RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

Rotation update on manifold in probabilistic NRSFM for robust 3D face modeling

Chengchao Qu[1,2]*, Hua Gao[3] and Hazım Kemal Ekenel[4]

Abstract

This paper focuses on recovering the 3D structure and motion of human faces from a sequence of 2D images. Based on a probabilistic model, we extensively studied the rotation constraints of the problem. Instead of imposing numerical optimizations, the inherent geometric properties of the rotation matrices are taken into account. The conventional Newton's method for optimization problems was generalized on the rotation manifold, which ultimately resolves the constraints into unconstrained optimization on the manifold. Furthermore, we also extended the algorithm to model within-individual and between-individual shape variances separately. Evaluation results give evidence to the improvement over the state-of-the-art algorithms on the Mocap-Face dataset with additive noise, as well as on the Binghamton University A 3D Facial Expression (BU-3DFE) dataset. Robustness in handling noisy data and modeling multiple subjects shows the capability of our system to deal with real-world image tracks.

Keywords: Non-rigid structure from motion, Manifold optimization, Newton's method, PLDA, Face model

1　Introduction

Recovering scene geometry and camera motion from sequences of 2D monocular images has seen significant success for the 3D geometry of static objects. The widely used rigid factorization method was first introduced by Tomasi and Kanade [1]. Orthonormality constraints are adopted on the rotation matrices in order to recover structure and motion in a single step. Unfortunately, most biological objects and natural scenes are deformable. 3D rigid motions, i.e., camera rotation and translation, along with non-rigid deformations, e.g., stretching and bending, are mixed altogether in their image measurements. Hence, extending the existing rigid algorithms to the non-rigid scenario turns out to be a far more challenging task than it appears to be.

It is known that the problem of non-rigid structure from motion (NRSFM) is generally underconstrained and thus intractable, if each point of the object moves arbitrarily. In practice, however, many objects, e.g., faces, deform under certain rules. A possible approach is to learn an application-specific 3D model of non-rigid structure from the training data to constrain deformation [2]. Another possibility is to hard-code and learn a model incrementally [3]. Some approaches [4–7] were proposed from another perspective to remove the need of such a prior model, which is not available in most real-world situations. The shape model, i.e., shape bases, is treated as unknowns to be solved, with only the orthonormality constraints on camera rotations being utilized. Xiao et al. [8] proved that only enforcing the orthonormality constraints is not enough for the factorization-based method; therefore, they introduced the basis constraints to reduce ambiguity.

In this work, two major contributions have been made. We first investigated the geometric properties of the orthonormality constraints and generalized the Newton's optimization method to the underlying manifold of the camera rotation matrices. That means, non-linear optimization can be carried out on the manifold without any imprecise approximations. We used a probabilistic principal component analysis (PPCA)-based framework [9] to model NRSFM as it is more robust to noise than the closed-form factorization techniques. Our second contribution is about dealing with multiple subjects. The current NRSFM algorithms mostly focus on the reconstruction of a single subject. While dealing with

*Correspondence: qu@kit.edu
[1]Vision and Fusion Laboratory (IES), Karlsruhe Institute of Technology (KIT), Adenauerring 4, 76131 Karlsruhe, Germany
[2]Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (Fraunhofer IOSB), Fraunhoferstr. 1, 76131 Karlsruhe, Germany
Full list of author information is available at the end of the article

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 2 of 12

data containing multiple subjects, no difference is taken into account, when modeling between-individual variation (e.g., face model of different identities) and within-individual variation (e.g., facial expression of the same identity). For that reason, we extended the PPCA-based framework to the probabilistic linear discriminant analysis (PLDA) [10] model to improve reconstruction performance on data with multiple subjects.

The remainder of this paper is organized as follows. Previous research on NRSFM is reviewed in Section 2. Section 3 presents the probabilistic NRSFM model [9] and our novel manifold optimization technique on the orthonormality constraints. Section 4 discusses the experimental results of our algorithm. Finally, we conclude our work in Section 5.

2 Related work

Modern structure from motion (SFM) algorithms employ the factorization method for orthographic camera projection proposed by Tomasi and Kanade [1]. The rank theorem ensures that the input matrix can be factorized into two matrices, one corresponds to the camera motion, and the other represents the shape. Although the resulting matrices from singular value decomposition (SVD) are not unique, they only differ by a linear transformation. By imposing metric constraints, a decent solution of the SFM problem for rigid objects can be achieved.

In the seminal work of Bregler et al. [11] and Torresani et al. [6] for solving NRSFM, they assumed that the 3D shape of an object can be explained as a linear combination of deformation shapes applied to a dominant rigid component. In this way, the non-rigid motion recovery is formulated as a factorization problem and the low rank of the image measurements is analyzed. In general, this model assumes that the number of basis shapes should be known, an inaccurate choice that can lead to performance drop. Theoretically, if the number is underestimated, it is not sufficient to represent all variations of the object; otherwise, the extra degree of freedom is unconstrained and is unlikely to generalize well, which starts fitting to noise.

Using the linear representation, Xiao et al. [8] proposed a closed-form scheme for solving the NRSFM problem. They proved in the previous work that by imposing orthonormality constraints alone on camera rotations, the increased degree of freedom will cause ambiguity. The additional basis constraints will determine the shape bases uniquely. In [12], Xiao and Kanade pointed out that even enforcing both sets of linear metric constraints above could still lead to ambiguity, if there exist degenerate bases, which are not of full rank three. However, by exploiting the rank three constraints inherently, Akhter et al. [13] analytically proved that orthonormality constraints alone are sufficient to recover the exact structure. Ambiguity solely lies in the transformation of linear basis

vectors, which does not affect the 3D structure reconstruction. Dai et al. [14] proved this claim by solving the NRSFM problem without any prior using matrix trace norm minimization.

Torresani et al. [9] proposed a probabilistic deformation model based on PPCA and suggested that it reveals better reconstruction result than the conventional linear model. In their work, 3D shapes are drawn from non-uniform probability distribution functions (PDFs) with a Gaussian prior on each shape in the subspace instead of the common linear subspace model, which is a specific usage of PPCA. The parameters of the PDF are unknown in advance, which will be optimized using the expectation-maximization (EM) algorithm together with the 3D shapes and rigid motions. An advantage of PPCA over the simple deterministic subspace model is that degeneracy of closed-form solutions does not occur so that the ambiguity problem figured out by Xiao et al. in [12] does not happen here. However, the rotation matrices are approximated by using a single Gauss–Newton step with a fixed updating step length, which can lead to a considerable performance drop in the rotation reconstruction if no proper metric on the manifold is defined.

Over the last years, more research on NRSFM has also been done using various forms of non-linear optimization techniques to minimize the 3D reprojection error. In order to overcome the degeneracy problem, some additional heuristic constraints were introduced. Shaji and Chandran [15] introduced a canonical Riemannian metric on the product span subspace of the rotation matrices and articulated shape weights. The Newton's algorithm is generalized to the product manifold to recover those parameters, while the Wiberg algorithm is employed to solve the shape update. It differs from our approach in that our framework uses a probabilistic model with a posterior objective function over the latent variables, which is more robust to noise introduced by tracking error or manual labeling. Section 4 shows the robustness of our model with extreme conditions of noise.

Other than recovering the whole 3D shapes and motion parameters like in almost all the existing applications, Rabaud and Belongie [16] presented a manifold learning approach that only focuses on an embedding of frames within the input image sequence. The intuition is as follows: given enough image frames, a non-rigid deformed 3D shape can be observed several times in different view angles. If some of the frames share a low 3D reconstruction error, they are highly likely to represent a similar 3D shape, otherwise it means a poorly matched set of frames. Following this principle, triplets of frames are compared to exploit all repetitions in possible shape deformations. Then the generalized non-metric multi-dimensional scaling framework is used to estimate the weight of each deformation shape. Bundle adjustment is employed as a

further optimization step, which minimizes the repro-jection error. This closed-form approach can reconstruct accurate 3D shape on a clean synthetic dataset; how-ever, with the amount of noise added, their performance drops very fast and approaches that of PPCA. Tao and Matuszewski [17] also employed manifold learning-based diffusion maps to handle highly deformable objects.

By exploiting the temporal smoothness of the shape tra-jectories across the images, Akhter et al. [18] addressed the NRSFM problem in trajectory space, which is the dual problem to the conventional spatial shape bases. By describing the 3D point trajectory linearly using object independent discrete cosine transform (DCT) vectors, unknowns in estimation are reduced, and stable recon-struction is achieved as a result. Gotardo and Martinez extended the temporal dependence to iteratively obtain higher-frequency DCT in [19] and explicitly modeled the complementary spaces of rank three in [20]. Valmadre and Lucey [21] formulated the regularization of the trajec-tory basis with a temporal filter. Recently, Park et al. [22] simplified the global motion estimation of the trajectory basis with the aid of a few stationary points in the scene. Despite the robust performance on various motion cap-ture datasets, limitation of its application is also obvious, i.e., the object deformation should be temporally con-tinuous and smooth. Otherwise, higher-frequency DCT vectors are needed, which significantly increases the rank of the trajectory matrix factorization and will eventually lead to degeneration and unstable performance. In con-trast, the primary problem in shape space does not suffer from this.

In this paper, we demonstrate a probabilistic, iterative alternating approach to solve the NRSFM problem. In contrast to Torresani et al. [9], the conventional New-ton's method is generalized on the rotation manifold to solve the optimal rotation matrix for each optimization iteration. The orthonormality constraints are naturally guaranteed by the metric update step without the need of being projected back after constrained optimizations on the Euclidean space. Additionally, a generic PLDA model that takes into account the commonness across all sub-jects, as well as the specific characteristics between the subjects, can be learned. On datasets with more than one subject, better individual reconstruction is achieved even if insufficient number of frames are available for each subject.

3 NRSFM model

Most of the state-of-the-art NRSFM algorithms make use of a linear subspace model to represent the shape model. A linear combination of deformation shapes is thereby applied to a dominant rigid component. Let the $3P \times 1$ matrix $\bar{\mathbf{s}}$ be the mean shape and the $3P \times K$ matrix $\mathbf{V}$ and the K-dimensional vector $\mathbf{z}_t$ be the remaining basis shapes

and their weights, respectively, where P is the number of landmarks in each image frame and K the number of artic-ulation shapes apart from the mean shape. The 3D shape of the tth frame is represented as

$$\mathbf{s}_t = \bar{\mathbf{s}} + \mathbf{V}\mathbf{z}_t. \tag{1}$$

Note that shapes are stacked in matrix $\mathbf{V}$ so that each column represents a basis shape. Camera rotation in frame t is denoted by the 2×3 matrix $\mathbf{R}_t$. Due to the inevitable presence of internal and external noise in image tracks or labeling, a zero-mean Gaussian noise $\mathbf{n}_t$ with variance σ^2 is also added. If we align the images to the cen-ter and drop the translations, the 2D observation matrix under the orthographic camera model can be factorized into

$$\mathbf{p}_t = \mathbf{R}_t(\bar{\mathbf{s}} + \mathbf{V}\mathbf{z}_t) + \mathbf{n}_t. \tag{2}$$

This probabilistic formulation of the conventional prin-cipal component analysis (PCA) was addressed by Tipping and Bishop in [23]. It has a simple linear probabilistic assumption that all marginal and conditional distributions are Gaussian. PPCA is closely related to factor analysis [24], in which a statistical model is used to describe the relation between the observed vector $\mathbf{p}_t$ and the corre-sponding latent variables $\mathbf{z}_t$.

In Eq. (2), the weight coefficients $\mathbf{z}_t$ are formulated as an independent and identically distributed (i.i.d.) Gaussian prior

$$\mathbf{z}_t \sim \mathcal{N}(0; \mathbf{I}). \tag{3}$$

These unobserved or latent variables are marginalized out instead of being explicitly calculated. Since there only exists linear transformations in Eq. (2), the measurement matrix $\mathbf{p}_i$ is also Gaussian distributed [9] with the form

$$\mathbf{p}_t \sim \mathcal{N}(\mathbf{R}_t\bar{\mathbf{s}}; \mathbf{R}_t\mathbf{V}\mathbf{V}^\top\mathbf{R}_t^\top + \sigma^2\mathbf{I}). \tag{4}$$

3.1 Shape update

The PPCA model can be estimated iteratively by the EM algorithm [9]. In the expectation step (E-step), the posterior distribution over $\mathbf{z}_t$ is defined as

$$\begin{aligned} q(\mathbf{z}_t) &= p(\mathbf{z}_t|\mathbf{p}_t, \mathbf{R}_t, \bar{\mathbf{s}}, \mathbf{V}, \sigma^2) \\ &= \mathcal{N}(\mathbf{z}_t|\boldsymbol{\mu}_t; \boldsymbol{\Sigma}_t). \end{aligned} \tag{5}$$

Over this distribution, the first two moments of $\mathbf{z}_t$ are given

$$\boldsymbol{\mu}_t = \mathbb{E}[\mathbf{z}_t], \tag{6}$$

$$\boldsymbol{\phi}_t = \mathbb{E}[\mathbf{z}_t\mathbf{z}_t^\top] = \boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^\top. \tag{7}$$

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 4 of 12

In the following maximization step (M-step), the expected negative log-likelihood function

$$
\begin{aligned}
\mathcal{L} &= -\mathbb{E}\left[\sum_t \log p(\mathbf{p}_t|\mathbf{R}_t, \bar{\mathbf{s}}, \mathbf{V}, \sigma^2)\right] \\
&= \frac{1}{2\sigma^2}\sum_t \mathbb{E}\left[||\mathbf{p}_t - \mathbf{R}_t(\bar{\mathbf{s}} + \mathbf{V}\mathbf{z}_t)||^2\right] \\
&\quad + JT\log(2\pi\sigma^2)
\end{aligned}
\tag{8}
$$

is minimized. The shape bases $\{\bar{\mathbf{s}}, \mathbf{V}\}$ and the noise parameter σ^2 can be updated individually in closed form by setting their partial derivative to zero [9] with the help of the expectations in Eqs. (6) and (7).

However, the camera rotation parameter $\mathbf{R}_t$ is subject to orthonormality constraints; hence, closed-form update like the other parameters is not possible. Torresani et al. [9] approximated the solution with a single Gauss–Newton step on the Euclidean space, which is inaccurate and has a theoretically low convergence rate. In the upcoming section, we propose our optimization technique on the manifold.

3.2 Motion update on manifold

In [9], a twist vector $\boldsymbol{\xi}$ is employed to hold the result of the single Gauss–Newton step. The exponential map of the skew-symmetric matrix $\hat{\boldsymbol{\xi}}$ is then set as the updating vector $\boldsymbol{\Delta}$. Note that without defining an appropriate metric on the manifold, a manually selected and fixed updating step length is implemented, which declines the performance obviously, when faced complex setups.

3.2.1 Newton's method on SO(3)

As we consider the orthographic camera model, the camera motion matrix $\mathbf{R}_t$ in Eq. (2) is obtained by projecting a 3D rotation matrix to 2D with an orthographic projection matrix

$$
\boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},
\tag{9}
$$

so that the mapping

$$
\mathbf{R} = \boldsymbol{\Pi}\mathbf{Q}
\tag{10}
$$

from 3D to 2D is satisfied. The rotation matrix $\mathbf{Q}$ is an orthogonal matrix with a determinant one, which lies exactly on the manifold of the special orthogonal group

$$
SO(3) = \left\{ \mathbf{Q} \in \mathbb{R}^{3\times3} : \mathbf{Q}^\top\mathbf{Q} = \mathbf{I}, \det(\mathbf{Q}) = 1 \right\}.
\tag{11}
$$

Hence, instead of putting an approximate algebraic or numeric constraint on the Euclidean space $\mathbb{R}^N$ and projecting them back onto the $SO(3)$ manifold, an unconstrained optimization on the manifold is a natural generalization and is expected to perform better.

To start with, we consider to be on the normal Euclidean space. The Newton's method iteratively finds the stationary points of differentiable functions. Provided that $f(x)$ is a twice-differentiable function, the update sequence x_n can be approximated by the Taylor series expansion up to the second order and rewritten as

$$
x_{k+1} = x_k - \left[f''(x_k)\right]^{-1} f'(x_k).
\tag{12}
$$

Given a quadratic function $f(x)$, the optimal point can be found even in a single step. So on the Euclidean space, the first and second order derivatives of the objective function are needed. Edelman et al. [25] proved that for Stiefel manifolds (set of all orthonormal k-frames in $\mathbb{R}^n$, $V_k(\mathbb{R}^n) = \left\{\mathbf{A} \in \mathbb{R}^{n\times k} : \mathbf{A}^\top\mathbf{A} = \mathbf{I}\right\}$), e.g., $SO(3)$, their canonical Riemannian structure makes possible to generalize a Riemannian Newton's method on them. Besides the gradient and Hessian, the definition of the update along the geodesic of the manifold must be known to ensure that the update is valid, because unlike on the Euclidean space, the update path is no longer a straight line but rather a geodesic curve, which stays on the surface of the manifold all the time and defines the shortest path between two points on the surface. The update step is illustrated in Fig. 1.

We define the objective function F with respect to the rotation matrix $\mathbf{Q}$ for the manifold optimization as follows

$$
F(\mathbf{Q}) = \mathbb{E}\left[\|\mathbf{p} - \boldsymbol{\Pi}\mathbf{Q}(\bar{\mathbf{s}} + \mathbf{V}\mathbf{z})\|_F^2\right].
\tag{13}
$$

Since $\mathbf{Q} \in SO(3)$, following [26], its tangent vector $\boldsymbol{\Delta} \in T(SO(3))$ is given by

$$
\boldsymbol{\Delta} = \mathbf{Q}\hat{\mathbf{u}},
\tag{14}
$$
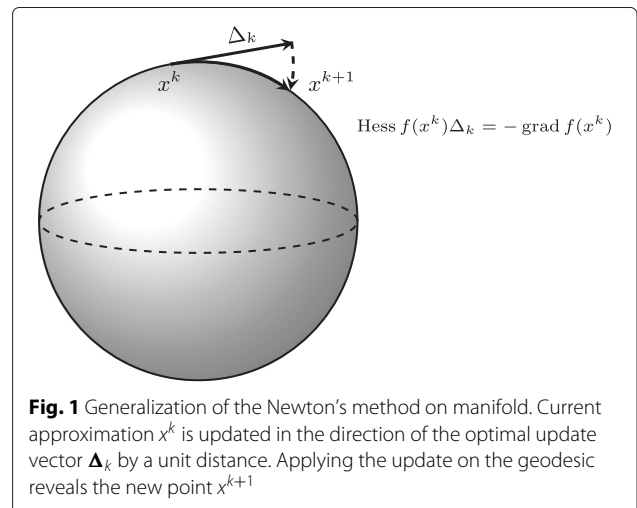
Fig. 1 Generalization of the Newton's method on manifold. Current approximation x^k is updated in the direction of the optimal update vector $\boldsymbol{\Delta}_k$ by a unit distance. Applying the update on the geodesic reveals the new point x^{k+1}

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 5 of 12

where $\hat{\mathbf{u}}$ is the skew-symmetric matrix of vector $\mathbf{u}$ in the form of

$$
\hat{\mathbf{u}} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}. \tag{15}
$$

For the Riemannian manifold, the metric can simply be induced from the Euclidean metric as

$$
g(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = \frac{1}{2} \mathrm{tr}(\boldsymbol{\Delta}_1^\top \boldsymbol{\Delta}_2). \tag{16}
$$

The explicit formula for geodesics on $SO(3)$ at $\mathbf{Q}$ in direction $\boldsymbol{\Delta}$ is then

$$
\begin{aligned}
\mathbf{Q}(t) &= \exp(\mathbf{Q}, \boldsymbol{\Delta} t) = \mathbf{Q}\exp(\hat{\omega}t) \\
&= \mathbf{Q}\left(\mathbf{I} + \hat{\omega}\sin(t) + \hat{\omega}^2(1-\cos(t))\right),
\end{aligned} \tag{17}
$$

where $t \in \mathbb{R}$, $\omega = \mathbf{Q}^\top\boldsymbol{\Delta} \in \mathfrak{so}(3)$ ($\mathfrak{so}(3)$ is the Lie algebra of $SO(3)$). The last equation is called the Rodrigues' rotation formula [27].

3.2.2 Gradient and Hessian

To obtain the gradient and Hessian, we first derive the first and second order derivative for the geodesic $\mathbf{Q}(t)$ with respect to t:

$$
\begin{aligned}
\left.\frac{\mathrm{d}\mathbf{Q}(t)}{\mathrm{d}t}\right|_{t=0} &= \mathbf{Q}\hat{\omega}\cos(t) + \mathbf{Q}\hat{\omega}^2\sin(t)\Big|_{t=0} \\
&= \mathbf{Q}\hat{\omega} \\
&= \mathbf{Q}(\mathbf{Q}^\top\boldsymbol{\Delta}) \\
&= \boldsymbol{\Delta}
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
\left.\frac{\mathrm{d}^2\mathbf{Q}(t)}{\mathrm{d}t^2}\right|_{t=0} &= -\mathbf{Q}\hat{\omega}\sin(t) + \mathbf{Q}\hat{\omega}^2\cos(t)\Big|_{t=0} \\
&= \mathbf{Q}\hat{\omega}^2 \\
&= \mathbf{Q}(\mathbf{Q}^\top\boldsymbol{\Delta})(\mathbf{Q}^\top\boldsymbol{\Delta}) \\
&= \boldsymbol{\Delta}(\mathbf{Q}^\top\boldsymbol{\Delta}) \\
&= -\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\mathbf{Q}
\end{aligned} \tag{19}
$$

Note that the last step in Eq. (19) is derived from the property of tangent space on the Stiefel manifold that $\mathbf{Q}^\top\boldsymbol{\Delta}$ is a skew-symmetric matrix with

$$
\mathbf{Q}^\top\boldsymbol{\Delta} + \boldsymbol{\Delta}^\top\mathbf{Q} = \mathbf{0}. \tag{20}
$$

Given the geodesic definition, we derive the gradient and Hessian in direction $\boldsymbol{\Delta} \in T(SO(3))$:

$$
\begin{aligned}
\mathrm{d}F(\boldsymbol{\Delta}) &= \left.\frac{\mathrm{d}F(\mathbf{Q}(t))}{\mathrm{d}t}\right|_{t=0} \\
&= \boldsymbol{\Pi}\mathbf{Q}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\dot{\mathbf{Q}}^\top\boldsymbol{\Pi}^\top - \mathbf{p}\boldsymbol{\mu}^\top\mathbf{V}^\top\dot{\mathbf{Q}}^\top\boldsymbol{\Pi}^\top|_{t=0} \\
&= \boldsymbol{\Pi}\mathbf{Q}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\boldsymbol{\Delta}^\top\boldsymbol{\Pi}^\top - \mathbf{p}\boldsymbol{\mu}^\top\mathbf{V}^\top\boldsymbol{\Delta}^\top\boldsymbol{\Pi}^\top
\end{aligned} \tag{21}
$$

$$
\begin{aligned}
\mathrm{Hess}\,F(\boldsymbol{\Delta},\boldsymbol{\Delta}) &= \left.\frac{\mathrm{d}^2 F(\mathbf{Q}(t))}{\mathrm{d}t^2}\right|_{t=0} \\
&= \boldsymbol{\Pi}\dot{\mathbf{Q}}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\dot{\mathbf{Q}}^\top\boldsymbol{\Pi}^\top \\
&\quad + \boldsymbol{\Pi}\mathbf{Q}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\ddot{\mathbf{Q}}^\top\boldsymbol{\Pi}^\top \\
&\quad - \mathbf{p}\boldsymbol{\mu}^\top\mathbf{V}^\top\ddot{\mathbf{Q}}^\top\boldsymbol{\Pi}^\top|_{t=0} \\
&= \boldsymbol{\Pi}\boldsymbol{\Delta}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\boldsymbol{\Delta}^\top\boldsymbol{\Pi}^\top \\
&\quad - \boldsymbol{\Pi}\mathbf{Q}\mathbf{V}\boldsymbol{\phi}\mathbf{V}^\top\mathbf{Q}^\top\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{\Pi}^\top \\
&\quad + \mathbf{p}\boldsymbol{\mu}^\top\mathbf{V}^\top\mathbf{Q}^\top\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{\Pi}^\top
\end{aligned} \tag{22}
$$

For any arbitrary pair of vectors $\mathbf{X}, \mathbf{Y} \in T(SO(3))$, polarization [26] helps compute $\mathrm{Hess}\,F(\mathbf{X}, \mathbf{Y})$ with

$$
\begin{aligned}
\mathrm{Hess}\,F(\mathbf{X},\mathbf{Y}) = \frac{1}{4}\,&(\mathrm{Hess}\,F(\mathbf{X}+\mathbf{Y},\mathbf{X}+\mathbf{Y}) \\
&-\mathrm{Hess}\,F(\mathbf{X}-\mathbf{Y},\mathbf{X}-\mathbf{Y})).
\end{aligned} \tag{23}
$$

3.2.3 Algorithm summary

With the requirements for generalizing Newton's method being ready, the optimal updating vector on the manifold can be found by modifying the original Newton Eq. (12) to

$$
\boldsymbol{\Delta} = -\mathrm{Hess}^{-1}\mathbf{G}, \tag{24}
$$

assuming that the Hessian is non-degenerate. It is the same as finding a vector $\boldsymbol{\Delta}$ that satisfies for all vector fields $\mathbf{Y}$

$$
\mathrm{Hess}\,F(\mathbf{Y},\boldsymbol{\Delta}) = g(-\mathbf{G},\mathbf{Y}) = -\mathrm{d}F(\mathbf{Y}), \tag{25}
$$

where $\mathbf{G} = \nabla F$ stands for the gradient. The Hessian can be uniquely determined by using an orthonormal basis $\{\mathbf{E}^k\}, k = 1, 2, 3$ into Eq. (25) as

$$
\mathrm{Hess}\,F(\mathbf{E}^k,\boldsymbol{\Delta}) = -\mathrm{d}F(\mathbf{E}^k). \tag{26}
$$

For simplicity, the standard basis $\mathbf{e}_k$ for $\mathbb{R}^3$ is chosen so that $\mathbf{E}^k = \mathbf{Q}\hat{\mathbf{e}}_k \in T(SO(3))$. Thus, the 3×3 Hessian matrix $\mathbf{H}$ and the three-dimensional gradient vector $\mathbf{g}$ can be obtained:

$$
\mathbf{H}_{kl} = \mathrm{Hess}\,F(\mathbf{E}^k, \mathbf{E}^l), \tag{27}
$$

$$
\mathbf{g}_k = \mathrm{d}F(\mathbf{E}^k),\ k, l = 1, 2, 3 \tag{28}
$$

Then, we solve for the vector $\mathbf{u} = [u_1, u_2, u_3]^\top \in \mathbb{R}^3$ using

$$
\mathbf{u} = -\mathbf{H}^{-1}\mathbf{g}. \tag{29}
$$

Finally, the desired updating vector $\boldsymbol{\Delta} = \mathbf{Q}\hat{\mathbf{u}}$ is obtained. The last step is to update the current rotation along the geodesic in the direction of this vector. The algorithm is summarized in Algorithm 1.

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 6 of 12

Algorithm 1 Minimize $F(\mathbf{Q}) = \mathbb{E}\left[\|\mathbf{p} - \mathbf{Q}(\bar{\mathbf{s}} + \mathbf{V}\mathbf{z})\|_F^2\right]$

---

1. At the point $\mathbf{Q}$, compute the optimal update vector
   $\mathbf{\Delta} = -\text{Hess}^{-1}\mathbf{G}$

   1: Choose basis tangent vectors
      $\mathbf{E}^k = \mathbf{Q}\hat{\mathbf{e}}_k \in T(SO(3))$ with $\mathbf{e}_k$ for $1 \leq k \leq 3$ being the standard basis for $\mathbb{R}^3$
   2: Compute $\mathbf{H}_{kl} = \text{Hess}\, F(\mathbf{E}^k, \mathbf{E}^l)$, $1 \leq k, l \leq 3$
   3: Compute $\mathbf{g}_k = \text{d}\, F(\mathbf{E}^k)$, $1 \leq k \leq 3$
   4: Compute $\mathbf{u} = (u_1, u_2, u_3)^\top$ such that $\mathbf{u} = -\mathbf{H}^{-1}\mathbf{g}$
   5: The optimal updating vector
      $\mathbf{\Delta} = -\text{Hess}^{-1}\mathbf{G} = \mathbf{Q}\hat{\mathbf{u}}$

2. Update the rotation $\mathbf{Q}$

   1: Move $\mathbf{Q}$ in the direction $\mathbf{\Delta}$ along the geodesic to

   $$\exp(\mathbf{Q}, \mathbf{\Delta}t) = \mathbf{Q}\left(\mathbf{I} + \hat{\omega}\sin(t) + \hat{\omega}^2(1 - \cos(t))\right),$$

   where $t = \sqrt{\frac{1}{2}\text{tr}(\mathbf{\Delta}^\top \mathbf{\Delta})}$ and $\omega = \mathbf{Q}^\top\mathbf{\Delta}/t$

3.3 NRSFM with PLDA

PLDA was presented by Prince and Elder in [10] as a probabilistic estimation for deterministic linear discriminant analysis (LDA) [28]. This model separately models the between-individual and within-individual variations among different subjects. Unlike PCA, which only takes into account the whole data distribution, LDA seeks the maximum separability of classes along the direction that has the highest ratio of the variance between the classes to the variance within the classes [29]. Thus, for our PLDA model, the single shape subspace $\mathbf{V}$ in Eq. (1) is replaced by the between-individual subspace $\mathbf{F}$ and the within-individual subspace $\mathbf{K}$ as follows

$$\mathbf{s}_{ij} = \bar{\mathbf{s}} + \mathbf{F}\mathbf{h}_i + \mathbf{K}\mathbf{w}_{ij}, \qquad (30)$$

where i denotes the ith individual of the total I subjects and j denotes the jth image of J images belonging to this person. Compared to the original definition in Eq. (1), the latent variables $\mathbf{z}_t$ now consist of two parts. The first part, $\mathbf{h}_i$, indicates the parameter for the between-individual subspace $\mathbf{F}$, which remains constant for all J images of individual i, while the second part $\mathbf{w}_{ij}$ describes how each image varies in the within-individual subspace $\mathbf{K}$. Given this advanced shape model, the latent identity variables $\mathbf{h}_i$ guarantee that a great part of the commonness in the same subject is preserved and taken into account at runtime.

In order to estimate the PLDA parameters, an EM algorithm that is similar to PPCA is presented by Prince and Elder [10] with modifications in the E-step. The main point is to ensure that all J images share the same latent identity variable $\mathbf{h}_i$ despite the image-specific latent variables $\mathbf{w}_{ij}$. Therefore, the calculation of these J images is

done in the single step and the corresponding equations in Eq. (1) are stacked up into a composite matrix system

$$\begin{bmatrix} \mathbf{s}_{i1} \\ \mathbf{s}_{i2} \\ \vdots \\ \mathbf{s}_{iJ} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{s}} \\ \bar{\mathbf{s}} \\ \vdots \\ \bar{\mathbf{s}} \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{K} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{K} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i1} \\ \mathbf{w}_{i2} \\ \vdots \\ \mathbf{w}_{iJ} \end{bmatrix}, \qquad (31)$$

or equivalently

$$\mathbf{s}_i = \bar{\mathbf{s}}' + \mathbf{A}\mathbf{y}_i. \qquad (32)$$

Accordingly, the expectations of the new latent variables $\mathbf{y}_i$ for the E-step changes from Eqs. (6) and (7) to

$$\boldsymbol{\mu}_i' = \mathbb{E}[\mathbf{y}_i], \qquad (33)$$

$$\boldsymbol{\phi}_i' = \mathbb{E}\left[\mathbf{y}_i\mathbf{y}_i^\top\right] = \boldsymbol{\Sigma}_i + \boldsymbol{\mu}_i'\boldsymbol{\mu}_i'^\top. \qquad (34)$$

As for the M-step, most of the existing PPCA updates remain unchanged, if we replace the original shape matrix $\mathbf{V}$ with $[\mathbf{F}\quad\mathbf{K}]$ and each of the latent variables $\mathbf{z}_{ij}$ with $\begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{ij} \end{bmatrix}$. Accordingly, the objective log-likelihood function to be minimized in Eq. (8) can be modified to

$$\mathcal{L} = -\mathbb{E}\left[\sum_t \log p(\mathbf{p}_t|\mathbf{R}_t, \bar{\mathbf{s}}, [\mathbf{F}\quad\mathbf{K}], \sigma^2)\right]$$

$$= \frac{1}{2\sigma^2}\sum_t \mathbb{E}\left[\|\mathbf{p}_t - \mathbf{R}_t(\bar{\mathbf{s}} + [\mathbf{F}\quad\mathbf{K}]\begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{ij} \end{bmatrix})\|^2\right]$$

$$+ JT\log(2\pi\sigma^2). \qquad (35)$$

We apply the Gauss-Newton step [9] as well as our manifold extension of Newton's method to optimize the objective function.

4 Experiments

In this section, extensive experiments are conducted to validate the proposed approaches. Rotation recovery using the Newton's method on the manifold is first assessed on different datasets. Subsequently, performance of PLDA on generated data with multiple subjects is presented.

4.1 Setup

For our experiments, the evaluation criteria is the same as in [9], i.e., the sum of squared differences between estimated 3D shapes to ground truth depth: $\|\hat{\mathbf{s}}_{1:T} - \mathbf{s}_{1:T}\|_F^2$, with the camera rotation $\mathbf{R}$ also being applied to the 3D shape. As the ground truth for camera rotation is not given, we are not able to measure the absolute performance gain from our algorithm explicitly. However, the decreased reconstruction error implicitly assesses the effectiveness of rotation estimation in our algorithm.

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 7 of 12

Moreover, additive zero-mean Gaussian noise is imposed to analyze the robustness of reconstruction. The noise level is plotted as the ratio of the noise variance to the norm of the 2D measurements: $JT\sigma^2/\|\mathbf{p}_{1:T}\|_F$. The noise levels range from 0 to 30 % with 2 % step, and the trials for each noise level are averaged over 10 runs. Our test is carried out on two face datasets, i.e., the Vicon motion capture data Mocap-Face [9] and the Binghamton University 3D Facial Expression (BU-3DFE) dataset [30].

The Mocap-Face dataset [9] contains a single video, which captures a single male subject with 40 markers attached to his face. The video contains 316 frames in total. Sample frames from this dataset can be seen in Fig. 2. Throughout the video sequence, the subject made limited changes of facial expression and head pose. Note that the tracking is very accurate using the markers.

The BU-3DFE dataset [30] is originally created for 3D facial expression analysis. The complete dataset consists of 100 subjects, covering different ethnic groups. Seven facial expressions are performed at four intensity levels by each subject. We randomly select 300 frames from 100 subjects for our test, in contrast to the Mocap-Face dataset [9], where there exists only one subject. This is a more practical setup, since in many computer vision datasets, e.g., for face alignment [31, 32], only static images of multiple subjects are provided, where only a few samples of the same subject is available. Separate application of NRSFM for each single subject is then impossible. Random poses are generated by projecting the 3D landmarks to 2D. Note that temporal smoothness of the shapes is not valid in this dataset. The dataset provides manual annotation of 83 marker points as shown in Fig. 3. Due to labeling noise and inconsistency, this dataset contains noise in the original measurements.

4.2 Evaluation of manifold PPCA
We first give quantitative results for the recovery of the 3D camera rotation between our algorithm (manifold PPCA) and the baseline PPCA [9], as well as the other state-of-the-art approaches, point trajectory approach (PTA) [18] and column space fitting (CSF2) [20].

#### 4.2.1 Reconstruction results on Mocap-Face
In the first experiment without noise on Mocap-Face [9], our approach achieves slightly better performance

than PPCA, while both having an effective reconstruction result under 3 % error, as is plotted in Fig. 4a. Qualitative results are also shown in Fig. 5a, which yields similar outcome. In comparison, performance of methods in trajectory space is more sensitive to the number of DCT bases. Starting from $K = 8$, both PTA and CSF2 degrade abruptly.

In real life, there are no markers and the automatic point detectors are usually not stable. To assess the performance of the system in a real-world case, it is necessary to test it on the noisy data. We evaluate our system with additive Gaussian noise at different noise levels. As can be observed from Fig. 4c, at the beginning, all algorithms have almost the same error rate up to 6 % noise. With more noise added, PPCA starts to undergo a significantly steeper curve than our approach. Starting from 20 % noise level, our result gets 50 % lower error rate than PPCA. That is most likely because with more noise, it is more difficult for the rotation approximation in PPCA to find the right updating direction. Despite achieving the lowest error in the above noise-free experiment, the state-of-the-art CSF2 surprisingly fails to hold up well against noise, which approaches PPCA as the second worst. The same trajectory-based PTA is more stable, thanks to smoother DCT bases. Shaji and Chandran [15] also evaluated on the Mocap-Face dataset [9] with additive noise. From their plot, the performance degrades very quickly with noise level over 20 %. However, our probabilistic approach does not suffer from this problem. Additionally, the variance of the results of each noise level is also shown in the figure, in which we observe that the manifold PPCA also reduces error variances. That means our approach performs much more stably under noisy circumstances.

4.2.2 Reconstruction results on BU-3DFE
Since the BU–3DFE dataset [30] is a more difficult setup, the performance is lower compared to the test on the Mocap-Face dataset [9]. The purpose of this test is to see how well a generic face model can be generated using different NRSFM approaches. As can be seen in Fig. 4b, the recovered models cannot fit all instances as well as on the Mocap-Face dataset [9]. But again, the error level of our attempt is in overall ca. 8 % lower than that of PPCA regardless of the choice of K, which demonstrates a relative performance gain of 30 to 40 %. As can be observed
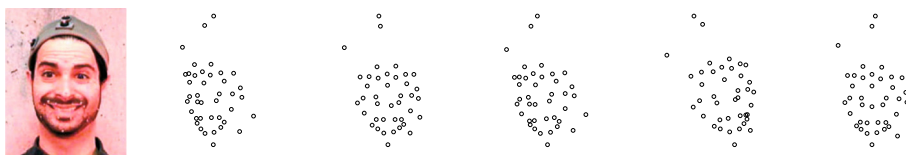
Fig. 2 The Mocap-Face dataset, captured by 40 markers attached to the face of the subject in the left image, 316 frames in total [9]

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 8 of 12
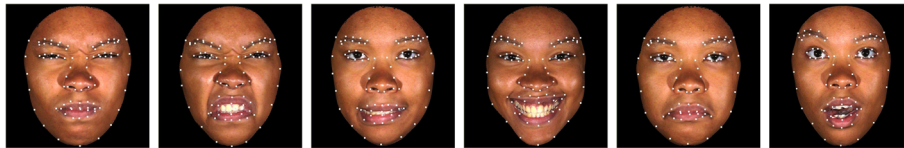
Fig. 3 The BU-3DFE dataset, with 83 points and 300 frames of multiple subjects, plus zero-mean Gaussian random pose changes [30]

qualitatively in Fig. 5b, PPCA's rotation approximation limits its result to getting better rotation estimate in frame 50. It also has difficulties to recover the contour of the faces correctly in frame 250, whereas our system clearly does better. It is also interesting to test the approaches in trajectory bases, where the smooth shape deformation assumption is no longer valid on the BU-3DFE dataset. As expected, CSF2 performs worse than the manifold PPCA, which does not take into account the temporal prior. Selection of K also has no influence to the error rate, unlike in Fig. 4a. If we add Gaussian noise to the data (see Fig. 4d), manifold PPCA and PTA degrades slightly slower than PPCA and CSF2, similar to that on the Mocap-Face dataset [9]. These results reveal that, when modeling more complicated shapes, an optimal rotation estimation using manifold optimization techniques is superior. Another advantage of our manifold-based approach is that it is more robust in noisy environments in general.
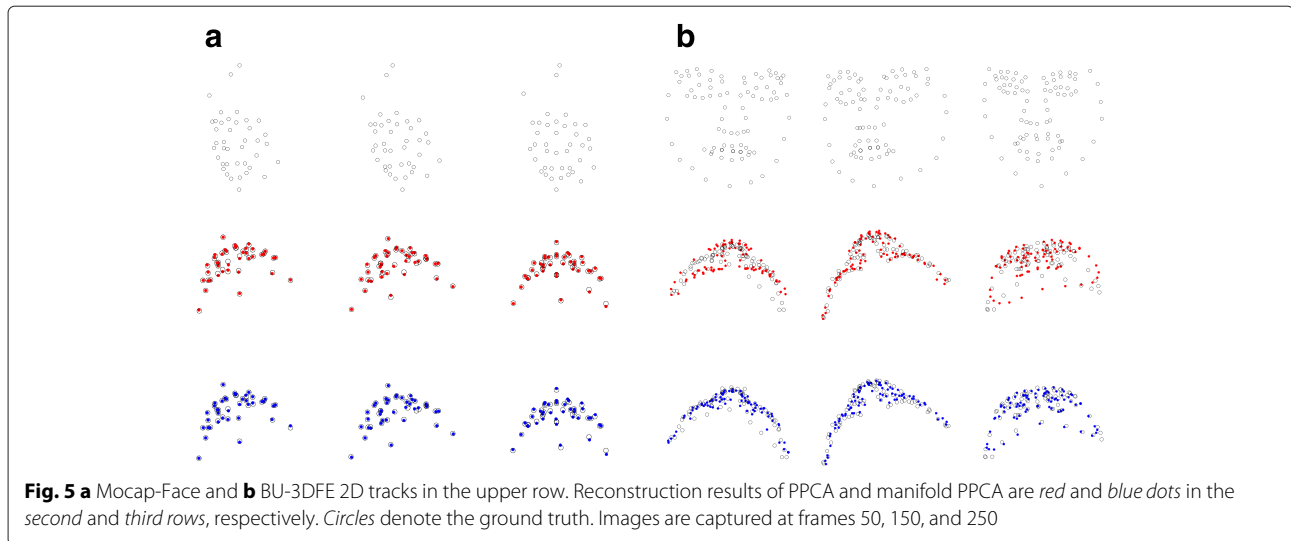
4.3 Evaluation of PLDA

In the second part of our experiments, we generate datasets with multiple subjects from the original BU-3DFE dataset [30] to evaluate the PLDA variant of the NRSFM algorithm in comparison with PPCA. We consider two setups with different number of subjects involved. For the first setup, we select six subjects and each subject has 50 images. For the second setup, there are 12 subjects with 25 images, respectively. Thus, the total number of frames is still 300. Similar to the PPCA experiments with additive noise, we randomly generate five input datasets for each setup to obtain statistically significant results.

For all tests in this section, we directly compare the results of PPCA and PLDA as well as the influence of imposing Newton's method on the manifold for recovery of the rotation matrix to both probabilistic frameworks. Since in Section 4.2.2, trajectory-based methods
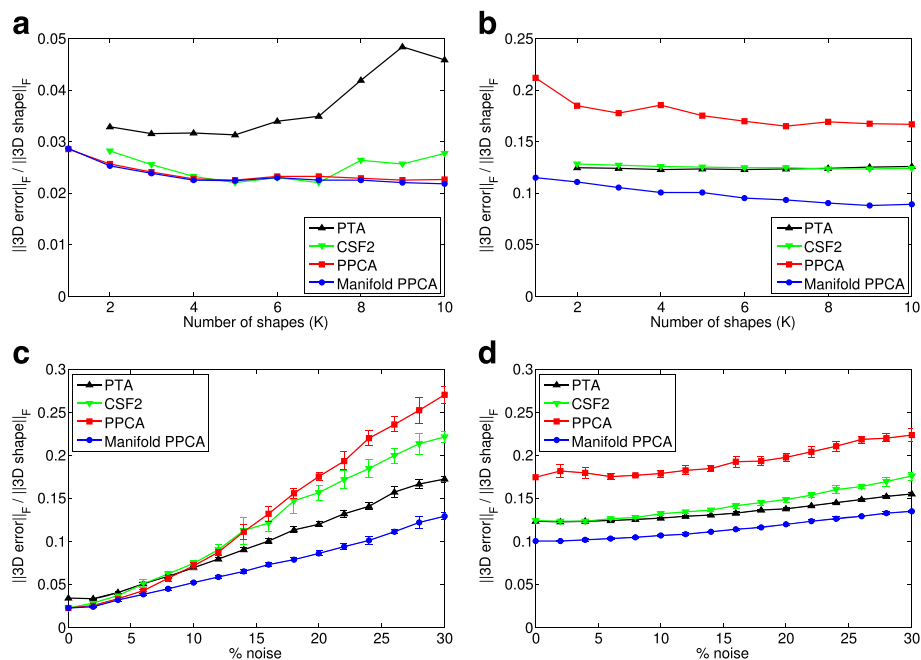
Fig. 4 The first row shows the reconstruction error as a function of bases (*K*) without adding noise manually on **a** Mocap-Face and **b** BU-3DFE. The second row shows the reconstruction error with additive Gaussian noise up to 30 % on **c** Mocap-Face and **d** BU-3DFE

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 9 of 12

Fig. 5 a Mocap-Face and **b** BU-3DFE 2D tracks in the upper row. Reconstruction results of PPCA and manifold PPCA are *red* and *blue dots* in the *second* and *third rows*, respectively. *Circles* denote the ground truth. Images are captured at frames 50, 150, and 250

are proven not to generalize well when temporal smoothness does not hold, the results of PTA and CSF2 are not included. The curves of both PPCA approaches are plotted in dashed lines while the PLDA results are plotted in solid lines.

In the test case without additive noise, we fix the number of between-subject shape bases $F = 3$ and vary the within-subject bases G in PLDA from 1 to 7, compared to the only shape bases K in PPCA that equals the sum of F and G. We first notice that with the help of PLDA, the performance for both rotation recovery techniques has got further improvement, independently from the input data with 6 subjects (Fig. 6a) or 12 subjects (Fig. 6b). Especially for the baseline PPCA algorithm with Gauss-Newton rotation approximation, there is nearly 10 % less error in the reconstruction in both setups. Moreover, the error variance also drops hugely to an acceptable level. For our Newton's rotation recovery method on the manifold, only little improvement in reconstruction error is observed; however, the performance without PLDA already delivers a satisfactory result. As a result, all of our proposed methods, i.e., manifold PPCA, PLDA and manifold PLDA, manage to make significant performance enhancement. The similar outcome of the manifold PPCA and PLDA, which employ different extensions for the optimization problem, indicates that the achieved result may have approximated to the performance limit of the dataset using the probabilistic framework.

When zero-mean Gaussian noise is added (Fig. 6c, d), although the gaps between the manifold PLDA and PLDA remain close, the lower average error rate and the stability with notably less error deviation at some noise levels again demonstrate the effectiveness and necessity of our better rotation recovery. We also observe that both PLDA-based methods degrade faster than those PPCA-based methods

with the amount of imposed Gaussian noise starting from ca. 15 to 20 %. We conclude that its reason is probably because PLDA-based approaches need to estimate more parameters in the E-step than PPCA in each iteration, which makes the additional noise and uncertainty a decisive deficit factor for the approach. But overall, introducing PLDA does help to further decrease the error reconstruction rate with or without additive noise.

Qualitative experiments are also conducted in order to review the effect of applying PLDA on datasets with more than one subject, as can be seen in Fig. 7. We know that from Eq. (30), the between-individual linear shape model of PLDA consists of the global mean shape $\bar{\mathbf{s}}$ plus the subject-specific shape term $\mathbf{Fh}_i$. The frame-specific shape term $\mathbf{Kw}_{ij}$ serves solely as within-individual variance and is therefore omitted in this experiment. Thus, in Fig. 7, the reconstruction of every single subject for each dataset given by the first two terms in Eq. (30) is shown. The 3D ground truth is obtained by averaging all 50 shape vectors for the corresponding subject. As expected, the reconstruction result is fairly satisfactory, thanks to the characteristics of PLDA. Even for some unique faces as in Fig. 7e, g, their contours and facial features are still well modeled, which again proves the capability and tolerance of our approach.

In Fig. 8, we illustrate the reconstructed bases $\mathbf{F}$ and $\mathbf{K}$. The effect of varying the between-individual ($\mathbf{F}$) and within-individual ($\mathbf{K}$) shape bases learned by PLDA between ± 3 standard deviations from the mean value is analyzed. With the first three bases of $\mathbf{F}$, different eye, face contour, and mouth types are modeled, respectively, in Fig. 8b, c, and d. It is interesting to see that the variations in the figures are more related to identification of the subjects than to the expressions. With the bases of $\mathbf{K}$ instead, different facial expressions present in BU-3DFE

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 10 of 12
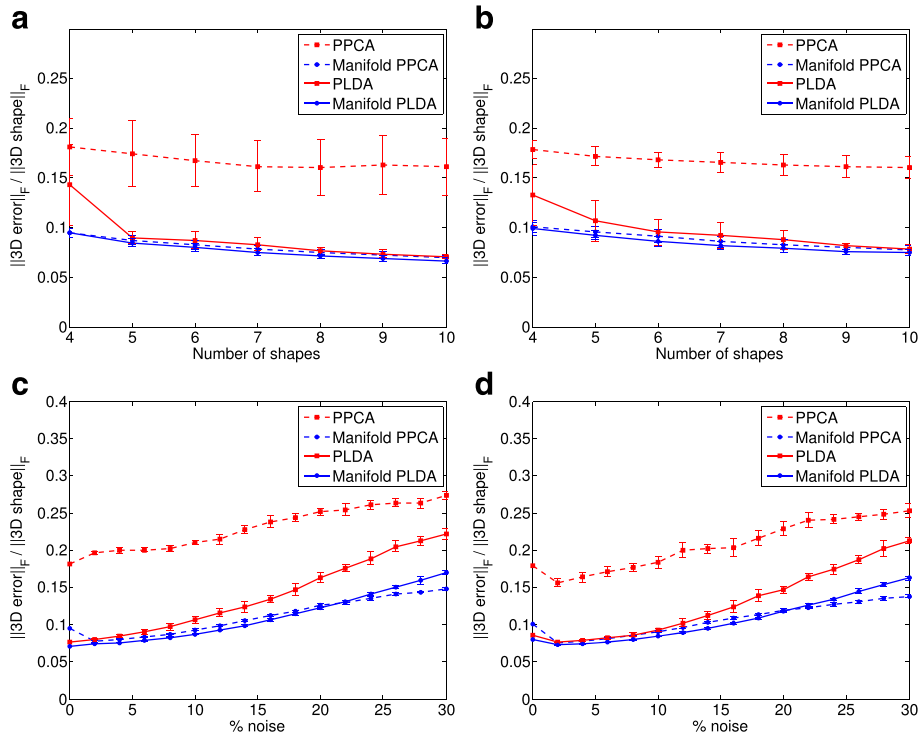
Fig. 6 The *first row* shows the reconstruction error without adding noise manually using PPCA and PLDA. The *second row* shows the reconstruction error with additive Gaussian noise up to 30 %. **a** and **c** in the *left column* are evaluated on datasets with 6 subjects, and **b** and **d** in the *right column* are evaluated on datasets with 12 subjects
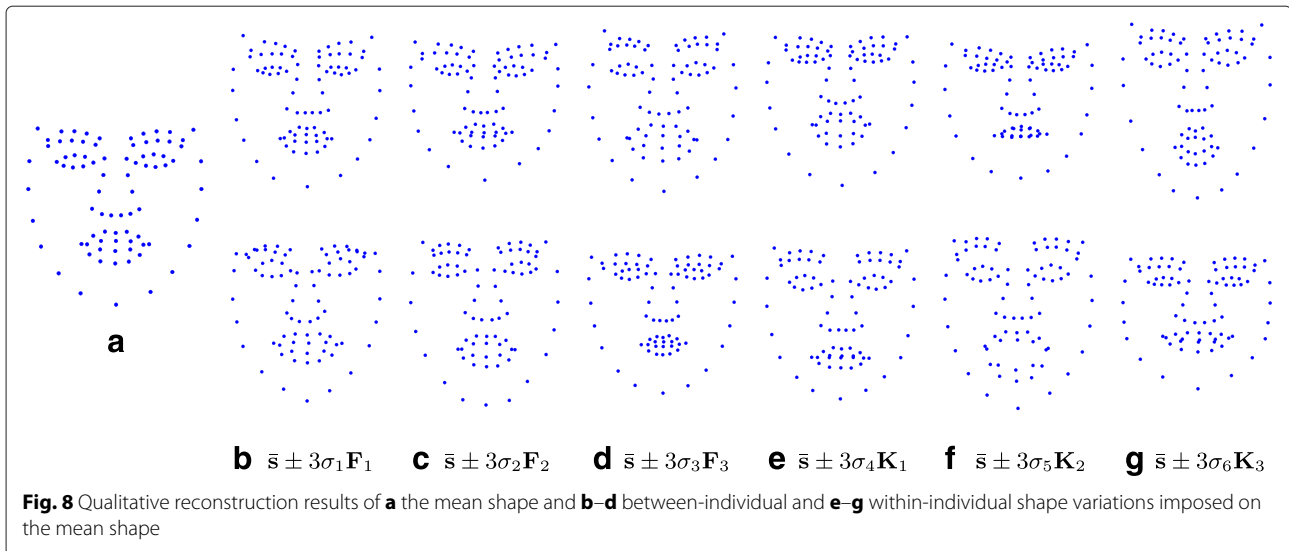
[30] are well recognizable. For example, Fig. 8e shows opening and closing mouths. Evolution from angry to fear is illustrated in Fig. 8f and from surprise to happiness in Fig. 8g, respectively. Those results fully meet our expectations and conform to the characteristics of PLDA, which provide optimally and more meaningfully reconstructed shape bases than those given by PPCA.

5 Conclusions

In this work, we have presented a novel solution to unleash the orthonormality constraints of the camera rotation matrix in the NRSFM problem. Without requiring conducting complex approximations, performing rotation update on the $SO(3)$ manifold implicitly ensures the validity of the constraints. In the experiments
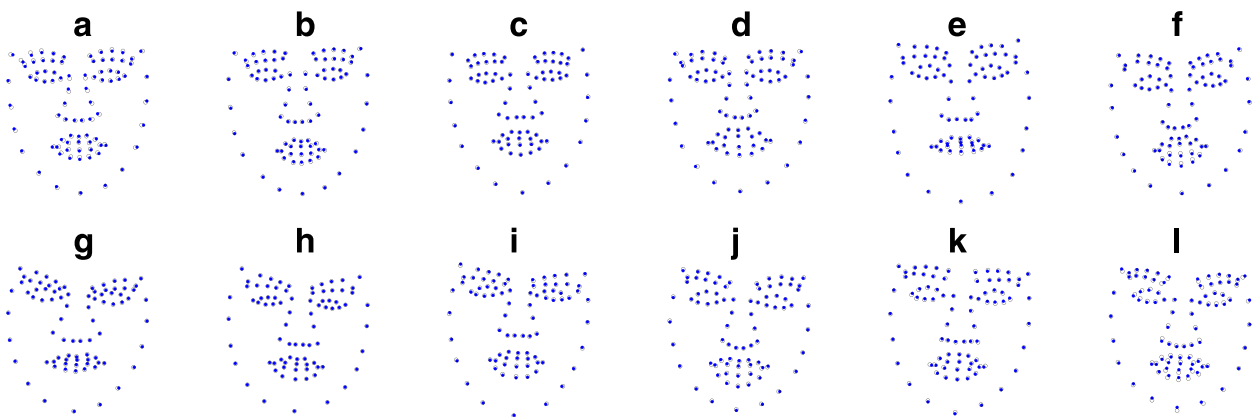
Fig. 7 a–l Qualitative reconstruction results of different subjects using PLDA. Each *row* corresponds to a generated dataset consisting of six subjects based on BU-3DFE

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 11 of 12

$$\mathbf{b} \quad \bar{\mathbf{s}} \pm 3\sigma_1 \mathbf{F}_1 \qquad \mathbf{c} \quad \bar{\mathbf{s}} \pm 3\sigma_2 \mathbf{F}_2 \qquad \mathbf{d} \quad \bar{\mathbf{s}} \pm 3\sigma_3 \mathbf{F}_3 \qquad \mathbf{e} \quad \bar{\mathbf{s}} \pm 3\sigma_4 \mathbf{K}_1 \qquad \mathbf{f} \quad \bar{\mathbf{s}} \pm 3\sigma_5 \mathbf{K}_2 \qquad \mathbf{g} \quad \bar{\mathbf{s}} \pm 3\sigma_6 \mathbf{K}_3$$

Fig. 8 Qualitative reconstruction results of **a** the mean shape and **b**–**d** between-individual and **e**–**g** within-individual shape variations imposed on the mean shape

on the Mocap-Face dataset [9] with additional noise, which contains only one subject, our approach performs significantly better by reducing up to 50 % reconstruction error. Furthermore, the proposed PLDA approach successfully extends the existing probabilistic framework to separately model between-subject and within-subject shape variations during the alternating optimization for datasets with multiple identities. On the BU-3DFE dataset [30] with multiple subjects and manually annotated landmarks, we clearly outperform the baseline approaches in all tests. To conclude, we have shown that the proposed approaches are robust against noise, which indicates that they are more capable of dealing with real-world data. In addition to its robustness, our approaches generalize better on datasets with multiple subjects.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
In this work, two major contributions have been made. We first investigated the geometric properties of the orthonormality constraints and generalized the Newton's optimization method to the underlying manifold of the camera rotation matrices. That means, non-linear optimization can be carried out on the manifold without any imprecise approximations. We used a PPCA-based framework [9] to model NRSFM as it is more robust to noise than the closed-form factorization techniques. Our second contribution deals with multiple subjects. Current NRSFM algorithms mostly focus on the reconstruction of a single subject. While dealing with data containing multiple subjects, no difference is taken into account, when modeling between-individual variation (e.g., face model of different identities) and within-individual variation (e.g., facial expression of the same identity). For that reason, we extend the PPCA-based framework to the PLDA [10] model to improve reconstruction performance on data with multiple subjects.

Author details
[1] Vision and Fusion Laboratory (IES), Karlsruhe Institute of Technology (KIT), Adenauerring 4, 76131 Karlsruhe, Germany. [2] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (Fraunhofer IOSB), Fraunhoferstr. 1, 76131 Karlsruhe, Germany. [3] Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), EPFL STI IEL LTS5, ELD 241 (Bâtiment ELD), Station 11, CH–1015 Lausanne, Switzerland. [4] Faculty of Computer and Informatics, Istanbul Technical University (ITU), 34469, Maslak, Istanbul, Turkey.

References
1. C Tomasi, T Kanade, Shape and motion from image streams under orthography: a factorization method. Int. J. Comput. Vis. **9**(2), 137–154 (1992)
2. V Blanz, T Vetter, in *Proceedings of the Annual International Conference on Computer Graphics and Interactive Techniques*. A morphable model for the synthesis of 3D faces, (1999), pp. 187–194
3. S Ullman, Maximizing rigidity: The incremental recovery of 3-D structure structure from rigid and nonrigid motion. Perception. **13**, 255–274 (1983)
4. M Brand, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Morphable 3D models from video, vol. 2, (2001), pp. 456–463
5. M Brand, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. A direct method for 3D factorization of nonrigid motion observed in 2D, vol. 2, (2005), pp. 122–128
6. L Torresani, DB Yang, EJ Alexander, C Bregler, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Tracking and modeling non-rigid objects with rank constraints, vol. 1, (2001), pp. 493–500
7. L Torresani, A Hertzmann, in *Proceedings of the European Conference on Computer Vision*. Automatic non-rigid 3D modeling from video, (2004), pp. 299–312
8. J Xiao, J Chai, T Kanade, A closed-form solution to non-rigid shape and motion recovery. Int. J. Comput. Vis. **67**(2), 233–246 (2006)
9. L Torresani, A Hertzmann, C Bregler, Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors. IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 878–892 (2008)
10. SJD Prince, JH Elder, in *Proceedings of the IEEE International Conference on Computer Vision*. Probabilistic linear discriminant analysis for inferences about identity, (2007), pp. 1–8
11. C Bregler, A Hertzmann, H Biermann, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Recovering non-rigid 3D shape from image streams, vol. 2, (2000), pp. 690–696

Qu *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:45

Page 12 of 12

12. J Xiao, T Kanade, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Non-rigid shape and motion recovery: degenerate deformations, vol. 1, (2004), pp. 668–675

13. I Akhter, Y Sheikh, S Khan, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. In defense of orthonormality constraints for nonrigid structure from motion, (2009), pp. 1534–1541

14. Y Dai, H Li, M He, A simple prior-free method for non-rigid structure-from-motion factorization. Int. J. Comput. Vis. **107**(2), 101–122 (2014)

15. A Shaji, S Chandran, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Riemannian manifold optimisation for non-rigid structure from motion, (2008), pp. 1–6

16. V Rabaud, S Belongie, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Linear embeddings in non-rigid structure from motion, (2009), pp. 2427–2434

17. L Tao, BJ Matuszewski, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Non-rigid structure from motion with diffusion maps prior, (2013), pp. 1530–1537

18. I Akhter, Y Sheikh, S Khan, T Kanade, in *Advances in Neural Information Processing Systems*. Nonrigid structure from motion in trajectory space, (2008), pp. 41–48

19. PFU Gotardo, AM Martinez, Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10), 2051–2065 (2011)

20. PFU Gotardo, AM Martinez, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Non-rigid structure from motion with complementary rank-3 spaces, (2011), pp. 3065–3072

21. J Valmadre, S Lucey, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. General trajectory prior for non-rigid reconstruction, (2012), pp. 1394–1401

22. HS Park, T Shiratori, I Matthews, Y Sheikh, 3D trajectory reconstruction under perspective projection. Int. J. Comp. Vis. **115**(2), 115–135 (2015)

23. ME Tipping, CM Bishop, Probabilistic principal component analysis. J. R. Stat. Soc. **61**, 611–622 (1999)

24. DJ Bartholomew, *Latent Variable Models and Factor Analysis*. (Charles Griffin & Co. Ltd., London, 1987)

25. A Edelman, TA Arias, ST Smith, The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1999)

26. Y Ma, S Košecká, S Sastry, Optimization criteria and geometric algorithms for motion and structure estimation. Int. J. Comput. Vis. **44**(3), 219–249 (1999)

27. RM Murray, Z Li, SS Sastry, *A Mathematical Introduction to Robotic Manipulation*. (CRC Press, Boca Raton, FL, 1994)

28. PN Belhumeur, J Hespanha, DJ Kriegman, Eigenfaces *vs*, Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)

29. AM Martínez, AC Kak, PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 228–233 (2001)

30. L Yin, X Wei, Y Sun, J Wang, MJ Rosato, in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*. A 3D facial expression database for facial behavior research, (2006), pp. 211–216

31. R Gross, I Matthews, J Cohn, T Kanade, S Baker, in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*. Multi-PIE, (2008), pp. 1–8

32. K Messer, J Matas, J Kittler, J Luettin, G Maître, in *Proceedings of the International Conference on Audio and Video-based Biometric Personal Verification*. XM2VTSDB: The extended M2VTS database, (1999), pp. 72–77