Fast Active Appearance Model Search Using Canonical Correlation Analysis

René Donner, Michael Reiter, Georg Langs, Philipp Peloschek, and Horst Bischof

Abstract—A fast AAM search algorithm based on *canonical correlation analysis (CCA-AAM)* is introduced. It efficiently models the dependency between texture residuals and model parameters during search. Experiments show that CCA-AAMs, while requiring similar implementation effort, consistently outperform standard search with regard to convergence speed by a factor of four.

Index Terms—Image processing and computer vision, active appearance models, statistical image models, subspace methods, medical imaging.

1 Introduction

Active appearance models (AAMs) [7] learn characteristics of objects during a training phase by building a compact statistical model representing shape and texture variation of the object. The use of this a priori knowledge enables the AAM search to yield good results even on difficult and noisy data. AAMs have been employed in various domains like face modeling [11], studying human behavior [17], and medical imaging tasks, like segmentation of cardiac MRIs [21] or the diaphragm in CT data [1], and registration in functional heart imaging [25]. In [24], an extensive overview of existing applications is given.

1.1 AAM Search

The goal of the AAM search is to find the model parameters that generate a synthetic image as close as possible to a given input image and to use the resulting AAM parameters for interpretation [7]. Matching the model and target image is treated as an optimization problem, i.e., the problem of minimizing the texture residual with regard to model parameters. Provided that the model is roughly aligned with the target image, the relation of texture residuals and parameter updates can be modeled a priori (by offline training) within a certain class of objects [7].

In the original AAM search approach proposed by Cootes et al. [5], the mapping from error images to AAM parameters is modelled by a linear regression approach (linear least-squares estimates). In the later proposed optimization approach [7], the regression estimates were replaced by a simplified Gauss-Newton procedure, where the Jacobian matrix is evaluated only once (offline) by numerical differentiation from training data. Throughout this paper, we will refer to this as standard approach. Both approaches are similar in the sense that they assume that the error surface can be approximated reasonably well by a quadratic function. The main advantage of the latter approach [7] is that, during training, not all difference images have to be stored in memory.

Various approaches to increase convergence speed and AAM search result accuracy have been proposed. A review of various search techniques is given in [4]. ShapeAAMs [6] update only pose and shape parameters during search, while gray-level parameters are computed directly from the sample. They converge faster but the failure rate increases. Direct appearance models (DAMs) [15] predict shape parameters directly from texture. The convergence speed of AAMs for tracking applications is investigated in [10].

These modifications of the original approach improve the convergence speed and the quality of the results by either reducing the number of parameters that are to be optimized in a sensible way (DAMs and ShapeAAMs), by saving computation time by reducing the synthesizing steps necessary for the error function calculation (AAM tracking), or by reducing noise in the regression training images (DAMs). However, the basic parameter updating scheme during search is based on the standard approach [7] for all of these methods. They represent heuristic approaches to AAM fitting, which trade accuracy for efficiency by assuming a constant Jacobian during search.

An alternative is to perform analytic gradient descent. To avoid the resulting inefficiency during search, in [19], an inverse compositional approach (ICIA) was proposed. It treats shape and appearance variation independently and is based on a variant of the Lucas-Kanade image alignment algorithm [18], which performs a Gauss-Newton optimization. The texture warp is composed of incremental warps and, thus, the assumption that the Jacobian used for parameter prediction stays constant becomes valid. In [12], an extension to this approach (simultaneous ICA) was proposed for combined AAMs. The analytical calculation of the Jacobian is based on the mean and partial derivatives of the AAM parameters.

The approach presented in this paper utilizes training images, allowing CCA to extract additional regression-relevant information which may be discarded by the purely generative PCA-based model.

1.2 A Fast CCA-Based Search

In this paper, we present an AAM search algorithm based on *canonical correlation analysis (CCA)* [14]. CCA is a statistical method for factor analysis in two signal spaces. It determines linear combinations of variables (canonical variates) in each of the two signals, which are pairwise maximally correlated. The directions of maximum correlation (canonical factors) capture relevant signal components, constituting the functional relation of the two signals.

There exist a number of related regression techniques, such as *Partial Least Squares* [13], *Reduced Rank Wiener Filtering* (see, for example, [9]. CCA, in particular, has some very attractive properties (for example, it is invariant with regard to affine transformations—and, thus, scaling—of the input variables) and cannot only be used for regression purposes, but whenever we need to establish a relation between two sets of measurements (e.g., finding corresponding points in stereo images [2]). In signal processing, CCA is used for optimal reduced-rank filtering [16], where the goal is data reduction, robustness against noise, and high computational efficiency. It has also been successfully applied to pattern classification [23], appearance-based 3D pose estimation [20], and stereo vision [2].

Our approach follows the original AAM training procedures proposed by Cootes et al. [5], [7]. Essentially, we use a linear regression model of the texture residual vector $\mathbf{r} \in \mathbb{R}^p$ and corresponding AAM parameter displacements $\delta\mathbf{p} \in \mathbb{R}^q$ (p is the size of the synthetic image and q is the number of parameters used in the model). In our approach, however, we use reduced-rank estimates obtained by CCA, instead of standard linear least-squares regression estimates.

- *R. Donner and G. Langs are with the Pattern Recognition and Image Processing Group, Institute of Computer Aided Automation, Vienna University of Technology, Favoritenstr. 9/183/2, A-1040 Vienna, Austria, and the Computer Vision Group, Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria. E-mail: {donner, langs}@prip.tuwien.ac.at.*
- *M. Reiter is with the Pattern Recognition and Image Processing Group, Institute of Computer Aided Automation, Vienna University of Technology, Favoritenstr. 9/183/2, A-1040 Vienna, Austria. E-mail: rei@prip.tuwien.ac.at.*
- *P. Peloschek is with the Department of Radiology, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Wien, Austria. E-mail: philipp.peloschek@meduniwien.ac.at.*
- *H. Bischof is with the Computer Vision Group, Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria. E-mail: bischof@icg.tu-graz.ac.at.*

Manuscript received 4 Oct. 2005; revised 10 Feb. 2006; accepted 10 Feb. 2006; published online 11 Aug. 2006.
Recommended for acceptance by S. Baker.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0536-1005.

The motivation of CCA is twofold. First, in the standard approach, the regression matrix consisting of a large number ($p \times q$) of parameters has to be estimated from a limited number of (noisy) training images. The rank constraint therefore leads to more reliable regression parameter estimates [3], [8]. Second, we assume that the true regression matrix is of lower rank than $\min(p, q)$ because the texture residuals contain regression-irrelevant components, including noise and uncorrelated (higher order) components, which cannot be captured by the linear model.

We will show experimentally on different types of data that, indeed, CCA provides more accurate parameter updates that lead to faster convergence of the AAM search.

1.3 Paper Structure

In Section 2, AAMs, their standard training approach, and CCA are presented. The proposed CCA-based search algorithm is explained in detail in Section 3. In Section 4, results are shown and conclusions are drawn in Section 5.

2 METHODS

2.1 Active Appearance Models

The concept of active appearance models as described in [7] is based on the idea of combining both shape and texture information of the objects to be modeled. First, the shape vectors $\mathbf{x^i} = (x_1^i, \ldots, x_n^i, y_1^i, \ldots, y_n^i)^T, i = 1 \ldots N$ of the N training images are aligned using Procrustes analysis. The images are warped to the mean shape $\overline{\mathbf{x}}$ and normalized, yielding the texture vectors $\mathbf{g^i}$. By applying principal component analysis to the normalized data, linear models are obtained for both shape, $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s\mathbf{b}_s$, and texture, $\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g\mathbf{b}_g$, where $\overline{\mathbf{x}}, \overline{\mathbf{g}}$ are the mean vectors, $\mathbf{P}_s, \mathbf{P}_g$ are sets of orthogonal modes of variation (the eigenvectors resulting from PCA), and $\mathbf{b}_s, \mathbf{b}_g$ are sets of model parameters.

A given object can thus be described by $\mathbf{b}_s$ and $\mathbf{b}_g$. As $\mathbf{P}_s, \mathbf{P}_g$ may still be correlated, PCA is applied once more using the following concatenated vector:

$$\mathbf{b} = \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} = \begin{pmatrix} \mathbf{W}_s\mathbf{P}_s^T(\mathbf{x} - \overline{\mathbf{x}}) \\ \mathbf{P}_g^T(\mathbf{g} - \overline{\mathbf{g}}) \end{pmatrix},$$

where $\mathbf{W}_s$ is a diagonal scaling matrix derived from the value ranges of the eigenvalues of the shape and texture eigenspaces. This yields the final combined linear model $\mathbf{b} = \mathbf{P}_c\mathbf{c}$, where $\mathbf{P}_c = (\mathbf{P}_{cs}^T, \mathbf{P}_{cg}^T)^T$.

Shape free images and the corresponding shapes defining the deformation of the texture can be expressed directly using $\mathbf{c}$ by $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s\mathbf{W}_s^{-1}\mathbf{P}_{cs}\mathbf{c}$ and $\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g\mathbf{P}_{cg}\mathbf{c}$. To enable the model to deal with rotation, scaling, and translation the additional model parameters $\mathbf{t}$, capturing scaling and rotation and $\mathbf{u}$, modelling image contrast and brightness, are introduced. The resulting AAM model represents shape and texture variation of image content utilizing a single parameter vector $\mathbf{p} = (\mathbf{c}^T|\mathbf{t}^T|\mathbf{u}^T) \in \mathbb{R}^q$.

2.2 Standard AAM Search Approach

Provided we have a trained AAM where model parameters $\mathbf{p}$ generate synthetic images $\mathbf{I}_{model}(\mathbf{p})$, the standard search for an optimal match minimizes the difference between a given image $\mathbf{I}_{image}$ and the reconstructed image $\mathbf{I}_{model}(\mathbf{p})$. The search for the model parameters $\mathbf{p}$ can be guided by using knowledge about how the difference images correlate with the parameter displacements. This knowledge is obtained during training.

During each search step, the current image residual between the model texture $\mathbf{g}_m(\mathbf{p})$ and the sampled image patch $\mathbf{g}_s(\mathbf{p})$ (warped to the mean shape) is computed using

$$\mathbf{r}(\mathbf{p}) = \mathbf{g}_s(\mathbf{p}) - \mathbf{g}_m(\mathbf{p}). \tag{1}$$

The search procedure aims at minimizing the sum of square (pixel) error

$$\frac{1}{2}\mathbf{r}(\mathbf{p})^T\mathbf{r}(\mathbf{p}). \tag{2}$$

Following the standard Gauss-Newton optimization method one approximates (linearizes) (1) using the first-order Taylor expansion

$$\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) \approx \mathbf{r}(\mathbf{p}) + \frac{\partial\mathbf{r}}{\partial\mathbf{p}}\delta\mathbf{p},$$

with the ijth element of matrix $\frac{\partial\mathbf{r}}{\partial\mathbf{p}}$ being $\frac{\partial r_i}{\partial p_j}$.

Building the derivative of (2) with regard to p and setting it to zero gives

$$\delta\mathbf{p} = -\mathbf{R}\,\mathbf{r}(\mathbf{p}), \tag{3}$$

where

$$\mathbf{R} = \left(\frac{\partial\mathbf{r}}{\partial\mathbf{p}}^T\frac{\partial\mathbf{r}}{\partial\mathbf{p}}\right)^{-1}\frac{\partial\mathbf{r}}{\partial\mathbf{p}}^T = \left(\frac{\partial\mathbf{r}}{\partial\mathbf{p}}\right)^{\dagger},$$

with † denoting the pseudoinverse and $\mathbf{R}$ has size $q \times k$. Instead of recalculating $\frac{\partial\mathbf{r}}{\partial\mathbf{p}}$ at every step, it is computed once during training using numeric differentiation.

During training, each parameter is displaced from its optimal value in h steps from -1 to +1 standard deviations, and a weighted average of the resulting difference images over the training set is built:

$$\frac{dr_i}{dp_j} = \sum_h \omega(\delta p_{jh})\frac{(r_i(\mathbf{p} + \delta p_{jh}) - r_i(\mathbf{p}))}{\delta p_{jh}}.$$

During the actual search, each iteration updates the model parameters using $\mathbf{p}_{next}(s) = \mathbf{p}_{current} + s\delta\mathbf{p}_{predicted}$, with $\delta\mathbf{p}_{predicted} = -\mathbf{R}\,\mathbf{r}_{current}$ and s being a scaling factor sequentially chosen from $\mathbf{s}_{steps} = \langle 1, 0.5, 1.5, 0.25, 0.1, 2, 0.025, 0.01\rangle$, as proposed in [7]. At each of these scaling steps, the image patch is compared to the syntetisized image $\mathbf{I}_{model}$, which is computationally expensive.

Let $E(\mathbf{p}_{current}) = |\mathbf{r}(\mathbf{p}_{current})|^2 = |\mathbf{g}_s - \mathbf{g}_m|^2$ be the error of the current model. An iteration is declared successful for the first step s to produce an error $E(\mathbf{p}_{next}(s)) < E(\mathbf{p}_{current})$. $\mathbf{p}_{current}$ is then set to $\mathbf{p}_{next}(s)$ and the search continues with the next iteration. If no $\mathbf{p}_{next}(s)$ better than $\mathbf{p}_{current}$ can be found, convergence is declared and $\mathbf{p}_{current}$ is the best estimate for the model parameters. As will be shown in Section 4, our approach eliminates the need for using different step sizes, as the parameter predictions are more accurate.

2.3 Canonical Correlation Analysis and Linear Regression

Canonical Correlation Analysis is a very powerful tool that is especially well suited for relating two sets of measurements (signals). Like *principal components analysis* (PCA), CCA also reduces the dimensionality of the original signals, since only a few factor-pairs are normally needed to represent the relevant information; unlike PCA, however, CCA takes into account the relationship between two signal spaces (in the correlation sense), which makes them better suited for regression tasks than PCA.

Given two zero-mean random variables $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$, CCA finds pairs of directions $\mathbf{w}_x$ and $\mathbf{w}_y$ that maximize the correlation between the projections $x = \mathbf{w}_x^T\mathbf{x}$ and $y = \mathbf{w}_y^T\mathbf{y}$ (in the context of CCA, the projections x and y are also referred to as *canonical variates*). More formally, the directions can be found as maxima of the function

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\mathbf{w}_x^T\mathbf{x}\mathbf{y}^T\mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T\mathbf{x}\mathbf{x}^T\mathbf{w}_x]E[\mathbf{w}_y^T\mathbf{y}\mathbf{y}^T\mathbf{w}_y]}},$$

$$\rho = \frac{\mathbf{w}_x^T\mathbf{C}_{xy}\mathbf{w}_y}{\sqrt{\mathbf{w}_x^T\mathbf{C}_{xx}\mathbf{w}_x\mathbf{w}_y^T\mathbf{C}_{yy}\mathbf{w}_y}},$$

whereby $\mathbf{C}_{xx} \in \mathbb{R}^{p\times p}$ and $\mathbf{C}_{yy} \in \mathbb{R}^{q\times q}$ are the *within-set covariance matrices* of $\mathbf{x}$ and $\mathbf{y}$, respectively, while $\mathbf{C}_{xy} \in \mathbb{R}^{p\times q}$ denotes their *between-set covariance matrix*. A number of at most $k = \min(p, q)$ factor pairs $\langle \mathbf{w}_x^i, \mathbf{w}_y^i\rangle, i = 1, \ldots, k$ can be obtained by successively

solving $\mathbf{w}^i = (\mathbf{w}_x^{iT}, \mathbf{w}_y^{iT})^T = \arg\max_{(\mathbf{w}_x^i, \mathbf{w}_y^i)}\{\rho\}$ subject to $\rho(\mathbf{w}_x^j, \mathbf{w}_y^i) = \rho(\mathbf{w}_x^i, \mathbf{w}_y^j) = 0$ for $j = 1, \ldots, i-1$. The factor pairs $\mathbf{w}^i$ can be obtained as solutions (i.e., eigenvectors) of a generalized eigenproblem (for details, see, e.g., [20]). The extremum values $\rho(\mathbf{w}^i)$, which are referred to as *canonical correlations*, are obtained as the corresponding eigenvalues.

Alternatively, $\mathbf{w}^i$ can be obtained by Singular Value Decomposition (SVD) of the cross-correlation matrix of prewhitened input data. Let $\mathbf{A} = \mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}$ and $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ be the SVD of $\mathbf{A}$, where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_q)$ are orthogonal matrices and $\mathbf{D}$ is a diagonal matrix with singular values. The canonical factors can be obtained as $\mathbf{w}_x^i = \mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{u}_i$ and $\mathbf{w}_y^i = \mathbf{C}_{yy}^{-\frac{1}{2}}\mathbf{v}_i$. It is instructive to compare CCA to the full-rank solution of standard multivariate linear regression (MLR), where the regression parameters $\mathbf{W}$ are given by the *Wiener filter*: $\mathbf{W} = E[\mathbf{x}\mathbf{x}^T]^{-1}E[\mathbf{x}\mathbf{y}^T] = \mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}$. In contrast to MLR, the CCA solution is computed using only the leading singular vectors of the cross-correlation matrix of prewhitened variables $\mathbf{x}, \mathbf{y}$ which are made explicit by SVD. Thus, CCA can be used to compute a (reduced) rank-n regression parameter matrix by using only $n < k$ factor pairs.

Thereby, in contrast to standard multivariate regression, CCA takes advantage of the correlations between the response variables to improve predictive accuracy [3]. Note also that, in contrast to the Wiener filter, the additional prewhitening of $\mathbf{y}$ makes CCA invariant with regard to scaling of $\mathbf{x}, \mathbf{y}$.

Analogously to standard multivariate regression, CCA can also directly be formulated as a linear least squares problem. An iterative (online) CCA algorithm based on the such a formulation is described in [2].

3 A Fast AAM Search Based on CCA

In the standard AAM search algorithm, a linear function (see (3)) is used to map texture residuals (difference images) $\mathbf{r}(\mathbf{p}) \in \mathbb{R}^p$ to corresponding parameter displacements $\delta\mathbf{p} \in \mathbb{R}^q$ (approximating $\mathbf{r}(\mathbf{p})$ by a first-order Taylor series expansion). In our algorithm, we extract linear features of $\mathbf{r}(\mathbf{p})$ by CCA of $\mathbf{r}(\mathbf{p})$ and $\mathbf{p}$.

3.1 CCA-AAM Training

During training, instead of computing $\mathbf{R}$ by numeric differentiation, we create training data for CCA. More precisely, the training data is generated as follows: Given the original training images that were used to build the AAM, for each training image, we generate a set of synthetic images by perturbing the optimal AAM match, i.e., $\mathbf{r}(\mathbf{p}_{opt} + \delta\mathbf{p})$, where the optimum parameter vector $\mathbf{p}_{opt}$ is obtained by mapping the training image texture and shape into the model eigenspace and the components of $\delta\mathbf{p}$ are randomly drawn from uniform distributions from -1 to +1 standard deviation. An overall number of m residual vectors with m corresponding parameter displacement vectors is obtained. We denote the set of random displacement vectors by $\mathbf{P} \in \mathbb{R}^{q\times m}$ and the set of corresponding texture residuals by $\mathbf{G} \in \mathbb{R}^{p\times m}$.

Applying CCA to these two data sets yields empirical canonical factors pairs $\mathbf{W}_g = (\mathbf{w}_g^1, \ldots, \mathbf{w}_g^{k^*})$ and $\mathbf{W}_p = (\mathbf{w}_p^1, \ldots, \mathbf{w}_p^{k^*})$, respectively, where $i = 1k^* \leq k$.

These are the (derived) linear combinations which are best predicted by $\mathbf{r}$. By discarding directions with low canonical correlation, i.e., those variates which are poorly predicted by $\mathbf{r}$, we expect to improve overall predictive accuracy and robustness against noise [3] (see also Section 1). The optimal number of factors k^* is estimated from a separate validation set. After employing CCA, we perform regression on the leading canonical projections $\mathbf{G}_{proj} = \mathbf{W}_g^T\mathbf{G}$ and $\mathbf{P}$. These projections are then used to compute the $p \times k^*$ transformation matrix $\mathbf{l} = \mathbf{P}\mathbf{G}_{proj}^{\dagger}$, where $\mathbf{G}_{proj}^{\dagger} = (\mathbf{G}_{proj}^T\mathbf{G}_{proj})^{-1}\mathbf{G}_{proj}^T$.

(a) (b)

Fig. 1. Types of data used for evaluation. (a) Face images and (b) hand x-rays of metacarpal bones.

3.2 CCA-AAM Search

During search, a new displacement prediction has to be obtained at each iteration. Instead of using (3), the prediction $\delta\mathbf{p}_{predicted}$ can be obtained as $\delta\mathbf{p}_{predicted} = \mathbf{l}\mathbf{r}_{proj}$, where $\mathbf{r}_{proj} = \mathbf{W}_g^T\mathbf{r}_{current}$.

As $\mathbf{R}_{cca} = \mathbf{l}\mathbf{W}_g^T$ can be precomputed during training, the final formulation of the prediction function is

$$\delta\mathbf{p}_{predicted}(\mathbf{r}_{current}) = \mathbf{R}_{cca}\mathbf{r}_{current}, \tag{4}$$

allowing for an AAM search utilizing the correlations between parameter displacement and image difference as captured by CCA. The computation of these predictions is as fast as for the standard approach, therefore, one step of an iteration of CCA search is as fast as one step using (3). For practical application, incremental PCA can be used to lower memory requirements. An appealing side-effect is that overfitting is avoided, which would otherwise be accomplished using regularization techniques for CCA [20]. The outlined approach can also be applied to other kinds of image features (e.g., gradient images or edges).

4 Experiments

4.1 Setup

Experiments were conducted on 36 face images [22] and 36 metacarpal bone images manually annotated by a medical expert (Fig. 1). The algorithm was evaluated using 4-fold cross validation. Following the standard AAM training scheme, a set of difference images and corresponding parameter displacements were obtained by randomly perturbing the AAM modes in the interval -1 to +1 standard deviation. While the calculation of $\mathbf{R}$ (cf. (3)) by numerical differentiation requires separate variation of each AAM mode, CCA-AAM training allows simultaneous variation of all modes.

To compare search performance in both cases, AAM search was performed on the test data using varying lengths of $\mathbf{s}_{steps}$. Scaling factors available during search are chosen by using the first y elements of $\mathbf{s}_{steps}$. As a performance measure, we use the total number of steps accumulated over all iterations (cf. Section 2.2).

Searches were initialized using equal initialization (randomly generated by translations of up to 10 pixels and mean shape and texture) for both approaches. The data for each of the result bars plotted was provided by 180 search results.

4.2 Faster Training

CCA-AAM training needs fewer synthetic difference images. Using 24 modes for face data and 18 for bone data, 6,480 synthetic face images and 4,860 synthetic bone images were generated for standard training. For CCA training, no improvement could be observed when using more than 200 synthetic difference images. Thus, although the computation of the CCA is expensive, training is still considerably faster than standard training. For a Matlab implementation on a PowerMac G5 1.8GHz, a speed-up factor of 4.9 and 3.5 was achieved.

4.3 Faster Convergence with Equal Accuracy

In Fig. 2, the mean landmark error (point to point distance) over the corresponding number of overall search steps until conver-

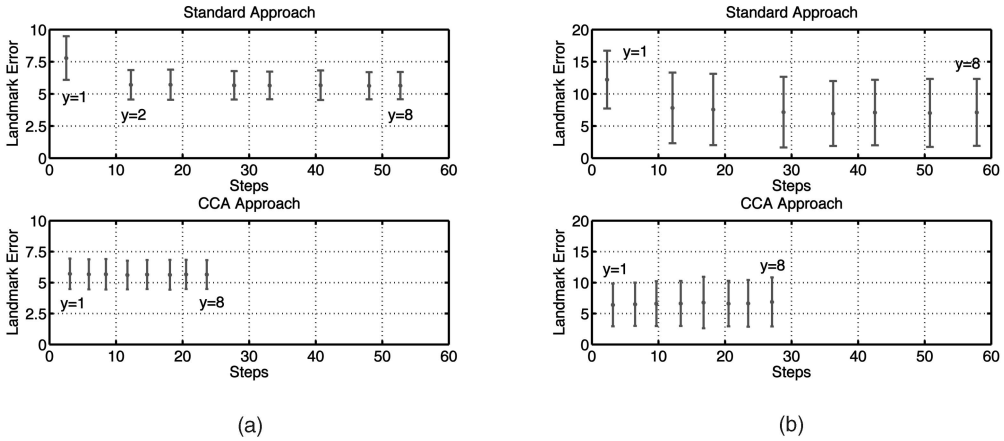

Fig. 2. Comparison of landmark errors. The eight bars correspond to y, i.e., the length of s_{steps}, ranging from 1 to 8, from left to right. Note how the CCA yields better (bones) or equal (faces) results faster (at $\approx$ 3 steps) than the standard approach (at $\approx$ 12 steps). (a) Faces and (b) metacarpal bones.

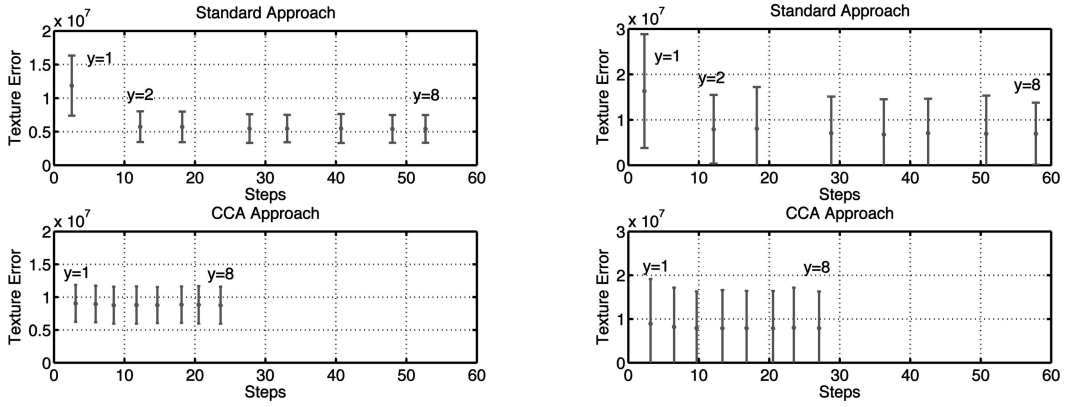

Fig. 3. Comparison of texture errors. The eight bars correspond to y (length of s_{steps}), ranging from 1 to 8, from left to right. Again, the CCA approach yields its best results already at $y = 1$ at $\approx$ 3 steps while the standard approach needs $\approx$ 12 steps for equal error levels. (a) Faces and (b) metacarpal bones.

gence is depicted. Error bars are one standard deviation. The eight results plotted correspond to y ranging from 1 to 8 as stated above.

In contrast to full rank of 24 (faces) and 18 (bones), CCA employed ranks of 10 and 9, respectively. It can be observed that the CCA convergence speed with almost equal final accuracy is considerably better than the one of the standard approach. The time for an iteration is dominated by the warping in each of the texture comparison steps. The calculation of the parameter prediction ((4) or (3), respectively) amounts for only 9 percent of the calculation time for a single step. We thus utilize the necessary number of steps as the distinctive measure of convergence speed.

Already with $\mathbf{s}_{steps} = \langle 1 \rangle$, i.e., no scaling of $\delta\mathbf{p}$, the CCA approach yields its best result in 3.07 steps for the faces data and 3.16 steps for the metacarpals, respectively. The standard approach needs at least $\mathbf{s}_{steps} = \langle 1, 0.5 \rangle$ to perform equally well, requiring 12.23 and 12.11 steps. The CCA approach is thus 3.98 and 3.83 times faster. The mean texture errors in Fig. 3 show a similar picture. Again, the CCA approach yields its best results already at $y = 1$. The standard approach is dependent on the availability of further scaling factors ($\mathbf{s}_{steps} = \langle 1, 0.5 \rangle$) to equal this performance. The results are summarized in Table 1.

4.4 Influence of Rank Reduction

In a separate experiment, the influence of rank reduction by CCA was investigated. In Fig. 4, the dependency of the mean landmark errors after search convergence is depicted for rank k set to $1, 4, \ldots, 24$ for the face data set. It can be seen that, for $k = 7$, the search yields the lowest landmark errors, and, for $k = 13$, the lowest texture errors. The number of necessary steps is lower than for full rank in both cases.

5 Conclusion

CCA-AAMs introduce a search algorithm based on *canonical correlation analysis (CCA)*. CCA efficiently models the dependencies between image residuals and parameter correction. Taking advantage of the correlations between these two signal spaces, CCA makes sensible rank reduction possible. It accounts for noise in the training data and thereby yields significant improvements of the AAM search performance in comparison to the standard search approach. After computing CCA, linear regression is performed on a small number of linear features which leads to a more accurate parameter prediction during search, eliminating the need for the expensive variable step size search scheme employed in the standard approach.

Empirical evaluation on two data sets shows that the CCA-AAM search approach is up to four times faster than the standard approach. As fewer training samples are needed, training is up to five times faster. Our approach can be adopted in most of the existing extensions of the original AAM search approach based on linear regression.

Future research will focus on the use of nonlinear CCA (Kernel CCA) and the application of CCA to obtain more compact and more descriptive active appearance models.

TABLE 1
Result Summary

Faces	Standard	CCA
Training samples	6480	200
Mean landmark error	5.7	5.7
Mean texture error ($\cdot 10^6$)	5.7	9.1
Necessary search steps	12.23	3.07
Search speed-up	1.00	**3.98**

Bones	Standard	CCA
Training samples	4860	200
Mean landmark error	7.8	6.4
Mean texture error ($\cdot 10^6$)	7.9	8.9
Necessary search steps	12.11	3.16
Search speed-up	1.00	**3.83**

Mean landmark and texture errors and the corresponding number of search steps for both data sets.

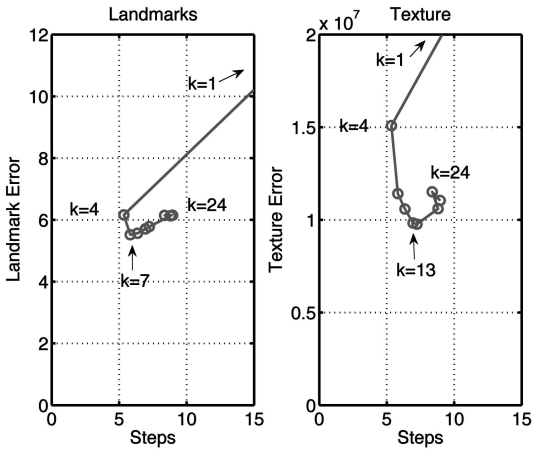

Fig. 4. Influence of CCA regression rank. Mean landmark error against the number of steps for different choices of k (number of factors for CCA regression) for the face data set.

ACKNOWLEDGMENTS

This work has been supported by the Austrian Science Fund under Grant P17083-N04 (AAMIR). Part of this work has been carried out as part of the K-plus Competence center ADVANCED COMPUTER VISION funded under the K-plus program. René Donner and Michael Reiter both contributed equally to this work. The authors are grateful to Thomas Melzer for his valuable comments.

REFERENCES

[1] R. Beichel, G. Gotschuli, E. Sorantin, F. Leberl, and M. Sonka, "Diaphragm Dome Surface Segmentation in CT Data Sets: A 3D Active Appearance Model Approach," *SPIE: Medical Imaging,* M. Sonka and J. Fitzpatrick, eds., vol. 4684, pp. 475-484, 2002.

[2] M. Borga, "Learning Multidimensional Signal Processing," ser. Linkoping Studies in Science and Technology, Dissertations, No. 531, Dept. of Electrical Eng., Linkoping Univ., Linkoping, Sweden, 1998.

[3] L. Breiman and J.H. Friedman, "Predicting Multivariate Responses in Multiple Linear Regression," *J. Royal Statistical Soc., Series A,* vol. 59, no. 1, pp. 3-54, 1997.

[4] T. Cootes and P. Kittipanya-ngam, "Comparing Variations on the Active Appearance Model Algorithm," *Proc. British Machine Vision Conf.,* vol. 2, pp. 837-846, 2002.

[5] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision (ECCV),* pp. 484-498, 1998, citeseer.nj.nec.com/cootes98active.html.

[6] T. Cootes, G. Edwards, and C. Taylor, "A Comperative Evaluation of Active Appearance Model Algorithms," *Proc. British Machine Vision Conf.,* vol. 2, pp. 680-689, 1998.

[7] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001, citeseer.nj.nec. com/cootes98active.html.

[8] K. Diamantaras and K. Hornik, "Noisy Principal Component Analysis," *Proc. Conf. Measurement '93,* J. Volaufova and V. Witkowsky, eds., pp. 25-33, May 1993.

[9] K.I. Diamantaras and S. Kung, *Principal Component Neural Networks.* John Wiley & Sons, 1996.

[10] F. Dornaika and J. Ahlberg, "Efficient Active Appearance Model for Real-Time Head and Facial Feature Tracking," *Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures (AMFG '03),* pp. 173-180, 2003.

[11] G. Edwards, C. Taylor, and T. Cootes, "Interpreting Face Images Using Active Appearance Models," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 300-305, 1998.

[12] R. Gross, I. Matthews, and S. Baker, "Generic vs. Person Specific Active Appearance Models," *Image and Vision Computing,* vol. 23, no. 11, pp. 1080-1093, Nov. 2005.

[13] A. Höskuldsson, "PLS Regression Methods," *J. Chemometrics,* vol. 2, pp. 211-228, 1988.

[14] H. Hotelling, "Relations between Two Sets of Variates," *Biometrika,* vol. 8, pp. 321-377, 1936.

[15] X. Hou, S. Li, H. Zhang, and Q. Cheng, "Direct Appearance Models," *Proc. Computer Vision and Pattern Recognition Conf.,* vol. 1, pp. 828-833, 2001.

[16] Y. Hua, M. Nikpour, and P. Stoica, "Optimal Reduced-Rank Estimation and Filtering," *IEEE Trans. Signal Processing,* vol. 49, no. 3, pp. 457-469, Mar. 2001.

[17] N. Johnson, A. Galata, and D. Hogg, "The Acquisition and Use of Interaction Behaviour Models," *Proc. IEEE Computer Vision and Pattern Recognition (CVPR '98),* pp. 866-871, 1998.

[18] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 674-679, 1981.

[19] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 135-164, Nov. 2004.

[20] T. Melzer, M. Reiter, and H. Bischof, "Appearance Models Based on Kernel Canonical Correlation Analysis," *Pattern Recognition,* vol. 39, no. 9, pp. 1961-1973, 2003.

[21] S.C. Mitchell, J.G. Bosch, B.P.F. Lelieveldt, R.J. van der Geest, J.H.C. Reiber, and M. Sonka, "3-D Active Appearance Models: Segmentation of Cardiac MR and Ultrasound Images," *IEEE Trans. Medical Imaging,* vol. 21, no. 9, pp. 1167-1178, 2002.

[22] M.M. Nordstrøm, M. Larsen, J. Sierakowski, and M.B. Stegmann, "The IMM Face Database—An Annotated Dataset of 240 Face Images," technical report, Informatics and Mathematical Modelling, Technical Univ. of Denmark, DTU, May 2004.

[23] A. Pezeshki, L.L. Scharf, M.R. Azimi-Sadjadi, and Y. Hua, "Underwater Target Classification Using Canonical Correlations," *Proc. MTS/IEEE Oceans Conf.,* 2003.

[24] M.B. Stegmann, "Generative Interpretation of Medical Images," master's thesis, Technical Univ. of Denmark, 2004.

[25] M.B. Stegmann, H. Olafsdottir, and H.B.W. Larsson, "Unsupervised Motion-Compensation of Multi-Slice Cardiac Perfusion MRI," *Medical Image Analysis,* vol. 9, no. 4, pp. 394-410, Aug. 2005.