Recognition from Appearance Subspaces Across Image Sets of Variable Scale

Ognjen Arandjelović
http://mi.eng.cam.ac.uk/~oa214

Trinity College
University of Cambridge
CB2 1TQ, UK

Abstract

Linear subspace representations of appearance variation are pervasive in computer vision. In this paper we address the problem of robustly matching them (computing the similarity between them) when they correspond to sets of images of different (possibly greatly so) scales. We show that the naïve solution of projecting the low-scale subspace into the high-scale image space is inadequate, especially at large scale discrepancies. A successful approach is proposed instead. It consists of (i) an interpolated projection of the low-scale subspace into the high-scale space, which is followed by (ii) a rotation of this initial estimate within the bounds of the imposed "downsampling constraint". The optimal rotation is found in the closed-form which best aligns the high-scale reconstruction of the low-scale subspace with the reference it is compared to. The proposed method is evaluated on the problem of matching sets of face appearances under varying illumination. In comparison to the naïve matching, our algorithm is shown to greatly increase the separation of between-class and within-class similarities, as well as produce far more meaningful modes of common appearance on which the match score is based.

1 Introduction

One of the most commonly encountered problems in computer vision is that of matching appearance. Whether it is images of local features [7], views of objects [6] or faces [14], textures [13] or rectified planar structures (buildings, paintings) [10], the task of comparing appearances is virtually unavoidable in a modern computer vision application. A particularly interesting and increasingly important instance of this task concerns the matching of *sets* of appearance images, each set containing examples of variation corresponding to a single class.

A ubiquitous representation of appearance variation within a class is by a linear subspace [4, 5]. The most basic argument for the linear subspace representation can be made by observing that in practice the appearance of interest is constrained to a small part of the image space. Domain-specific information may restrict this even further e.g. for Lambertian surfaces seen from a fixed viewpoint but under variable illumination [2, 3, 8] or smooth objects across changing pose [12, 15]. What is more, linear subspace models are also attractive for their low storage demands – they are inherently compact and can be learnt incrementally [9]. Indeed, throughout this paper we assume that the original data from which subspaces are estimated is not available.

A problem which arises when trying to match two subspaces – each representing certain appearance variation – and which has not as of yet received due consideration in the

literature, is that of matching subspaces embedded in different image spaces, that is, corresponding to image sets od different scales. This is a frequent occurrence: an object we wish to recognize may appear larger or smaller in an image depending on its distance, just as a face may, depending on the person's height and positioning relative to the camera. In most matching problems in the computer vision literature, this issue is overlooked. Here we address it in detail and show that a naïve approach to normalizing for scale in subspaces results in inadequate matching performance. Thus, we propose a method which *without any assumptions on the nature of appearance* that the subspaces represent, constructs an optimal hypothesis for a high-resolution reconstruction of the subspace corresponding to low-resolution data.

In the next section, we first briefly overview the linear subspace representation, followed by a description of the naïve normalization. The proposed solution is described in this section as well. In Section 3 the two approaches are empirically analyzed in detail. The main contribution and conclusions of the paper are summarized in Section 4.

2   Matching Subspaces across Scale

Consider a set $X \subset \mathbb{R}^d$ containing vectors which represent rasterized images:

$$X = \left\{ \mathbf{x}_1, \ldots, \mathbf{x}_N \right\} \tag{1}$$

where d is the number of pixels in each image. It is assumed that all of the images represented by members of X have the same aspect ratio, so that the same indices of different vectors correspond spatially to the same pixel location. A common representation of appearance variation described by X is by a linear subspace of dimension D, where usually it is the case that $D \ll d$. If $\mathbf{m}_X$ is the estimate of the mean of the samples in X:

$$\mathbf{m}_X = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i, \tag{2}$$

then $\mathbf{B}_X \in \mathbb{R}^{d \times D}$, a matrix with columns consisting of orthonormal basis vectors spanning the D-dimensional linear subspace embedded in a d-dimensional image space, can be computed from the corresponding covariance matrix $\mathbf{C}_X$:

$$\mathbf{C}_X = \frac{1}{N-1} \sum_{i=1}^{N} \left(\mathbf{x}_i - \mathbf{m}_X \right) \left(\mathbf{x}_i - \mathbf{m}_X \right)^T. \tag{3}$$

Specifically, an insightful interpretation of $\mathbf{B}_X$ is as the row and column space basis of the best rank-D approximation to $\mathbf{C}_X$:

$$\mathbf{B}_X = \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{d \times D} \\ \mathbf{B}^T \mathbf{B} = I}} \min_{\substack{\Lambda \in \mathbb{R}^{D \times D} \\ \Lambda_{ij} = 0, i \neq j}} \left\| \mathbf{C}_X - \mathbf{B} \, \Lambda \, \mathbf{B}^T \right\|_F^2, \tag{4}$$

where $\|.\|_F$ is the Frobenius norm of a matrix.

2.1   The "Naïve Solution".

Let $\mathbf{B}_X \in \mathbb{R}^{d_l \times D}$ and $\mathbf{B}_Y \in \mathbb{R}^{d_h \times D}$ be two basis vectors matrices corresponding to appearance variations of image sets containing images with d_l and d_h pixels respectively. Without loss

of generality, let also $d_l < d_h$. As before, we shall assume that all images both within each set, as well as across the two sets, are of the same aspect ratio. We wish to compute the similarity of sets represented by orthonormal basis matrices $\mathbf{B}_X$ and $\mathbf{B}_Y$.

Subspaces spanned by the columns of $\mathbf{B}_X$ and $\mathbf{B}_Y$ cannot be compared directly as they are embedded in different image spaces. Instead, let us model the process of an isotropic downsampling of a d_h-pixel image down to d_l pixels with a linear projection realized though a projection matrix $\mathbf{P} \in \mathbb{R}^{d_l \times d_h}$. In other words, for a low-resolution image set $X \subset \mathbb{R}^{d_l}$:

$$X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \tag{5}$$

there is a high-resolution set $X^* \subset \mathbb{R}^{d_h}$, such that:

$$X^* = \left\{ \mathbf{x}_i^* \mid \mathbf{x}_i = \mathbf{P}\,\mathbf{x}_i^* \; ; \; i = 1, \ldots, N \right\}. \tag{6}$$

The form of the projection matrix depends on (i) the projection model employed (e.g. bilinear, bicubic etc.) and (ii) the dimensions of high and low scale images; see Figure 1 for an illustration. Throughout this paper we used bilinear projection.
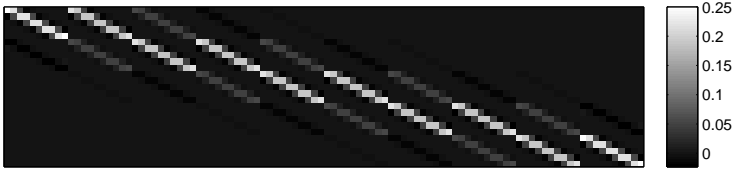

Figure 1: The projection matrix $\mathbf{P} \in \mathbb{R}^{25 \times 100}$ modelling the process of downsampling a 10×10 pixel image to 5×5 pixels, using a bilinear projection model, shown as an image. For the interpretation of image intensities see the associated grey level scale on the right.

Under the assumption of a linear projection model, the least-square error reconstruction of the high-dimensional data can be achieved with a linear projection as well, in this case by $\mathbf{P}_R$ which can be computed as:

$$\mathbf{P}_R = \mathbf{P}^T \left(\mathbf{P}\,\mathbf{P}^T\right)^{-1}. \tag{7}$$

Since we assume that the original data from which $\mathbf{B}_X$ was estimated is not available, an estimate of the subspace corresponding to X^* can be computed by re-projecting each of the basis vectors (columns) of $\mathbf{B}_X$ into $\mathbb{R}^{d_h}$:

$$\tilde{\mathbf{B}}_X^* = \mathbf{P}_R\,\mathbf{B}_X. \tag{8}$$

Note that in general $\tilde{\mathbf{B}}_X^*$ is not an orthonormal matrix i.e. $\tilde{\mathbf{B}}_X^{*T} \tilde{\mathbf{B}}_X^* \neq \mathbf{I}$. Thus, after re-projecting the subspace basis, we orthogonalize it using the Householder transformation [11], producing the high-dimensional subspace basis estimate $\mathbf{B}_X^*$ which can be directly compared with $\mathbf{B}_Y$.

Limitations of the Naïve Solution. The process of downsampling an image inherently causes a loss of information. In re-projecting the subspace basis vectors, information gaps

are "filled in" through interpolation. This has the effect of constraining the spectrum of variation in the high-dimensional reconstructions to the bandwidth of the low-dimensional data. Compared to the genuine high-resolution images, the reconstructions are void of high frequency detail which usually plays a crucial role in discriminative problems.

2.2 Proposed Solution

We seek a constrained correction to the subspace basis $\mathbf{B}_X^*$. To this end, consider a vector $\mathbf{x}_i^*$ in the high-dimensional image space $\mathbb{R}^{d_h}$, which when downsampled maps onto $\mathbf{x}_i$ in $\mathbb{R}^{d_l}$. As before, we model this as a linear projection effected by a projection matrix $\mathbf{P}$:

$$\mathbf{x}_i = \mathbf{P}\,\mathbf{x}_i^*. \tag{9}$$

Writing the reconstruction of $\mathbf{x}_i^*$, computed as described in the previous section, as $\mathbf{x}_i^* + \mathbf{c}_i$, it has to hold:

$$\mathbf{x}_i = \mathbf{P}\left(\mathbf{x}_i^* + \mathbf{c}_i\right), \tag{10}$$

or, equivalently:

$$\mathbf{0} = \mathbf{P}\,\mathbf{c}_i, \tag{11}$$

In other words, the correction term $\mathbf{c}_i$ has to lie in the nullspace of $\mathbf{P}$. Let $\mathbf{B}_c$ be a matrix of basis vectors spanning the nullspace which, given its meaning in the proposed framework, we shall refer to as the *ambiguity constraint subspace*. Then the actual appearance in the high-dimensional image space corresponding to the subspace $\mathbf{B}_X \in \mathbb{R}^{d_l \times D}$ is not spanned by the D columns of $\mathbf{B}_X^*$ but rather some D orthogonal directions in the span of the columns of $\left[\mathbf{B}_X^* \mid \mathbf{B}_c\right]$, as illustrated in Figure 2.
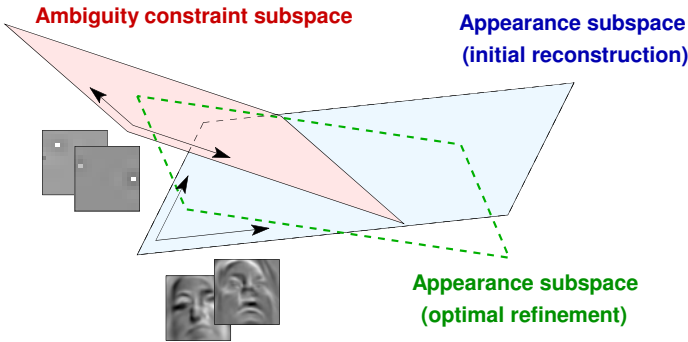

Figure 2: A conceptual illustration of the main idea: the initial reconstruction of the class subspace in the high dimensional image space is refined through rotation within the constraints of the ambiguity constraint subspace.

Let $\mathbf{B}_{Xc}$ be a matrix of orthonormal basis vectors computed by orthogonalizing $\left[\mathbf{B}_X^* \mid \mathbf{B}_c\right]$:

$$\mathbf{B}_{Xc} = \text{orth}\left(\left[\mathbf{B}_X^* \mid \mathbf{B}_c\right]\right) \tag{12}$$

Then we seek a matrix $\mathbf{T} \in \mathbb{R}^{(D+d_h-d_l)\times D}$ which makes the optimal choice of D directions from the span of $\mathbf{B}_{Xc}$:

$$\mathbf{B}_{Xc}\,\mathbf{T} = \mathbf{B}_{Xc}\left[\mathbf{t}_1 \mid \mathbf{t}_2 \mid \dots \mid \mathbf{t}_D\right]. \tag{13}$$

Here we define the optimal choice of $\mathbf{T}$ as the one that best aligns the reconstructed subspace with the subspace it is compared with, i.e. $\mathbf{B}_Y$. The matrix $\mathbf{T}$ can be constructed recursively, so let us consider how its first column $\mathbf{t}_1$ can be computed. The optimal alignment criterion can be restated as:

$$\mathbf{t}_1 = \arg\max_{\mathbf{t}_1'}\max_{\mathbf{a},\mathbf{b}} \frac{\left(\mathbf{B}_Y\,\mathbf{a}\right)\cdot\left(\mathbf{B}_{Xc}\,\mathbf{t}_1'\,\mathbf{b}\right)}{\|\mathbf{a}\|\,\|\mathbf{b}\|} \tag{14}$$

The right-hand side can be rewritten as:

$$\frac{\left(\mathbf{B}_Y\,\mathbf{a}\right)\cdot\left(\mathbf{B}_{Xc}\,\mathbf{t}_1\,\mathbf{b}\right)}{\|\mathbf{a}\|\,\|\mathbf{b}\|} = \frac{\mathbf{a}}{\|\mathbf{a}\|}\,\mathbf{B}_Y^{\,T}\,\mathbf{B}_{Xc}\,\mathbf{t}_1\,\frac{\mathbf{b}}{\|\mathbf{b}\|} = \frac{\mathbf{a}}{\|\mathbf{a}\|}\,\mathbf{U}\,\Sigma\,\mathbf{V}^T\,\mathbf{t}_1\,\frac{\mathbf{b}}{\|\mathbf{b}\|}, \tag{15}$$

where:

$$\mathbf{B}_Y^{\,T}\,\mathbf{B}_{Xc} = \overbrace{\left[\mathbf{u}_1 \mid \dots \mid \mathbf{u}_D\right]}^{\mathbf{U}} \overbrace{\begin{bmatrix} \sigma_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & 0 \\ 0 & 0 & \dots & \sigma_D & \dots & 0 \end{bmatrix}}^{\Sigma} \overbrace{\left[\mathbf{v}_1 \mid \dots \mid \mathbf{v}_{D+d_h-d_l}\right]^T}^{\mathbf{V}^T}, \tag{16}$$

is the Singular Value Decomposition of $\mathbf{B}_Y^{\,T}\,\mathbf{B}_{Xc}$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_D$. Then, from the right-hand side in Equation (15), by inspection the optimal directions of $\mathbf{a}$ and $\mathbf{b}$ are the first "output" direction $\mathbf{u}_1$ and the first "input" direction $\mathbf{v}_1$ respectively, and $\mathbf{t}_1 = \mathbf{v}_1$. The same process can be used to infer the remaining columns of $\mathbf{T}$, the i-th one being $\mathbf{t}_i = \mathbf{v}_i$.

Thus, the optimal reconstruction $\mathbf{B}_X'$ of $\mathbf{B}_X$ in the high-dimensional space, obtained by the constrained rotation of the naïve estimate $\mathbf{B}_X^*$, is given by the orthonormal basis matrix:

$$\mathbf{B}_X' = \mathbf{B}_{Xc}\left[\mathbf{v}_1 \mid \dots \mid \mathbf{v}_D\right] \tag{17}$$

The key steps of the algorithm are summarized in Figure 3.

Efficiency and Implementation Issues.   Before turning our attention to the empirical analysis of the proposed algorithm let us briefly highlight the low additional computational load imposed by the refinement of the re-constructed class subspace in the high-dimensional image space. Specifically, note that the output of Steps 1 and 3 in Figure 3 can be pre-computed, as it is dependent only on the *dimensions* of the low and high scale data, not the data itself. Orthogonalization in Step 2 is fast, as D – the number of columns in $\mathbf{B}_X$ – is small. Although

Input: Orthonormal subspace basis matrices $\mathbf{B}_X \in \mathbb{R}^{d_l}$, $\mathbf{B}_Y \in \mathbb{R}^{d_h}$
Projection model $\mathbf{P} \in \mathbb{R}^{d_l \times d_h}$

Output: Optimal reconstruction $\mathbf{B}'_X$ of the high-dimensional space corresponding to $\mathbf{B}_X$

1: Compute the reverse projection matrix
$$\mathbf{P}_R = \mathbf{P}^T \left(\mathbf{P}\,\mathbf{P}^T \right)^{-1}$$

2: Compute the initial naïve reconstruction
$$\mathbf{B}_X^* = \mathrm{orth}\left(\mathbf{P}_R\,\mathbf{B}_X \right)$$

3: Compute a basis of the ambiguity constraint subspace
$$\mathbf{B}_c = \mathrm{nullspace}\left(\mathbf{P} \right)$$

4: Compute a joint basis of the initial reconstruction and the ambiguity constraint subspace
$$\mathbf{B}_{Xc} = \mathrm{orth}\left(\left[\mathbf{B}_c \mid \mathbf{B}_X^* \right] \right)$$

6: Perform Singular Value Decomposition of $\mathbf{B}_Y{}^T \mathbf{B}_{Xc}$
$$\mathbf{B}_Y{}^T \mathbf{B}_{Xc} = \mathbf{U}\,\Sigma\,\mathbf{V}^T = \mathbf{U}\,\Sigma\,\left[\mathbf{v}_1 \mid \dots \mid \mathbf{v}_D \right]^T$$

7: Extract the orthonormal basis of the best reconstruction
$$\mathbf{B}'_X = \mathbf{B}_{Xc}\left[\mathbf{v}_1 \mid \dots \mid \mathbf{v}_D \right]$$

Figure 3: A summary of the proposed matching algorithm.

at the first sight more complex, the orthogonalization in Step 4 is also not demanding, as $\mathbf{B}_c$ is already orthonormal, so it is in fact only the D columns of $\mathbf{B}_X^*$ which need to be adjusted. Lastly, the Singular Value Decomposition in Step 6 operates on a matrix which has a high "landscape" eccentricity so the first D "input" directions can be computed rapidly, while Step 7 consists only of a simple matrix multiplication.

3 Experimental Analysis

The theoretical arguments put forward in the preceding sections were evaluated empirically on the problem of matching sets of images of faces. For this we used a collection of video sequences of pseudo-random head motion described in [1].

Protocol.   Evaluation was performed by constructing class models with downsampled face images in a single illumination setting. Thus each class represented by a linear subspace corresponds to a single person and captures his/her appearance in the training illumination. Images downsampled to 25×25, 20×20, 15×15, 10×10 and 5×5 pixels were used in turn. Training subspaces were then matched against subspaces estimated from higher scale data – we used 50×50 pixel images throughout – and each query subspace classified to the class of the highest similarity. The similarity between two subspaces was expressed by a number

in the range $[0,1]$, equal to the correlation of the two highest correlated vectors confined to them (as in Equation (14) in the previous section).

Class separation.   First, we looked into the effects of the method proposed in Section 2.2 on class separation, in comparison to the naïve method of Section 2.1. We quantified this as follows. For a given pair of training and "query" illumination conditions, we evaluated the similarity $\rho_{i,j}$ between all image sets i acquired in the training illumination and all sets j acquired in the query illumination. Thus, the mean confidences $\bar{e}_w$ and $\bar{e}_b$ of respectively the correct and incorrect matching assignments is given by:

$$\bar{e}_w = 1 - \frac{1}{100} \sum_{i=1}^{100} \rho_{i,i} \tag{18}$$

$$\bar{e}_b = 1 - \frac{1}{100 \times 99} \sum_{i=1}^{100} \sum_{\substack{j=1 \\ j \neq i}}^{100} \rho_{i,j}. \tag{19}$$

The corresponding separation is proportional to $\bar{e}_b$ and inversely proportional to $\bar{e}_w$:

$$\mu = \bar{e}_b \, \bar{e}_w^{-1}. \tag{20}$$

The separation was evaluated separately for all training-query illumination pairs for the naïve method and the proposed solution across different matching scales, and plotted in Figure 4.

Firstly, note that improvement was observed for all illumination combinations at all scales. Unsurprisingly, the most significant increase in class separation ($\approx$ 8.5-fold mean increase) was achieved for the most drastic difference in training and query sets, when subspaces embedded in a 25-dimensional image space – representing the appearance variation of images as small as 5×5 pixels – was matched against a subspace embedded in the image space of a 100 times greater dimensionality.

It is interesting to note that even at the more favourable scales of the low resolution input, although the mean improvement was less noticeable than at extreme scale discrepancies, the accuracy of matching in certain combinations of illumination settings still greatly benefited from the proposed method. For example, for low resolution subspaces representing appearance in 10×10 pixel images, the mean separation increase of 75.6% was measured; yet, for illuminations "1" and "2" – corresponding to the index 42 on the abscissa in Figure 4 (b) – the improvement was 473.0%.

The change effected on the inter-class and intra-class distances is illustrated in Figure 5, which shows a typical similarity matrix produced by the naïve and the proposed matching methods. The mean separation increase across different scales is shown in Figure 6 – 8.5-fold for 5×5 pixel images, 1.75-fold for 10×10, 1.25-fold for 15×15, 1.08-fold for 20×20 and 1.03-fold for 25×25.

Lastly, the inferred most similar modes of variation contained within two subspaces representing face appearance variation of the same person in different illumination conditions and at different training scales is shown in Figure 7. As the scale of low-resolution images is reduced, the naïve algorithm of Section 2.1 finds progressively worse matching modes with significant visual degradation in the mode corresponding to the low-resolution subspace. In contrast, the proposed algorithm correctly reconstructs meaningful high-resolution appearance even in the case of extremely low resolution images (5×5 pixels).
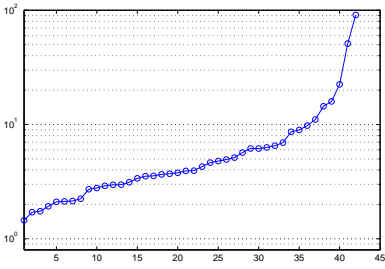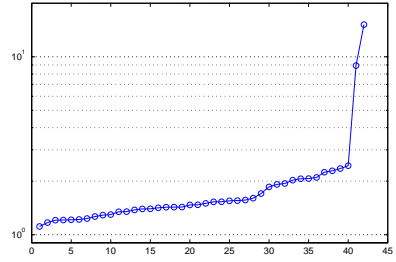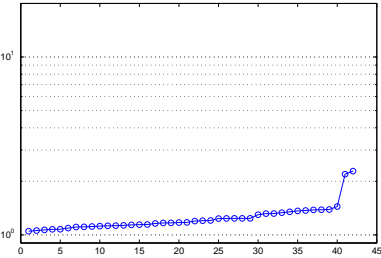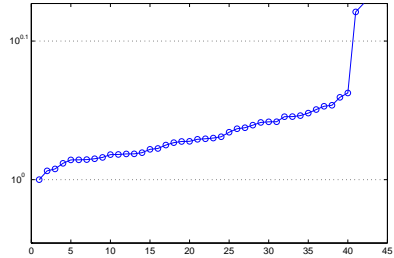
(a) $(5 \times 5) \longleftrightarrow (50 \times 50)$

(b) $(10 \times 10) \longleftrightarrow (50 \times 50)$

(c) $(15 \times 15) \longleftrightarrow (50 \times 50)$

(d) $(20 \times 20) \longleftrightarrow (50 \times 50)$

Figure 4: The increase in class separation μ (the ordinate; note that the scale is logarithmic) over different training-query illumination conditions (abscissa), achieved by the proposed method in comparison to the naïve subspace re-projection approach. Note that for clarity the training-query illumination pairs were ordered in increasing order of improvement for each plot; thus, the indices of different abscissae do not necessarily correspond.
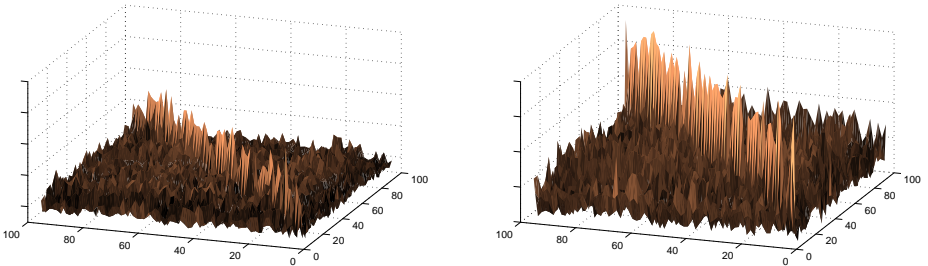

Figure 5: Typical similarity matrices resulting from the naïve (left) and the proposed (right) matching approaches. Our method produces a dramatic improvement in class separation as witnessed by the increased dominance of the diagonal elements of the similarity matrix.
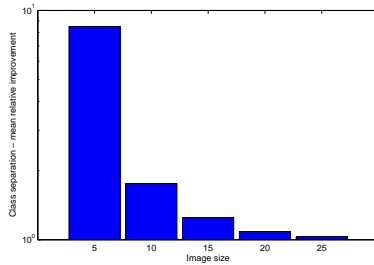
Figure 6: Mean class separation increase achieved, as a function of size of the low-scale images. Shown is the ratio of class separation when subspaces are matched using the proposed method and the naïve re-projection method described in Section 2.1. The rate of improvement decay is incrementally exponential, reaching 1 (no improvement) when the $d_l = d_h$.

(a) $(5 \times 5) \longleftrightarrow (50 \times 50)$ (b) $(10 \times 10) \longleftrightarrow (50 \times 50)$ (c) $(15 \times 15) \longleftrightarrow (50 \times 50)$

(d) $(20 \times 20) \longleftrightarrow (50 \times 50)$ (e) $(25 \times 25) \longleftrightarrow (50 \times 50)$

Figure 7: The inferred most similar modes of variation contained within two subspaces representing face appearance variation of the same person in different illumination conditions and at different training scales. In each subfigure, which corresponds to a different training-query scale discrepancy, the top pair of images represents appearance extracted by the naïve algorithm of Section 2.1 (as the left-singular and right-singular vectors of $\mathbf{B}_Y{}^T \mathbf{B}_X^*$); the bottom pair is extracted by the proposed method (as the left-singular and right-singular vectors of $\mathbf{B}_Y{}^T \mathbf{B}_{Xc}$).

4 Conclusion

In this paper we described a method for matching linear subspaces which represent appearance variations in images of different scales. Our approach consists of an initial re-projection of the subspace in the low-dimensional image space to the high-dimensional one, and subsequent refinement of the re-projection through a constrained rotation. Using facial appearance

images, on a large data set of 700 image sets in different illuminations, we showed that the proposed algorithm successfully reconstructs the personal subspace in the high-dimensional image space even for low-dimensional input corresponding to images as small as 5×5 pixels, improving average class separation by an order of magnitude.

References

[1] O. Arandjelović and R. Cipolla. Face recognition from video using the generic shape-illumination manifold. *In Proc. European Conference on Computer Vision (ECCV)*, 4: 27–40, May 2006.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(2):218–233, February 2003.

[3] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision (IJCV)*, 28 (3):245–260, July/August 1998.

[4] M. Bethge. Factorial coding of natural images: how effective are linear models in removing higher-order dependencies? *Journal of the Optical Society of America (JOSA A)*, 23(6):1253–1268, June 2006.

[5] P. Chen and D. Suter. An analysis of linear subspace approaches for computer vision and pattern recognition. *International Journal of Computer Vision (IJCV)*, 68(1):83–106, June 2006.

[6] M. Everingham, A. Zisserman, C. Williams, C. Van Gool, et al. The 2005 PASCAL visual object classes challenge. *In Selected Proc. First PASCAL Challenges Workshop*, 2006.

[7] V. Ferrari, T. Tuytelaars, , and L. Van Gool. Retrieving objects from videos based on affine regions. *In Proc. European Signal Processing Conference (EUSIPCO)*, pages 128–131, September 2004.

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(6):643–660, June 2001.

[9] P. Hall, D. Marshall, and R. Martin. Merging and splitting eigenspace models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(9):1042–1048, September 2000.

[10] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. 2nd edition, March 2004. ISBN 978-0-521-54051-3.

[11] A.S. Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM*, 5(4):339–342, October 1958.

[12] K. Lee, M. Ho, J. Yang, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(5):684–698, May 2005.

[13] R. Pradhan, Z. G. Bhutia, M. Nasipuri, and M. P. Pradhan. Gradient and principal component analysis based texture recognition system: A comparative study. *Fifth International Conference on Information Technology: New Generations*, pages 1222–1223, April 2009.

[14] Y. Su, S. Shan, X. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. *IEEE Transactions on Image Processing*, 18(8):1885–1896, August 2009.

[15] S. Zhou, G. Aggarwal, R. Chellappa, and D. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(2):230–245, February 2007.