MPCA: EM-BASED PCA FOR MIXED-SIZE IMAGE DATASETS

Feiyu Shi, Menghua Zhai, Drew Duncan, Nathan Jacobs

University of Kentucky
{fsh224, mzh234, drew, jacobs}@cs.uky.edu

ABSTRACT

Principal component analysis (PCA) is a widely used technique for dimensionality reduction which assumes that the input data can be represented as a collection of fixed-length vectors. Many real-world datasets, such as those constructed from Internet photo collections, do not satisfy this assumption. A natural approach to addressing this problem is to first coerce all input data to a fixed size, and then use standard PCA techniques. This approach is problematic because it either introduces artifacts when we must upsample an image, or loses information when we must downsample an image. We propose MPCA, an approach for estimating the PCA decomposition from multi-sized input data which avoids this initial resizing step. We demonstrate the effectiveness of this approach on simulated and real-world datasets.

***Index Terms*—** dimensionality reduction, expectation-maximization, nonlinear optimization

1. INTRODUCTION

Dimensionality reduction methods, such as principal component analysis (PCA) [1], are commonly used as a preprocessing step in image analysis applications [2, 3, 4]. When applied to an image dataset, such techniques typically assume that the images are well aligned and the same size (same number of pixels and aspect ratio). This assumption is violated, for example, when the object of interest is moving closer to the camera or when images are captured at different resolutions.

To handle image motion, several methods have been proposed that simultaneously align the images and perform dimensionality reduction. Jepson et al. [5] propose an online approach for estimating a low-rank appearance model for object tracking. De la Torre and Black [6] focus on face image datasets and solve this problem using nonlinear optimization. Peng et al. [7] propose a method for aligning linearly correlated images, with sparse errors, using a series of convex optimization problems. A basic assumption of these methods is that the motion is unknown but small, and that the images are of a similar size.

When image sizes are very different, it is important to account for how the images were created. We take inspiration from work on image super-resolution, and use a model of how the small-size images were generated from hypothetical full-size images. Subspace-based methods have been used for the problem of image super-resolution. Given a small-size image, Capel and Zisserman [8] proposed to reconstruct a full-size image using a PCA basis estimated from a training set of full-size images. Unlike this work, we do not assume an independent set of full-size images to learn a subspace. Vandewalle et al. [9] solve the super-resolution problem using a subspace method, but their focus is on solving the image alignment problem. We assume the transformations are known in advance and directly estimate the PCA decomposition from a dataset of images with diverse sizes.

We introduce a method for estimating a PCA decomposition of mixed-size data that extends Roweis's expectation-maximization (EM) algorithm [10]. Our method, MPCA, minimizes reconstruction error using nonlinear optimization in place of the linear operations used in the previous work. The key contributions of this work are: formulating the reconstruction error objective function for multi-size PCA, deriving the gradients of this objective function and integrating it into the EM-based PCA algorithm [10], and comparing this method to a commonly used naïve approach on synthetic and real datasets.

2. PCA FOR FIXED-SIZE DATA

We assume that input data elements can be represented by vectors. For instance, an $m \times n$ image with three color channels can be represented by a $d \times 1$ vector, where $d = m \times n \times 3$. Throughout the paper, we will focus on image data and assume that all images have been "flattened" using an equivalent method. Let $\mathbf{X} = \{X_i | X_i \in \mathbb{R}^{d_i \times 1}, i = 1 \ldots N\}$ denote the input dataset. If each data vector is the same size, where $d_i = d$, the dataset, $\mathbf{X}$, can be represented as a $d \times N$ matrix. The outcome of PCA [1] is a p-dimensional basis, $\mathbf{\Phi} \in \mathbb{R}^{d \times p}$, and a corresponding set of coefficients, $\mathbf{h} \in \mathbb{R}^{p \times N}$, that can be used to reconstruct the input dataset, X, with minimum error in the least-squared sense.

There are many methods for estimating a PCA decomposition of fixed-size data. The two most commonly used approaches involve taking the eigenvalue decomposition of the data covariance matrix or computing the singular value decomposition (SVD) of the mean-centered data matrix. Our

work extends an alternative EM-based algorithm [10], which we describe briefly for completeness. Given a mean-centered data matrix $\tilde{\mathbf{X}} = \{\tilde{X}_i | X_i - \bar{X}\}$ where $\bar{X} = \frac{1}{N}\sum X_i$, the algorithm alternates between the following steps:

$$\text{E-step: } \mathbf{h}^{new} = (\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\mathsf{T}\tilde{\mathbf{X}} \quad (1)$$

$$\text{M-step: } \mathbf{\Phi}^{new} = \tilde{\mathbf{X}}\mathbf{h}^\mathsf{T}(\mathbf{h}\mathbf{h}^\mathsf{T})^{-1}. \quad (2)$$

The resulting $\mathbf{\Phi}$ spans the principal component subspace, but is not guaranteed to be an orthogonal basis. To obtain the principal components, the following orthonormalization process is used:

$$\mathbf{\Phi}_{orth} = orth(\mathbf{\Phi}) \quad (3) \qquad [\mathbf{\Phi}', \mathbf{A}] = eig(\mathbf{h}\mathbf{h}^\mathsf{T}) \quad (5)$$

$$\mathbf{h} = \mathbf{\Phi}_{orth}^\mathsf{T}\tilde{\mathbf{X}} \quad (4) \qquad \mathbf{\Phi} = \mathbf{\Phi}_{orth}\mathbf{\Phi}' \quad (6)$$

where $orth(.)$ gives a orthonormal basis for the range of $\mathbf{\Phi}$ and $eig(.)$ computes the eigenvalue decomposition.

3. PCA FOR MULTI-SIZE DATA

When data vectors, X_i, are different sizes the data cannot be directly represented as a matrix, and none of the methods for computing PCA described in the previous section work as formulated. Such non-uniformity is common in datasets collected over large periods of time, such as from outdoor webcams, or from multiple individuals, such as Internet photo collections. There are many possible causes, including upgrading to a new sensor with higher resolution and inherent differences in scale, for example the image patches of objects at different distances from a camera. We begin our technical discussion by defining a generative model for multi-size image data.

3.1. Generative Model for Multi-Size Data

In image super-resolution, small-size images are presumed to be generated from a full-size image using a transformation model [8, 11, 12]. A typical transformation model is described as follows:

$$X_i = D_i B_i W_i Y + n_i, \text{ for } i = 1, \ldots, N, \quad (7)$$

where Y is a single full-size image, X_i is one of many small-size images, n_i is a noise term and D_i, B_i, W_i, are decimation, blurring and warping matrices, respectively.

In our problem, we assume that the transformation models are known and that there exist underlying full-size images. This gives us the following model:

$$X_i = S_i Y_i + n_i, \text{ for } i = 1, \ldots, N, \quad (8)$$

where Y_i denotes a full-size image and S_i is the known transformation model. The transformation matrices, $\{S_i\}$, are $m_i \times n$ matrices with $m_i < n$. Although a linear transformation operator, S_i, seems quite simple, it can model a wide variety of image transformations, including planar image transformations [13].

Our goal is to estimate a p-dimensional PCA basis, $\mathbf{\Phi}$, and a corresponding set of coefficients, $\mathbf{h}$, of the full-sized dataset, $\mathbf{Y} = \{Y_i\}$. We first describe the baseline approach, and then propose our EM-based multi-size PCA method.

3.2. Upsampling-Based PCA (UP-PCA)

As a baseline approach, we propose to independently transform each input image, X_i, into a corresponding full-size image, Y_i^{up}, and then use a standard techniques to estimate a PCA decomposition (UP-PCA). Let $S_i^\dagger$ be the inverse transformation operator that estimates a full-size image $Y_i^{up} = S_i^\dagger X_i$ from a small size image. For example, when the data transformation is a small scaling operation, the transformation, $S_i^\dagger$, can be bilinear interpolation (we use this for all experiments). Given the approximate full-size data matrix, Y^{up}, we use one of methods introduced in Sec. 2 to estimate a PCA decomposition, $\mathbf{\Phi}^{up}, \mathbf{h}^{up}$. In the end, UP-PCA minimizes the following reconstruction error:

$$E^{up}(\mathbf{\Phi}, \{h_i\}) = \frac{1}{dN}\sum_{i=1}^{N}\left\|(\bar{X} + \mathbf{\Phi}h_i) - S_i^\dagger X_i\right\|_2^2. \quad (9)$$

This approach performs well when the inverse of the transformation operations, $\{S_i^\dagger\}$, are easy to estimate, such as when image scale changes are relatively small. However, it fails dramatically when it is impossible to accurately estimate the full-size image, such as when scale changes are large.

4. EM-BASED PCA FOR MULTI-SIZE DATA

We propose an EM-based PCA approach (MPCA) that minimizes errors due to directly estimating the full-size image from a small-size image.

4.1. Multi-Size Average Image Estimation

The first step of MPCA is to estimate the average image. In UP-PCA, we first performed single-image super-resolution and averaged the resulting images. For MPCA, we estimate the average image directly from the small-sized images using a multi-image super-resolution technique. We obtain the average image, $\bar{X}^{ma}$, by minimizing the following objective function:

$$E^{ma}(\bar{X}) = \frac{1}{dN}\sum_{i=1}^{N}\left\|S_i\bar{X} - X_i\right\|_2^2. \quad (10)$$

We minimize this E^{ma} using standard nonlinear optimization methods together with the following analytical gradient:

$$\frac{\partial E^{ma}}{\partial \bar{X}} = \frac{2}{dN}\sum_{i=1}^{N} S_i^\mathsf{T}(S_i\bar{X} - X_i). \quad (11)$$

Algorithm 1: MPCA

Input: Multi-size dataset, $\mathbf{X}$

Output: PCA basis $\mathbf{\Phi}^{mp}$ and coefficients $\mathbf{h}^{mp}$

1 Center data: Compute $\bar{X}^{ma}$ and $\tilde{\mathbf{X}}$;

2 Initialization: $\mathbf{\Phi} \leftarrow \mathbf{\Phi}^{up}$, $\mathbf{h} \leftarrow \mathbf{h}^{up}$;

3 **while** $\Delta E^{mp} > \epsilon E^{mp}$ **do**

4 $\quad h_i \leftarrow \arg\min_{h_i} E^{mp}(\mathbf{\Phi}, h_i)$;

5 $\quad \mathbf{\Phi} \leftarrow \arg\min_{\mathbf{\Phi}} E^{mp}(\mathbf{\Phi}, h_i)$;

6 **end**

7 $\mathbf{\Phi} \leftarrow orth(\mathbf{\Phi})$;

8 $h_i \leftarrow \arg\min_{h_i} E^{mp}(\mathbf{\Phi}, h_i)$;

9 $\mathbf{\Phi}' \leftarrow eig(\mathbf{h}\mathbf{h}^{\mathsf{T}})$;

10 $\mathbf{\Phi}^{mp} \leftarrow \mathbf{\Phi}\mathbf{\Phi}'$, $h_i^{mp} \leftarrow \arg\min_{h_i} E^{mp}(\mathbf{\Phi}^{mp}, h_i)$;

4.2. Multi-Size EM Algorithm (MPCA)

We now return to our main goal, estimating a PCA decomposition of the original data. We begin by defining the reconstruction error for the multi-size case by incorporating the known image transformations, S_i, into the definition of reconstruction error:

$$E^{mp}(\mathbf{\Phi}, \{h_i\}) = \frac{1}{dN} \sum_{i=1}^{N} \left\| S_i(\bar{X}^{ma} + \mathbf{\Phi} h_i) - X_i \right\|_2^2 . \tag{12}$$

The main benefit of this reformulation over the upsampling method is that we no longer have to invert the transformation matrix S_i. Instead, we solve for the full-sized basis that minimizes the reconstruction error when converted into the size of the given data.

Inspired by the EM algorithm for PCA [10], we use nonlinear optimization (L-BFGS [14]) to alternately minimize E^{mp}:

$$h_i = \arg\min_{h_i} E^{mp}(\mathbf{\Phi}, h_i) \tag{13}$$

$$\mathbf{\Phi} = \arg\min_{\mathbf{\Phi}} E^{mp}(\mathbf{\Phi}, h_i), \tag{14}$$

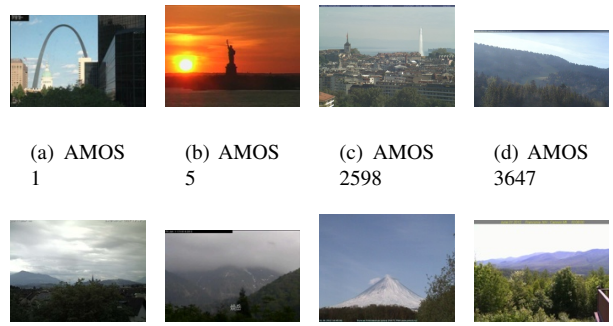
by using its partial derivatives:

$$\frac{\partial E^{mp}}{\partial h_i} = \frac{2}{dN}[\mathbf{\Phi}^{\mathsf{T}} S_i^{\mathsf{T}} S_i(\bar{X}^{ma} + \mathbf{\Phi} h_i) - \mathbf{\Phi}^{\mathsf{T}} S_i^{\mathsf{T}} X_i] \tag{15}$$

$$\frac{\partial E^{mp}}{\partial \Phi} = \frac{2}{dN} \sum_{i=1}^{N} S_i^{\mathsf{T}} S_i(\bar{X}^{ma} + \mathbf{\Phi} h_i) h_i^{\mathsf{T}} - S_i^{\mathsf{T}} X_i h_i^{\mathsf{T}} \tag{16}$$

As with the fixed-size method (Sec 2), we obtain the final PCA subspace $\mathbf{\Phi}^{mp}$, and its coefficients $\mathbf{h}^{mp}$ by othonormalizing $\mathbf{\Phi}$ using the similar method. The key difference is that we use (13) instead of (4) to obtain the final coefficients, $\mathbf{h}^{mp}$. See Alg. 1 for a complete algorithm description.

The alternating minimization typically converges quickly (in 10 to 20 iterations) with reasonable values for optimality conditions. However runtime depends on the eigenspectrum of the input data, an inherent problem for iterative approaches to estimating PCA decompositions [10]. If the hypothetical full-size data matrix, $\mathbf{Y}$, has several close singular values, then convergence can require many iterations. In such cases, it may be beneficial to combine algorithm steps 3–6 in a single nonlinear optimization to make assessing convergence easier.



(a) AMOS 1 (b) AMOS 5 (c) AMOS 2598 (d) AMOS 3647

(e) AMOS 4854 (f) AMOS 9746 (g) AMOS 9795 (h) AMOS 10050

Fig. 1. Example images from AMOS.

5. EVALUATION

We evaluated a MATLAB implementation of the proposed methods with real and synthetic datasets[1].

5.1. Metrics

We use two error metrics to evaluate the quantitative performance of our method. For all datasets, we evaluate reconstruction accuracy for a given basis dimensionality, p, using the peak signal-to-noise ratio (PSNR):

$$\text{PSNR} = 10 \log_{10}(1/E). \tag{17}$$

When working with artificially resized data, we also directly compare the ground-truth PCA basis, $\mathbf{\Phi}$, with the estimated PCA basis, $\hat{\mathbf{\Phi}}$:

$$e_{pc} = \left\| \mathbf{\Phi}^{\mathsf{T}} \hat{\mathbf{\Phi}} - \mathbf{I} \right\|_F . \tag{18}$$

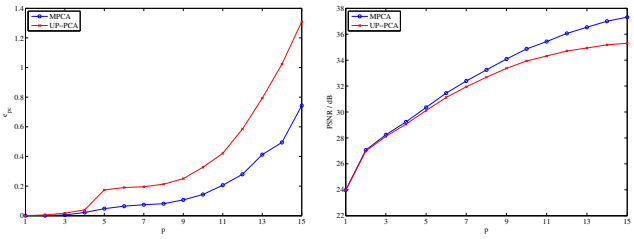
5.2. Experiments on Artificially Resized Image Data

We use a dataset of 8 outdoor scenes, each with 50 fixed-size images (Fig. 1), from the AMOS dataset [3] as a starting point. The image size for the scene with the smallest (largest) image is 153×103 (256×192). From these, we generate multiple synthetic multi-sized datasets. For a given scene, we generate each synthetic dataset by resizing each image by a random scaling factor drawn uniformly from $[r, 1]$. We do

[1] http://cs.uky.edu/~jacobs/mpca

(a) PCA (b) UP-PCA error (c) MPCA error

Fig. 2. A comparison of a the third principal component estimated by various methods on the AMOS-1 dataset ($r = .77$). (a) A ground-truth principal component. (b,c) A false-color image that represents the squared error between (a) and the third component estimated by UP-PCA and MPCA, respectively (dark blue pixels correspond to low error).



(a) e_{pc} (b) PSNR

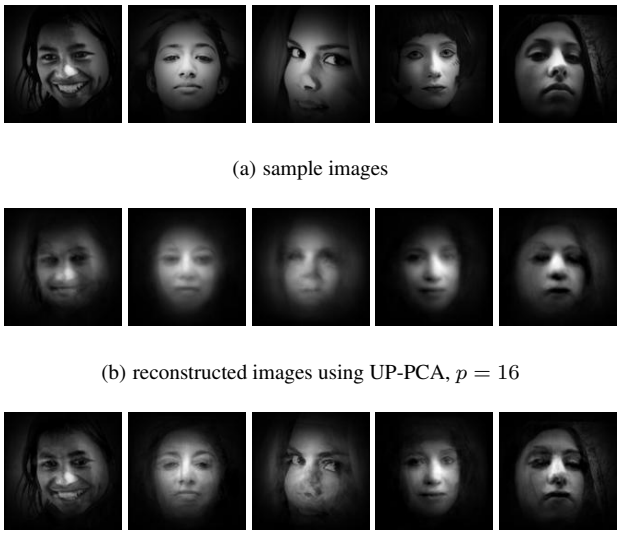
Fig. 3. Results of quantitative evaluation on the AMOS-1 dataset. MPCA gives more accurate components (a) and reconstructions (b) than UP-PCA.

this for 8 different values of r, and repeat this process for 30 iterations. We generate a total of 8×30 different synthetic datasets for each scene. For each PCA estimation method, we compute the first p principal components.

We visualize the errors of the third principal component estimated by UP-PCA and MPCA relative to ground truth (PCA on the full-sized images) for one of our generated multi-sized datasets. The error in the estimate using UP-PCA (Fig. 2(b)) is much larger than the error for MPCA (Fig. 2(c)), especially near the edges. This shows that MPCA generates more accurate estimates of the true principal components than UP-PCA. Fig. 3 compares the performance of UP-PCA and MPCA on the AMOS-1 dataset using our proposed metrics (averaged over 8×30 trials). The results of the other seven datasets are similar. Fig. 3(a) and Fig. 3(b) show that as p increases, e_{pc} or PSNR increases, and MPCA performs much better. This improvement increases as p grows larger. The average reconstruction improvement of each AMOS dataset ranges from .06 to .77 dB in PSNR, with an average improvement of .34 dB.

5.3. Experiments on Real Multi-size Image Data

To construct a real multi-sized dataset, we downloaded many images containing faces from Flickr, detected fiducial points using a commercial software package[2], and extracted a small surrounding patch. Each face patch was converted to gray, aligned using a similarity transform estimated from eye locations, and point-wise multiplied by a Gaussian to reduce the effect of background pixels. The resulting set of 500 patches range in size from 110×110 to 256×256. Fig. 4 shows reconstructed images obtained using the UP-PCA approach and the MPCA approach for 16 basis components. Images reconstructed using the MPCA components preserve detail while those reconstructed using UP-PCA components are more blurry.



(a) sample images

(b) reconstructed images using UP-PCA, $p = 16$

(c) reconstructed images using MPCA, $p = 16$

Fig. 4. Selected results of reconstructed images of face dataset.

6. CONCLUSIONS

We proposed an approach for estimating a PCA basis for a set of images of different sizes. The key contribution is extending an existing EM-based approach with nonlinear optimization to handle a variety of image formation models. We show quantitatively and qualitatively that this method outperforms a baseline approach on real and synthetic datasets. An interesting area for future work is in simultaneously estimating the image transformations and the PCA basis.

Acknowledgements

We gratefully acknowledge Robert Pless for his assistance with the AMOS dataset and Mohammad T. Islam for help preparing the face dataset.

[2] http://www.omron.com

7. REFERENCES

[1] Ian Jolliffe, *Principal component analysis*, Wiley Online Library, 2005.

[2] Matthew Turk and Alex Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] Nathan Jacobs, Nathaniel Roman, and Robert Pless, "Consistent temporal variations in many outdoor scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[4] Li Shen and Ping Tan, "Photometric stereo and weather estimation using internet images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[5] Allan D Jepson, David J Fleet, and Thomas F El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1296–1311, 2003.

[6] Fernando De la Torre and Michael J Black, "Robust parameterized component analysis," in *European Conference on Computer Vision*, 2006.

[7] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[8] David Capel and Andrew Zisserman, "Super-resolution from multiple views using learnt image models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[9] Patrick Vandewalle, Luciano Sbaiz, Joos Vandewalle, and Martin Vetterli, "Super-resolution from unregistered and totally aliased signals using subspace methods," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3687–3703, 2007.

[10] Sam Roweis, "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems*, 1998.

[11] Michael Elad and Arie Feuer, "Super-resolution reconstruction of image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 817–834, 1999.

[12] Bahadir K Gunturk, Aziz Umit Batur, Yucel Altunbasak, Monson H Hayes, and Russell M Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 597–606, 2003.

[13] Steven M Seitz and Simon Baker, "Filter flow," in *IEEE International Conference on Computer Vision*, 2009.

[14] Dong C Liu and Jorge Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1-3, pp. 503–528, 1989.