D image plane than rotation.

4.2 Perspective Projection

Orthographical projection is assumed throughout this paper. For objects located far from cameras, it’s a good approximation. Perspective or other projections can be handled in a similar way. It has a different form of projection matrix. In the future, we plan to extend this pose manifold concept to include perspective projection, which is good at representing near-camera objects.

4.3 Occlusion

In-depth rotations of a 3D object inevitably involve occlusions. Occluded pixels may cause the proposed pose manifold representation inconsistent dimensions of pose spaces of different images. Visible parts of the object in all input images can be chosen and the findings in this paper hold for any subset of the object. Restricting the 3D objects to be convex objects can alleviate the occlusion problem. In the future, we may group the input images into adjacent sets and estimate a “partial” pose manifold from each set. Then a “composite” pose manifold may be combined.

4.4 Correspondence

A dense correspondence is needed to estimate the pose manifold. In this paper, a sparse set of correspondence was manually marked and the dense correspondence was established by a TPS warping. Existing stereo matching algorithms can be a good option for such tasks. In turn, the findings of this paper may serve as a regularization term for stereo algorithms.

4.5 Relation to 3D modeling

All the operations reported in this paper are confined in 2D image planes, except synthesizing input data which involved 3D rotations. It is interesting to investigate its relationships to 3D modeling using stereo algorithms. The findings of this paper can be seen as a 2D representation of 3D rotations without explicit 3D models. The manifold constraints can be used to prune the corresponding points established using stereo matching techniques. Avoiding explicit 3D operations may also reduce computational complexity of the tasks.

4.6 Illumination

Beside viewpoint variations, illumination variations also affect the appearances of an object greatly. The illumination linear subspaces and manifolds were better established than pose (refer to [1, 2, 10]). In the future, we plan to combine these two subspace theories to include both illumination and pose variations in the proposed representations.

References

- [1] Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(2), 218–233 (2003)
- [2] Belhumeur, P.N., Kriegman, D.J.: What is the set of images of an object under all possible illumination conditions. *Int. J. Comput. Vis.* 28(3), 245–260 (1998)
- [3] Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(9), 1063–1074 (2003)
- [4] Blanz, V., Grother, P., Phillips, P.J., Vetter, T.: Face recognition based on frontal views generated from non-frontal images. In: *Proc. IEEE Conf. CVPR* (2005)
- [5] Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* 11(6), 567–585 (1989)
- [6] Caglioti, V.: On the space requirements of indexing 3D models from 2D perspective images. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 718–723 (2000)
- [7] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(6), 681–685 (2001)
- [8] Fermuller, C., Aloimonos, Y.: On the Geometry of Visual Correspondence. *Int. J. Comput. Vis.* 21(3), 223–247 (1997)
- [9] Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and light-fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(4), 449–465 (2004)
- [10] Jacobs, D.: The space requirements of indexing under perspective projections. *IEEE Trans. Pattern Anal. Mach. Intell.* 18(3) (1996)
- [11] Prince, S.J.D., Warrell, J., Elder, J.H., Felisberti, F.M.: Tied Factor Analysis for Face Recognition across Large Pose Differences. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(6), 970–984 (2008)
- [12] Sashua, A.: *Geometry and Photometry in 3D Visual Recognition*. Ph.D. MIT (1992)
- [13] Sashua, A.: On photometric issues in 3D visual recognition from a single 2D image. *Int. J. Comput. Vis.* 21(1-2), 99–122 (1997)
- [14] Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(12), 1615–1618 (2003)
- [15] Ullman, S., Basri, R.: Recognition by Linear Combinations of Models. *IEEE Trans. Pattern Anal. Mach. Intell.* 13(10), 992–1006 (1991)