Subspace Regression: Predicting a Subspace from One Sample

Minyoung Kim[†] Zengyin Zhang[†] Fernando de la Torre[†] Wende Zhang[‡]

[†] Robotics Institute, Carnegie Mellon University
[‡] Electrical & Controls Integration Lab, General Motors R&D

Abstract. Subspace methods have been extensively used to solve a variety of problems in computer vision including object detection, recognition, and tracking. Typically, subspaces are learned from a training set that contains different configurations of a particular object (e.g., variations on shape or appearance). However, in some situations it is not possible to have access to data with multiple configurations of an object. For instance, consider the problem of predicting a person-specific subspace of the pose variation from only a frontal face image, by learning a mapping between frontal images and the corresponding pose subspaces in training samples. We refer to this problem as *subspace regression*.

Subspace regression is a challenging problem for two main reasons: (i) it involves a mapping between high-dimensional spaces, (ii) it is unclear how to parameterize the mapping between one sample and a subspace. We propose four methods to learn a mapping from one sample to a subspace: Individual Mapping on Images, Direct Mapping to Subspaces, Regression on Subspaces, and Direct Subspace Alignment. We show the validity of our approaches to build a person-specific face subspace of pose or illumination, and its applications to face tracking and recognition.

1 Introduction

Since the early work of Sirovich and Kirby [1] parameterizing the human face using Principal Component Analysis (PCA) [2] and the successful eigenfaces of Turk and Pentland [3], many computer vision researchers have used subspace techniques to construct linear models of optical flow, shape or gray level for tracking [4,5], detection [3] and recognition [6]. The modeling power of subspace techniques is especially useful when applied to visual data, because there is a need for dimensionality reduction given the increase in the number of features. Typically, subspaces are learned from a set of registered training samples. For instance, consider the problem of building a person-specific image subspace that models the variation across pose. Building this subspace often requires a large number of training images sampled from the underlying person manifold across viewpoints, typically a set of images of all possible pose variations. Once the data are collected we can compute PCA (Fig. 1 (Left)) to learn a person-specific pose subspace. However, in general this procedure typically incurs a costly data

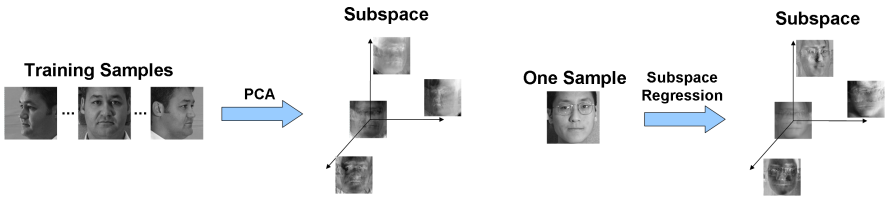


Fig. 1. Illustration of the proposed subspace regression approach. (Left) Traditional subspace learning. A person-specific pose subspace is learned by performing PCA on a set of training samples. (Right) Subspace regression can predict a person-specific pose subspace from only one sample (e.g., frontal image).

collection step: for every new test subject, one has to gather images of that same subject for all possible poses (often, illuminations and expressions).

A relatively unexplored problem in computer vision is how to learn a subspace from a single sample (i.e., one image). We refer to this problem as *subspace regression*. Fig. 1 (Right) illustrates the main goal of subspace regression. Predicting a subspace from only one sample is a challenging problem mainly due to the high dimensionality of the data, typically with fewer training samples than the number of features. Moreover, it is unclear which is the best way to effectively and efficiently parameterize the mapping and the subspace. We first suggest two relatively straightforward solutions which are based on learning regressors from a query sample (e.g., a frontal image) to samples of all other states[1] or the subspace *itself*. Despite their conceptual simplicity, these approaches have a potential problem of unreliable function estimation originated from very high-dimensional input/output space (e.g., image dimension of several thousands).

To address the issue of estimating a large number of parameters, we next propose *Regression on Subspaces* (ROS), a novel generative-discriminative approach that performs regression on the subspace coordinates. ROS can yield a reliable estimator with a significantly reduced number of parameters. We show that ROS can be seen as a tri-factor reduced-rank regression which aggressively reduces the number of parameters in a sensible manner. Additionally, using ROS parameterization we propose *Direct Subspace Alignment* (DSA), that directly finds mappings between the tri-factor parameterization of the subspace and the training subspaces. We demonstrate the validity of the four approaches to predict a person-specific subspace of pose or illumination, and its applications to face tracking and recognition.

The rest of the paper is organized as follows. After briefly reviewing the background on ridge regression and reduced-rank regression in Sec. 2, we propose the four subspace regression approaches in Sec. 3. We discuss prior work related to ours in Sec. 4, and demonstrate the efficacy of the proposed approaches in the experiments in Sec. 5.

[1] The *state* is defined as a possible mode of variation (e.g., pose).

2 Background

2.1 Ridge Regression

Let $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ be n training samples, $\mathbf{x}_i \in \Re^{p\times 1}$ being a p-dimensional input sample and $\mathbf{y}_i \in \Re^{d\times 1}$ a d-dimensional output sample. In the linear[2] regression the prediction function has a linear form, $\mathbf{f}(\mathbf{x}) = \mathbf{W}^\top \mathbf{x}$, where $\mathbf{W} \in \Re^{p\times d}$ is the matrix to be learned. We use the matrix notation for input, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, which is of dimension $(p \times n)$. Similarly, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ is of dimension $(d \times n)$. Following the regularized empirical risk minimization framework, we minimize:

$$\min_{\mathbf{W}} ||\mathbf{Y} - \mathbf{W}^\top \mathbf{X}||_F^2 + \lambda ||\mathbf{W}||_F^2, \tag{1}$$

where λ controls the degree of regularization. (1) is often called the ridge regression, which admits a closed-form solution: $\mathbf{W} = (\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I})^{-1}\mathbf{X}\mathbf{Y}^\top$.

2.2 Reduced-Rank Regression

Since its introduction in the early 1950s by Anderson [7], the reduced-rank regression (RRR) model has inspired a wealth of diverse applications in several fields including computer vision [8]. The RRR learns a mapping between input $\mathbf{x}$ and output $\mathbf{y}$ by minimizing:

$$\min_{\mathbf{A},\mathbf{C}} \sum_{i=1}^{n} ||\mathbf{y}_i - \mathbf{A}\mathbf{C}^\top \mathbf{x}_i||_2^2, \tag{2}$$

where $\mathbf{A} \in \Re^{d\times q}$ and $\mathbf{C} \in \Re^{p\times q}$. One may guess that $\mathbf{A}$ and $\mathbf{C}$ can be obtained from the singular value decomposition (SVD) of the learned $\mathbf{W}$ from (1). However, as shown in [9], this often yields results inferior results to simultaneously optimizing $\mathbf{A}$ and $\mathbf{C}$ from the least square optimization. And, it is known that the least square solution has no local minima as it reduces to learning the canonical correlation analysis (CCA) [10] embedding on $\mathbf{x}$ (to learn $\mathbf{C}$) followed by the least square regression estimation for $\mathbf{A}$ from the embedding of $\mathbf{x}$, i.e., $\mathbf{C}^\top \mathbf{x}$, to $\mathbf{y}$.

3 Subspace Regression

This section describes four methods that learn mappings between one sample and a subspace: Individual Mapping on Images (IMI), Direct Mapping to Subspaces (DMS), Regression on Subspaces (ROS), and Direct Subspace Alignment (DSA). Throughout the paper we assume that there are n instances (i.e. subjects) $(i = 1, \ldots, n)$, where for each instance i, there are $(K+1)$ samples (denoted by $\mathbf{x}_i^s$) of different states (e.g. pose) $s \in \{0, 1, \ldots, K\}$. These samples can be regarded as the $(K+1)$ realizable variations for the instance i. The state $s = 0$ is reserved

[2] The nonlinear extension is straightforward via the kernel tricks.

for indicating the reference state. Hence, the reference sample for i is $\mathbf{x}_i^0$. For instance, in the problem of predicting the person-specific subspace from a frontal image, the training data consist of n subjects, where for each subject i, we have $(K+1)$ images of different states (e.g., facial poses, illuminations, expressions). When the pose is only the state we consider (i.e., the person-specific subspace accounting for only the pose variability), one can let $s=0$ correspond to a frontal pose. We presume that each training image $\mathbf{x}_i^s$ is labeled with the subject identity i and the state s. In the testing phase, our goal is to predict the subspace of a new subject from a given frontal image.

3.1 Individual Mapping on Images (IMI)

In IMI, we learn for each state $s=1,\ldots,K$, a regression function that maps the reference sample $\mathbf{x}^0$ to the s-state sample $\mathbf{x}^s$. This enables us to estimate the subspace for a new instance (or subject) using PCA with the generated images from the learned regressors. Formally, we form the training data for regression as: $\{(\mathbf{x}_i^0, \mathbf{x}_i^s)\}_{i=1}^n$, for $s=1,\ldots,K$. In the reduced-rank model we solve:

$$\min_{\{\mathbf{A}_s\},\{\mathbf{C}_s\}} \sum_{s=1}^{K}\sum_{i=1}^{n} ||\mathbf{x}_i^s - \mathbf{A}_s\mathbf{C}_s^\top \mathbf{x}_i^0||_2^2. \tag{3}$$

Here, $\mathbf{A}_s \in \Re^{p\times l}$ and $\mathbf{C}_s \in \Re^{p\times l}$ for $s=1,\ldots,K$ are the parameters of the model, where $l(\ll p)$ is the reduced-rank dimension. Once we solve (3), given a query $\mathbf{x}_*^0$ of an unseen subject $*$, the synthesized sample at state s becomes:

$$\mathbf{x}_*^s = \mathbf{A}_s\mathbf{C}_s^\top \mathbf{x}_*^0. \tag{4}$$

In this way, we generate samples for all possible states, $\mathbf{x}_*^s$ for $s=1,\ldots,K$, from which we can build a person-specific subspace for the subject $*$ via PCA.

3.2 Direct Mapping to Subspaces (DMS)

DMS learns a direct regression between the reference sample ($\mathbf{x}^0$) and the subspace built with all other samples at different states ($\mathbf{x}^s$, for $s=1,..K$). In this setting, the training data can be formed as $\{(\mathbf{x}_i^0, (\boldsymbol{\mu}_i, \mathbf{B}_i))\}_{i=1}^n$, where the output point $(\boldsymbol{\mu}_i, \mathbf{B}_i)$ is the subspace of the instance (subject) i, typically learned via PCA from the training samples $\{\mathbf{x}_i^s\}_{s=0}^K$. We regard the output as the concatenated vector of the mean $\boldsymbol{\mu}_i$ and the vectorized basis matrix $\mathbf{B}_i$. Similar to the IMI approach, as the output consists of many variables (e.g., mean and eigenvectors), we consider reduced-rank regression.

3.3 Regression on Subspaces (ROS)

The above two approaches tend to perform poorly due to the high dimensionality of the output space (either images or subspaces). The reduced-rank regression has been adopted, but it still requires many parameters to achieve the desired

Algorithm 1 Regression on Subspace (Training)

Input: Samples $\{\mathbf{x}_i^s\}$ for $i = 1, \ldots, n$, $s = 0, \ldots, K$.
Output: Learned $\mathbf{g}_s \in \Re^{r \times r}$ for $s = 1, \ldots, K$.
1) For $s = 0, \ldots, K$, PCA-learn a state subspace $\mathcal{S}_s = (\mathbf{m}_s, \mathbf{A}_s)$ with data $\{\mathbf{x}_i^s\}_{i=1}^n$.
2) Project $\{\mathbf{x}_i^s\}$ onto the state subspace $\mathcal{S}_s$ (5).
3) Learn a regressor for each state $s = 1, \ldots, K$: $\min_{\mathbf{g}_s} \sum_{i=1}^n ||\mathbf{z}_i^s - \mathbf{g}_s^\top \mathbf{z}_i^0||_2^2$

Algorithm 2 Regression on Subspace (Testing)

Input: Test reference sample $\mathbf{x}_*^0$ and the learned $\{\mathbf{g}_s\}_{s=1}^K$.
Output: Learned person subspace for $*$, $(\boldsymbol{\mu}_*, \boldsymbol{B}_*)$.
1) Project $\mathbf{x}_*^0$ onto the 0-state subspace: $\mathbf{z}_*^0 = \mathbf{A}_0^\top (\mathbf{x}_*^0 - \mathbf{m}_0)$.
2) Apply regression on subspace for $s = 1, \ldots, K$: $\mathbf{z}_*^s = \mathbf{g}_s^\top \mathbf{z}_*^0$
3) Synthesize a sample for each state $s = 1, \ldots, K$: $\mathbf{x}_*^s = \mathbf{A}_s \mathbf{z}_*^s + \mathbf{m}_s$
4) PCA-learn a person-specific subspace $(\boldsymbol{\mu}_*, \boldsymbol{B}_*)$ with $\{\mathbf{x}_*^s\}_{s=1}^K$ and $\mathbf{x}_*^0$.

accuracy. In this section we propose a new method called *Regression on Subspaces* (ROS) that learns a high-dimensional mapping with significantly fewer parameters.

The overall concept of ROS is illustrated in Fig. 2. We first learn the state-specific PCA subspaces, each of which is learned from samples (images) of the same state across different instances (subjects). For the state s, we denote the r-dim state-specific subspace by $\mathcal{S}_s = (\mathbf{m}_s, \mathbf{A}_s)$, where $\mathbf{m}_s \in \Re^p$ is the mean and $\mathbf{A}_s \in \Re^{p \times r}$ contains the basis vectors in its columns. $\mathbf{A}_s$ is learned via PCA[3]. Notice that $\mathcal{S}_s$ captures the variability solely in the styles (i.e., subjects), not in the contents (i.e., states). Once we have learned the state-specific subspaces, one can project $\mathbf{x}_i^s$, the image of subject i in the state s, onto $\mathcal{S}_s$, yielding the subspace coordinate $\mathbf{z}_i^s \in \Re^r$ derived as:

$$\mathbf{z}_i^s = \mathbf{A}_s^\top (\mathbf{x}_i^s - \mathbf{m}_s) \tag{5}$$

Although the state-specific subspaces are learned individually and independently, one can relate them to one another by introducing certain restrictions to the subspaces. Here we implicitly impose such constraints by considering mappings from one subspace to another. More specifically, we presume that there exists a regression matrix $\mathbf{g}_s \in \Re^{r \times r}$ for each state s $(= 1, \ldots, K)$ that maps the 0-state coordinate $\mathbf{z}^0$ to the subspace-s coordinate $\mathbf{z}^s$. That is,

$$\mathbf{z}^s = \mathbf{g}_s^\top \mathbf{z}^0. \tag{6}$$

One can naturally treat $\mathbf{g}_s$ as a subspace morphing function from $\mathcal{S}_0$ to $\mathcal{S}_s$, accounting for how a frontal image representation can be transformed into the representation of the state s in a subject-generic manner. Even though either

[3] Ideally CCA can be optimal, however, in our experiments due to the small number of training data points, PCA often outperformed CCA.

IMI or DMS explicitly aims at such a goal, it suffers from the high dimensionality of both input and output.

On the other hand, ROS operates on the coordinates of the subspaces, being more robust to noise with significantly fewer parameters to be learned. By regarding the pairs $\{(\mathbf{z}_i^0, \mathbf{z}_i^s)\}_{i=1}^n$ as samples independently and identically drawn from an underlying probability distribution on the joint subspace $(\mathcal{S}_0, \mathcal{S}_s)$, one can learn $\mathbf{g}_s$ using the regression algorithms we discussed before.

Once we obtain the mappings among the state subspaces, we can synthesize out-of-state images $\{\mathbf{x}_*^s\}_{s=1}^K$ for a new subject $*$, given its 0-state image $\mathbf{x}_*^0$. Finally, with the synthesized images $\{\mathbf{x}_*^s\}_{s=1}^K$, one can learn a subspace for $*$ with the data $\{\mathbf{x}_*^s\}_{s=1}^K \cup \{\mathbf{x}_*^0\}$ using PCA. The overall training/testing algorithms are shown in Alg. 1 and 2.

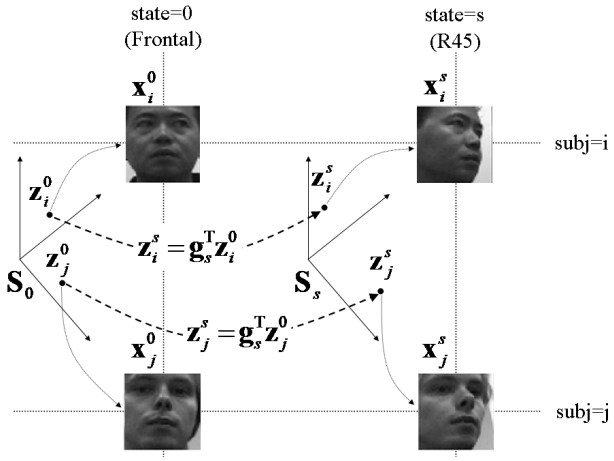


Fig. 2. Regression on Subspace (ROS) approach.

ROS as Tri-Factor Reduced-Rank Regression Recall from Sec 3.1 that the IMI solves the following reduced-rank regression:

$$\min_{\{\mathbf{R}_s\}} \sum_{s=1}^{K} \sum_{i=1}^{n} ||\mathbf{x}_i^s - \mathbf{R}_s \mathbf{x}_i^0||_2^2, \quad \text{where} \quad \mathbf{R}_s = \mathbf{A}_s \mathbf{C}_s^\top, \tag{7}$$

which can be seen as a *two-factor* parametrization of the regression coefficient $\mathbf{R}_s$. Although the reduced-rank regressor results in a smaller number of parameters by dyadic factorization, the ROS can be seen as a more aggressive factorization with significantly fewer parameters. To see this, we rewrite the ROS objective in a more general perspective. The optimization problem we impose for the ROS can be expressed as:

$$\min_{\{\mathbf{A}_s\},\{\mathbf{g}_s\}} \left[\sum_{s=0}^{K} \sum_{i=1}^{n} ||\mathbf{x}_i^s - \mathbf{A}_s \mathbf{A}_s^\top \mathbf{x}_i^s||_2^2 + \eta \sum_{s=1}^{K} \sum_{i=1}^{n} ||\mathbf{A}_s^\top \mathbf{x}_i^s - \mathbf{g}_s^\top \mathbf{A}_0^\top \mathbf{x}_i^0||_2^2 \right]. \tag{8}$$

Here, for expositional simplicity, we dropped the subspace means $\mathbf{m}_s$, assuming that the data points are centered.

The first term of (8) is responsible for the state-wise PCA subspace learning (i.e., $\mathbf{A}_s$ for $s = 0, 1, \ldots, K$), while the second term corresponds to the subspace regression from 0-state subspace ($\mathcal{S}_0$) to s-state subspace ($\mathcal{S}_s$). The two terms are balanced by the positive constant η. The way we solved this problem earlier can be seen as coordinate descent optimization: we first learn $\mathbf{A}_s$ while ignoring the second term, then we fix the learned $\mathbf{A}_s$'s, and optimize the second term with respect to $\{\mathbf{g}_s\}$ alone.

Now we replace $\mathbf{A}_s^\top \mathbf{x}_i^s$ in the first term by $\mathbf{g}_s^\top \mathbf{A}_0^\top \mathbf{x}_i^0$ in the second term, as we try to make the second term become 0. This then leads to:

$$\min_{\{\mathbf{A}_s\},\{\mathbf{g}_s\}} \sum_{s=1}^{K} \sum_{i=1}^{n} ||\mathbf{x}_i^s - \mathbf{A}_s \mathbf{g}_s^\top \mathbf{A}_0^\top \mathbf{x}_i^0||_2^2. \tag{9}$$

Note that (9) can be seen as a noise-free version of (8) by forcing the subspace regression error (i.e., the second term in (8)) to be 0. Compared to the reduced-rank regression, (9) further factorizes the post-multiplying matrix $\mathbf{C}_s$ as:

$$\mathbf{C}_s \rightarrow \mathbf{A}_0 \mathbf{g}_s, \tag{10}$$

which yields the *tri-factor* reduced-rank regression:

$$\mathbf{x}_i^s = \mathbf{A}_s \mathbf{g}_s^\top \mathbf{A}_0^\top \mathbf{x}_i^0. \tag{11}$$

This results in an important consequence in that one can fix $\mathbf{A}_s$ and $\mathbf{A}_0$, which causes otherwise a large number of parameters to be estimated. In fact, as there could always be a rotation matrix between $\mathbf{A}_s$ and $\mathbf{g}_s$, and another between $\mathbf{g}_s$ and $\mathbf{A}_0$ that can make the final mapping invariant, we let $\mathbf{A}_s$ and $\mathbf{A}_0$ be rather fixed. Our ROS forces that these matrices come from the state-wise PCA learning, which is quite intuitive as well. All we need to estimate are $\mathbf{g}_s$, hence the number of parameters becomes $K \times r^2$.

3.4 Direct Subspace Alignment (DSA)

In the previous sections we have considered different ways of learning the mappings between the reference image and the subspaces or other images. In this section we propose another new method called *Direct Subspace Alignment* (DSA). The main idea is to directly maximize the alignment score between the predicted subject subspace and the ground-truth subspace. To formulate the optimization problem, we need to parameterize the predicted subject subspace. DSA uses the tri-factor RRR ROS' parametrization, that is, $\mathbf{x}_i^s = \mathbf{A}_s \mathbf{g}_s^\top \mathbf{A}_0^\top \mathbf{x}_i^0$. DSA builds a parameterized subspace with the matrix $\mathbf{Q}_i$. $\mathbf{Q}_i$ is composed of the reference sample $\mathbf{x}_i^0$ and the K-state samples that are generated by the tri-factor RRR. That is:

$$\mathbf{Q}_i = [\mathbf{x}_i^0, \quad \mathbf{A}_1 \mathbf{g}_1^\top \mathbf{A}_0 \mathbf{x}_i^0, \quad \ldots, \quad \mathbf{A}_K \mathbf{g}_K^\top \mathbf{A}_0 \mathbf{x}_i^0]. \tag{12}$$

Note that $\mathbf{Q}_i \in \Re^{p\times(K+1)}$ is a function of the parameters $\{\mathbf{g}_s\}_{s=1}^K$ which will be estimated DSA optimizes $\{\mathbf{g}_s\}_{s=1}^K$ such that the eigenvector of $\mathbf{Q}_i\mathbf{Q}_i^\top$ are maximally correlated with the eigenvectors of the ground truth bases $\mathbf{B}_i = [\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_q^{(i)}]$. DSA optimizes:

$$\max_{\mathbf{g}_1,\cdots,\mathbf{g}_K} E(\mathbf{g}_1, \cdots, \mathbf{g}_K) = \sum_{i=1}^{n}\sum_{j=1}^{q}\left\{{\mathbf{b}_j^{(i)}}^\top \mathbf{eig}_j(\mathbf{Q}_i\mathbf{Q}_i^\top)\right\}^2. \tag{13}$$

We assume that the basis vectors are always sorted in decreasing eigenvalue order. The operator $\mathbf{eig}_j(\mathcal{A})$ returns $\mathcal{A}$'s eigenvector corresponding to the j^{th} largest eigenvalue. Thus $\mathbf{eig}_j(\mathbf{Q}_i\mathbf{Q}_i^\top)$ indicates the j^{th} eigenvector of $\mathbf{Q}_i\mathbf{Q}_i^\top$.

Notice that (13) is the sum of the squared cosine angles between the basis vectors of the target subject subspace and the parameterized subject subspace. Hence (13) explicitly maximizes the alignment between the principal directions on the training set and the eigenvectors of $\mathbf{Q}_i\mathbf{Q}_i^\top$ that are parameterized by low-dimensional matrices $\mathbf{g}_1, ..., \mathbf{g}_K$. Optimizing (13) with respect to $\mathbf{g}_1, ..., \mathbf{g}_K$ is a non-convex optimization problem. In this paper, we used a numerical gradient optimization using Matlab's fminunc() function.

4 Related Work

Although to the best of our knowledge there exists little prior work that tackles the subspace regression problem framed as ours, in the face recognition literature there has been a similar line of research that aims at predicting images at unseen poses from a frontal-pose image. In this section we will briefly review some recent work closely related to ours, and contrast them with our approaches.

Recently in [11,12], a learning-based face synthesis approach was proposed, which aimed to discover the correspondences between facial features of frontal and non-frontal poses via regression estimation. From the learned regressors, they can synthesize features of non-frontal poses. For instance, the authors in [12] considered the correspondence between facial landmark points (e.g., AAM shapes). Non-frontal images can be synthesized by warping the texture to a canonical shape by mapping triangles. Suppressing the difference in image features used (landmark points vs. PCA subspace), this approach is similar in spirit to IMI or ROS algorithms. However, a crucial advantage of our proposed ROS is that we can explicitly capture the intrinsic low-dimensional relationship between different poses, potentially leading to more robust solutions. Moreover, ROS has less number of parameters which makes it less prone to overfitting.

The relationship between frontal and non-frontal images has been often encoded using generative probabilistic models. In [13], a joint Gaussian probabilistic model was formed on the pair of images at different views (e.g., frontal and left-45-deg). Given a new test image, they treated the new pose as missing data and marginalized the Gaussian distribution to reconstruct the unseen pose. Alternatively, the authors in [14] proposed a generative model that creates a one-to-many mapping from an idealized identity space to the observed

data space, where the identity space is the latent space in the factor analysis framework representing each individual invariant with poses. These approaches are intuitively appealing, not contingent on 3D modeling, and computationally efficient. However, compared to our subspace regression methods, their approach is based on the *generative* modeling and learning of pairs of poses, which can often be outperformed by our *discriminative* regression-based approaches.

Some other approaches are based on the 3D representation of faces. In [15] they estimated the 3D shape of faces from the non-frontal input images, and generate frontal views of the reconstructed faces using 3D computer graphics. Despite their good performance in face recognition tasks, it is unclear how it can work in our scenario of having low quality and low resolution video.

5 Experimental Results

We conducted experiments on predicting a person-specific facial pose subspace from a frontal image, predicting a subspace for illumination from only one illumination, and applications to subspace-based face recognition and face tracking.

5.1 Predicting a Subspace for Pose

We consider the CMU PIE data set [16] composed of about 41,368 images of 68 people. The face of each person is taken under 13 different poses, 43 different illumination conditions, and 4 different expressions. The images are labeled with these states. We used a subset of these images by taking 720 images of 60 subjects with 12 different poses with the same expression and illumination conditions. Each face image is cropped into a tight bounding box using the ground-truth facial landmark points which are also provided by the data set. The images of all poses for some five subjects are shown in Fig. 3. All the images are under the same expression and illumination. We normalized the images into a same size (48×48). For each subject, we estimate a PCA subspace for a fixed dimension. Then we split the data randomly into 50/10 training/testing subjects. By revealing only a single frontal image for each subject in the test fold, we predict the subspaces of the test subjects.

To measure the goodness of subspace alignment (between the estimated subspace and the ground-truth subspace obtained from real samples), we use three quantitative error metrics: (i) the smallest principal angle [10] that can also be computed by the `subspace()` function in Matlab, (ii) the sum of the squared cosine angles between basis vectors, and (iii) the subspace distance defined in [17]. In the squared cosine angle metric, for two subspaces $\mathbf{B}_1 = [\mathbf{b}_1^{(1)}, \ldots, \mathbf{b}_q^{(1)}]$ and $\mathbf{B}_2 = [\mathbf{b}_1^{(2)}, \ldots, \mathbf{b}_q^{(2)}]$, we do the SVD decomposition: $\mathbf{B}_1 = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^\top$ and $\mathbf{B}_2 = \mathbf{U}_2 \boldsymbol{\Sigma}_2 \mathbf{V}_2^\top$. Then the sum of the squared cosine angle errors between two subspaces can be defined as (in fact, we take an average, and subtract it from 1):

$$d_1(\mathbf{B}_1, \mathbf{B}_2) = 1 - \frac{1}{q} \sum_{j=1}^{q} \left({\mathbf{u}_j^{(1)}}^\top \mathbf{u}_j^{(2)} \right)^2, \tag{14}$$

Fig. 3. Different pose images for the first five subjects in the CMU PIE data set [16]. The images are arranged in a way that each row corresponds to a particular subject while each column represents a specific state (pose in this case).

where $\mathbf{u}_j^{(m)}$ is the j^{th} column of $\mathbf{U}_m$ for $m = 1, 2$. The recent subspace distance of [17]is defined as follows:

$$d_2(\mathbf{B}_1, \mathbf{B}_2) = \sqrt{\frac{1}{2}\left|tr\left(\mathbf{B}_1\mathbf{B}_1^\top - \mathbf{B}_2\mathbf{B}_2^\top\right)\right|}. \tag{15}$$

Note that for all these three measures, the smaller numbers indicate better performance. We performed our four subspace regression approaches. The test errors are shown in Table 1 (Left). In this experiment, we set the subject subspace dimension $q = 4$, the state (pose) subspace dimension $r = 5$. As shown in the table, the more sophisticated ROS and DSA approaches outperform the fairly straightforward IMI and DMS approaches most of the time.

Table 1. Subspace prediction errors for pose (Left) and illumination (Right). The best ones are boldfaced. PA=principal angle, d_1=cosine angle, d_2=subspace distance.

Pose	IMI	DMS	ROS	DSA
PA	0.4740	0.9940	**0.4542**	**0.4542**
d_1	0.2514	0.5854	**0.1284**	0.2582
d_2	0.5902	0.8784	**0.5635**	**0.5635**

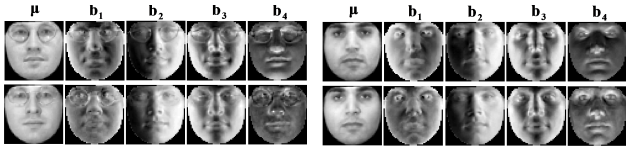
Illum.	IMI	DMS	ROS	DSA
PA	0.2172	0.4484	**0.1808**	0.2021
d_1	0.2090	0.3192	0.2068	**0.1741**
d_2	0.2434	0.7101	**0.2025**	0.2272

5.2 Predicting a Subspace for Illumination

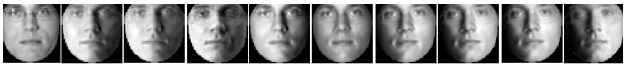
We next illustrate the capability of subspace regression to learn a person-specific illumination subspace. We used the frontal faces of 60 subjects under 19 different illumination conditions from the CMU PIE data set [16]. The 19 illumination states are selected so that they are roughly uniformly spread along the horizon. The reference state $s = 0$ corresponds to frontal lighting (e.g., the leftmost image in Fig. 4(c)). The subject subspace dimension is set to $q = 4$ (whereas the

ideal Lambertian surfaces should be three). From the 60 subjects, 50 subjects are selected for training, and the rest as testing. For ROS, the state subspace dimension is chosen as $r = 10$. All images are manually labeled with 66 landmarks, and warped to a canonical shape representation using triangulation.

Table 1 (Right) shows the test errors indicating the goodness of alignment with respect to the ground-truth subspace in terms of three error measures described in Sec. 5.1. Again ROS and DSA continue to outperform the straightforward regression approaches IMI and DMS. Fig. 4 depicts the learned subspace by ROS for some test subjects as well as the images generated from the predicted subspace. As shown, the mean and basis of the predicted subspace appear to be very similar to those of the ground-truth subspace. Also, the synthesized images sampled from the learned subspace look quite realistic.



(a) Subspace for subject A (b) Subspace for subject B



(c) Synthesized images from the ROS-learned subspace

Fig. 4. (a), (b) Subject subspaces for illumination variation. For two subjects (A and B), where (top) = the ground-truth subspaces, (bottom) = the predicted subject subspace by ROS. **(c)** Synthesized images from the ROS-learned subspace.

5.3 Face Tracking Across Pose

In this section we show how effectively the person-specific subspaces built from one sample can be applied to the problem of face tracking across pose. The face tracking scenario is within a vehicle with a camera mounted at a fixed position. We recorded videos that lasted for about 2 minutes yielding about 3000-frame long videos at 25 fps rates. Tracking in these videos is quite challenging because there are large variations in illumination and pose changes, since it is taken outdoors. One of the major challenges of this problem is that the driver's facial pose varies abruptly and dramatically. Moreover, the videos are noisy with low resolution and contrast. See Fig. 5 for few frames of the video.

In this context, we used subspace-based trackers [4,18] that can effectively deal with noise and large variation in appearance. In these approaches, the appearance model of the tracker is a subspace that can capture the variability of the target appearance, which is combined with efficient searching strategies, typically the sampling-based particle filtering [19]. To handle the appearance change

during tracking, the recent Incremental Visual Tracker (IVT) [18] updates the subspace model using the previously tracked images. However, as the training images are collected from the decision made by the current tracker, IVT's subspace learning can be seen as *self-training* or *unsupervised learning*. Although it is a reasonable approach to pursue given restricted information of a single image at the first frame, IVT is potentially unable to judge whether the current decision is correct or not in a principled manner, which is indeed the main cause of tracking drift.

On the other hand, the way we learn a person-specific subspace (i.e., the subspace regression) is highly advantageous as we directly predict a subspace from a single image given in the first frame of the video. Not only addressing the above-mentioned issues of the IVT, our approaches also avoid the computational overhead of updating the subspace at every frame since the subspace model is determined and fixed at the first frame. We design a tracker system that incorporates the subspace estimated by our subspace regression approaches into the particle filtering framework. The training data used for the subspace regression are obtained from the CMU PIE data set [16], where we use 50 subjects with 12 different poses (other conditions such as illumination and expression are not considered here).

In addition to comparing our approaches with IVT, we also present the performance of a fairly basic template matching tracker as a baseline comparison. We maintain a template image model for the face target, while similar to IVT, the template model is updated at every frame by computing a weighted average of current template and the tracked images, hence named as Adaptive Template Matching (ATM). The ATM is essentially identical to the IVT except that it maintains only the mean of the subspace.

We enforce the same settings for all competing methods for fair comparison. The initial location of a face is obtained from a face detector. For the tracking states, we used the axis-aligned bounding box representation, meaning that we keep track of three parameters: center position and scale. To provide quantitative tracking results, we manually labeled the face location for every 10th frame to form the ground-truth. The average root-mean-square (RMS) errors (in pixels) are shown in Table 2 (Left). Our IMI, DSA and ROS approaches achieved slightly better performance than the IVT even though we only take into account the pose variation in the subspace learning. Fig. 5 also depicts some selected frames that compare our approaches with the IVT.

5.4 Face Recognition from Arbitrary Pose

This section describes experiments on using subspace regression for face recognition across pose, and we compare it with previous work [13]. We used the CMU PIE data set [16], where 50 subjects are selected to train our methods for subspace regression, while we use the rest of 10 subjects for testing. The frontal-pose images for these 10 subjects serve as the gallery images, and the probe set is comprised of their images in 4 different poses: left/right 45/90-degree views.

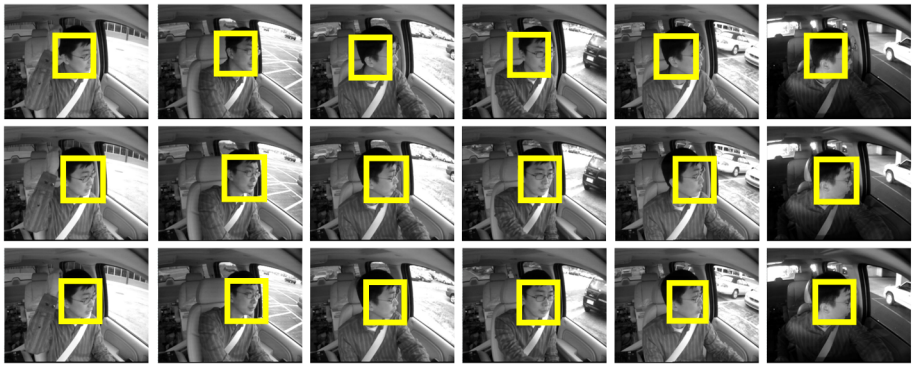


Fig. 5. Selected frames illustrating tracking performance of IVT (top row), DMS (middle), and ROS (bottom). Yellow bounding box represents the tracked face.

Table 2. Results of (a) face tracking and (b) face recognition from arbitrary poses.

(a) Tracking RMS Errors

ATM	42.99
IVT [18]	38.41
IMI	38.16
DMS	40.04
ROS	**37.98**
DSA	**37.95**

(b) Face Recognition Errors

Test Pose	L-45	R-45	L-90	R-90	Average
[13]'s Approach	0.20	0.40	0.40	0.30	0.325
IMI	0.20	0.30	0.40	0.50	0.350
DMS	0.30	0.40	0.30	0.50	0.375
ROS	0.10	0.20	0.30	0.30	**0.225**
DSA	0.20	0.20	0.30	0.30	0.250

We first apply the subspace regression approaches to each frontal image in the gallery set to predict the person-specific pose subspace. Given a new test image (i.e. non-frontal image) in the probe set, we project it onto each of the learned subspaces in the gallery images and compute the distance to the subspaces. This avoids the pose prediction step for test images. The test recognition errors for four different test poses are shown in Table 2 (Right).

The similar problem of face recognition from arbitrary poses has been studied previously (See the brief summary in the related work in Sec. 4). In the table, we also compared our methods with the recent approach of [13] that builds a probabilistic model for a pair of images at different poses, and predicts the latent test images based on the probability maximization. As shown in the result, the ROS and DSA approaches outperform [13]'s probabilistic method, which substantiates that our *discriminative* regression algorithm built on the low-dimensional intrinsic subspaces can be superior to the generative modeling approaches.

6 Concluding Remarks

This paper has addressed the novel problem of predicting a subspace from one sample, and we have illustrated its benefits in predicting person-specific pose and illumination subspaces. We have also successfully applied it to the problem of face recognition and tracking. By casting the problem as the subspace

regression, we have proposed four approaches, where we have observed that our approaches, especially the ROS and the DSA, are promising and perform well with a significantly reduced number of parameters. Interestingly, DSA is the optimal error function to optimize because it directly regresses on the subspace. However, it did not always achieve the best performance. Some drawbacks are that the optimization procedure is complex and prone to be trapped into local minima. Developing efficient optimization methods is left as future work.

References

[1] Sirovich, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. J. Opt. Soc. Am. A **4** (1987) 519–524
[2] Jolliffe, I.T.: Principal Component Analysis. New York: Springer-Verlag (1986)
[3] Turk, M., Pentland, A.: Eigenfaces for recognition. Journal Cognitive Neuroscience **3** (1991) 71–86
[4] Black, M.J., Jepson, A.D.: Eigentracking: Robust matching and tracking of objects using view-based representation. IJCV **26** (1998) 63–84
[5] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: ECCV. (1998)
[6] Bischof, H., Wildenauer, H., Leonardis, A.: Illumination insensitive recognition using eigenspaces. Computer Vision and Image Understanding **1** (2004) 86 – 104
[7] Anderson, T.W.: An Introduction to Multivariate Statistical Analysis. 2nd ed. Wiley, New York (1984)
[8] de la Torre, F., Black, M.J.: Dynamic coupled component analysis. In: CVPR. (2001)
[9] Stoica, P., Viberg, M.: Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. IEEE Trans. on Sig. Proc. **44** (1996) 3069–3078
[10] Hotelling, H.: Relations between two sets of variates. Biometrika **28** (1936) 321–377
[11] Sanderson, C., Bengio, S., Yongsheng, G.: On transforming statistical models for non-frontal face verification. Pattern recognition **39** (2006) 288–302
[12] Asthana, A., Sanderson, C., Gedeon, T., Goecke, R.: Learning-based face synthesis for pose-robust recognition from single image (2009) BMVC.
[13] Ni, J., Schneiderman, H.: Face view synthesis across large angles (2005) International Workshop on Analysis and Modeling of Faces and Gestures.
[14] Prince, S., Warrell, J., Elder, J., Felisberti, F.: Tied factor analysis for face recognition across large pose differences. IEEE Trans. on PAMI **30** (2008) 970–984
[15] Blanz, V., Grother, P., Phillips, P., Vetter, T.: Face recognition based on frontal views generated from non-frontal images (2005) CVPR.
[16] Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. IEEE Trans. on PAMI **25** (2003) 1615–1618
[17] Wang, L., Wang, X., Feng, J.: Intrapersonal subspace analysis with application to adaptive Bayesian face recognition. Pattern Recognition **38** (2005) 617–621
[18] Ross, D., Lim, J., Lin, R.S., Yang, M.H.: Incremental learning for robust visual tracking. IJCV **77** (2007) 125–141
[19] Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density (1996) ECCV.