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

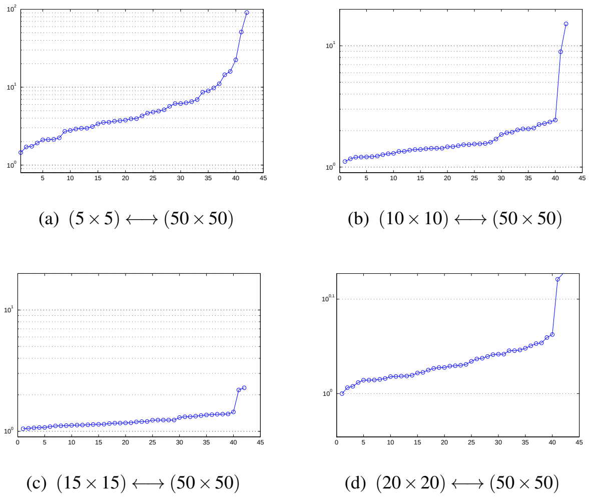
One of the most commonly encountered problems in computer vision is that of matching appearance. A particularly interesting and increasingly important instance of this task concerns the matching of *sets* of appearance images, each set containing examples of variation corresponding to a single class. A ubiquitous representation of appearance variation within a class is by a linear subspace. 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 or smooth objects across changing pose. What is more, linear subspace models are also attractive for their low storage demands – they are inherently compact and can be learnt incrementally. 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.

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 theoretical arguments put forward in the preceding sections were evaluated empirically on the problem of matching sets of images of faces. 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 – 50×50 pixel images were used throughout – and each query subspace classified to the class of the highest similarity.

Class separation was evaluated separately for all training-query illumination pairs for the naïve method and the proposed solution across different matching scales, and is shown plotted in Figure 1. The mean separation increase across different scales is shown in Figure 2 – 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 3. As the scale of low-resolution images is reduced, the naïve algorithm 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) $(5\times5)\longleftrightarrow(50\times50)$ (b) $(10\times10)\longleftrightarrow(50\times50)$

(c) $(15\times15)\longleftrightarrow(50\times50)$ (d) $(20\times20)\longleftrightarrow(50\times50)$

Figure 1: The increase in class separation (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. 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.

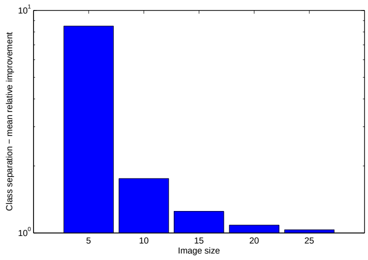
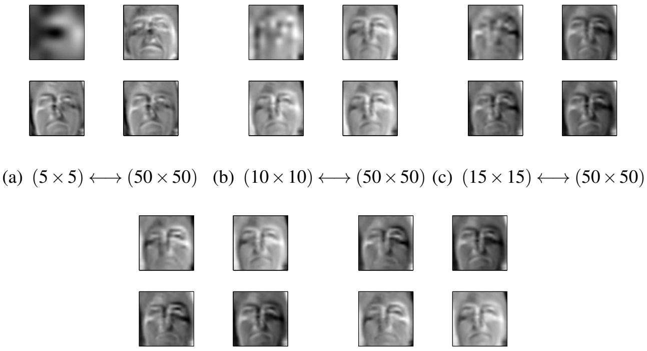


Figure 2: Mean class separation increase achieved, as a function of size of the low-scale images. Shown is the ratio of class separation when subspaces are matched using the proposed method and the naïve re-projection method. The rate of improvement decay is incrementally exponential, reaching 1 (no improvement) when low and high resolution image sizes are the same (50 pixels).



(a) $(5\times5)\longleftrightarrow(50\times50)$ (b) $(10\times10)\longleftrightarrow(50\times50)$ (c) $(15\times15)\longleftrightarrow(50\times50)$

(d) $(20\times20)\longleftrightarrow(50\times50)$ (e) $(25\times25)\longleftrightarrow(50\times50)$

Figure 3: 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; the bottom pair is extracted by the proposed method.