Pattern Recognition 47 (2014) 2662–2672

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Hallucinating optimal high-dimensional subspaces

Ognjen Arandjelović *

Centre for Pattern Recognition and Data Analytics (PRaDA), School of Information Technology, Deakin University, Geelong 3216 VIC, Australia

ARTICLE INFO

Article history:
Received 3 May 2013
Received in revised form
31 January 2014
Accepted 12 February 2014
Available online 20 February 2014

Keywords:
Projection
Ambiguity
Constraint
SVD
Similarity
Face

ABSTRACT

Linear subspace representations of appearance variation are pervasive in computer vision. This paper addresses the problem of robustly matching such subspaces (computing the similarity between them) when they are used to describe the scope of variations within sets of images of different (possibly greatly so) scales. A naïve solution of projecting the low-scale subspace into the high-scale image space is described first and subsequently shown to be 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 method is evaluated on the problem of matching sets of (i) face appearances under varying illumination and (ii) object appearances under varying viewpoint, using two large data sets. In comparison to the naïve matching, the proposed 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 [1], views of objects [2] or faces [3], textures [4] or rectified planar structures (buildings, paintings) [5], 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 [6,7]. 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 [8–10] or smooth objects across changing pose [11,12]. Moreover, linear subspace models are also attractive for their low storage demands – they are inherently compact and can be learnt incrementally [13–18]. Indeed, throughout this paper it is assumed 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 of different scales. This is a frequent occurrence: an object one wishes 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 it is addressed in detail and shown that a naïve approach to normalizing for scale in subspaces results in inadequate matching performance. Thus, a method is proposed 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, a brief overview of the linear subspace representation is given first, followed by a description of the aforementioned naïve scale normalization. The proposed solution is described in this section as well. In Section 3 the two approaches are compared empirically and the results are analysed in detail. The main contribution and conclusions of the paper are summarized in Section 4.

* Tel.: +61 3522 73079.
E-mail address: ognjen.arandjelovic@gmail.com
URL: http://mi.eng.cam.ac.uk/~oa214

http://dx.doi.org/10.1016/j.patcog.2014.02.006

2. Matching subspaces across scale

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

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

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 $D \gg d$. 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

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

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

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

where $\|\cdot\|_F$ 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, here it is assumed that all images both within each set, as well as across the two sets, are of the same aspect ratio. Thus, 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

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

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 Fig. 1 for an illustration.

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

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

Since it is assumed 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 $\mathbb{R}^{d_h}$:

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

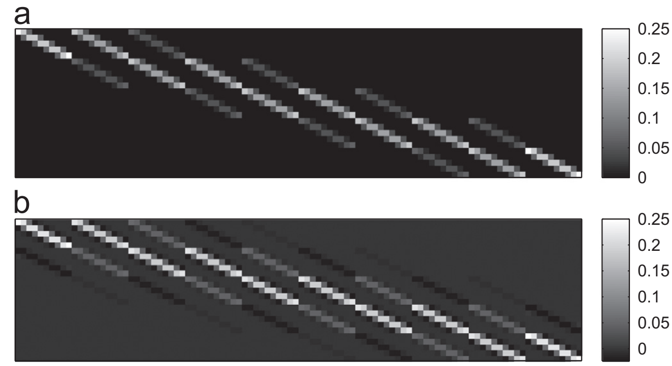

Fig. 1. The projection matrix $\mathbf{P} \in \mathbb{R}^{25 \times 100}$ modelling the process of downsampling a 10×10 pixel image to 5×5 pixels, using (a) bilinear and (b) bicubic projection models, shown as an image. For the interpretation of image intensities see the associated grey level scales on the right.

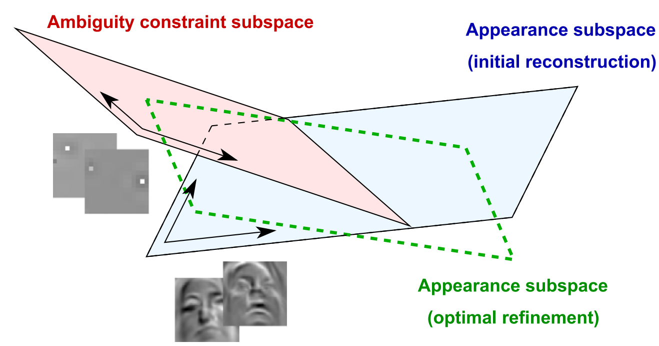

Fig. 2. A conceptual illustration of the main idea: the initial reconstruction of the class subspace in the high dimensional image space is refined through rotation within the constraints of the ambiguity constraint subspace.

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, it is orthogonalized using the Householder transformation [19], producing a high-dimensional subspace basis estimate $\mathbf{B}_X^*$ which can be compared directly with $\mathbf{B}_Y$.

2.1.1. 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, this is modelled 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

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

or, equivalently

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