An Active Appearance Model with a Derivative-Free Optimization

Jixia ZHANG[‡], Franck DAVOINE[†], Chunhong PAN[‡]
LIAMA Sino-French IT Lab.
CNRS[†], Institute of Automation of the Chinese Academy of Sciences[‡]
95, Zhongguancun Dong Lu, PO Box 2728 − Beijing 100190 − PR China
Franck.Davoine@gmail.com

Abstract

A new AAM *matching algorithm based on a derivative free optimization is presented and evaluated in this paper. For this, we use an efficient model-based optimization algorithm, the so called* NEWUOA *algorithm from M.J.D. Powell. We compare the new matching method performances against the standard one based on a fixed Jacobian matrix learned from a training set, and show significant improvements in terms of accuracy.*

1 Introduction

Since the seminal work of Edwards, Cootes and Taylor [7], Active Appearance Models (AAM) have been widely used for modeling and registering visual deformable objects, in applications ranging from medical imaging [18] to facial behavior recognition [9]. In the context of non-rigid face modeling, AAMs have been employed to encode in a compact way the variations in face appearance across shape and texture. Numerous extensions have been proposed to the original formulation [4] to make the AAM more accurate, more robust to occlusions or pose variations, or even to reduce the complexity of the learning or the matching processes. The models, learned from a set of suitably annotated examples, are matched to previously unseen faces by updating the model during search, using essentially two kinds of approaches.

The first one focuses on regression-relevant feature representations [20, 6, 16]. In the original AAM fitting algorithm [7], the error between the input image and the closest model instance (the texture residual vector $\mathbf{r}$) with regard to the model parameter updates $\delta\mathbf{p}$ was modeled by a linear regression approach. The method in [20] suggests that the shape can be estimated directly from the texture using a linear regression, considering they are sufficiently correlated. More recently, an efficient multivariate regressor from Haar-like features to the AAM parameter update space, learned through a boosting procedure has been proposed in [16]. This method boasts significant improvements in terms of accuracy. A fast algorithm was proposed in [6], based on canonical correlation analysis. This method learns a linear regression between the canonical projections of the texture residuals and random model parameter displacements. The authors show that, because of its better robustness to noise, the CCA provides more accurate parameter updates, leading to a reduced number of iterations in the matching procedure.

The second class of approaches treat the matching as a non linear optimization problem [3], and the usual way to drive the parameter fitting algorithm is to use a form of gradient descent. Cootes *et al.* proposed a simplified Gauss-Newton procedure with a fixed Jacobian matrix (and thus a constant updating matrix) computed offline by numeric differentiation on typical facial images. The fixedness of the updating matrix is justified assuming that the error surface around the true minimum can be reasonably well approximated by a quadratic function. And it has been widely shown that this method allows efficient matching to take place, even if there are several parameters to evaluate. Extensions to this method have been introduced, at the expense of computing time, by updating the Jacobian as the search progresses [2, 5]. A Gauss-Newton optimization method has been proposed in [1] by Baker and Matthews, introducing the inverse compositional image alignment algorithm. By reversing the role of the image and the model in the error function to minimize, they demonstrate that the Jacobian matrix, to a first order, becomes constant throughout the fitting algorithm. The resulting updating matrix is analytically derived from the cost function itself and not numerically estimated from a training set. This results in an improvement of the fitting accuracy and a reduction of parameter updatings compared to the basic version of the AAM. In [17], Baker *et al.* propose extensions: the project-out and the simultaneous inverse compositional approaches. In the first one, the shape parameters are iteratively updated, and the texture parameters are then estimated in a single step. This algorithm deals only with very subtle appearance variations. In the second one, the shape and texture parameters are estimated simultaneously, resulting in a fairly slow but more accurate algorithm.

Our motivation in this paper is to evaluate the efficiency of one of the best state-of-the-art derivative free optimization methods, the so called NEWUOA (NEW Unconstrained Optimization Algorithm) proposed by M.J.D. Powell [14], in the framework of AAM matching. This model-based trust region optimizer has been demonstrated as a very powerful tool to estimate the minimum of noisy, smooth and piecewise smooth objective functions; it outperformed geometry-based methods like the Nelder-Mead (or downhill simplex) algorithm and a pattern-search one, in many cases by a wide margin [11]. Furthermore this low-complexity optimizer performs very well in a recent comparison of model-based methods [13], and it has been successfully used in the context of rigid, linear and piecewise linear registration of medical images [19].

In order to motivate our approach, we compare its performances against the basic AAM based on a simplified Gauss-Newton procedure and usually considered as a good baseline indicator for evaluating new matching methods. Performances are compared in terms of accuracy on a variety of datasets.

2 Methods

We introduce in the two next subsections a brief overview of the AAM's parametrization and construction, and introduce the NEWUOA optimization algorithm.

2.1 Basic AAM search

A statistical appearance model describes an object of a predefined class by simultaneously encoding its intrinsic variation in shape and texture. In the context of active appearance models for face analysis [3], assuming that the shape and the texture spaces follow a Gaussian probability distribution, the facial appearance variability is linearly modeled with a Principal Component Analysis (PCA). By shape, we mean a set of landmark points sampled on the main contours and features of the face. By texture, we mean the pattern of intensities enclosed within the convex hull of the shape.

First, the N facial textures of all the training original faces are warped to the Procrustes mean shape, yielding a new set of N shape-free textures. A PCA is then applied to shape and shape-free texture data, denoted respectively by $\mathbf{s}$ and $\mathbf{g}$: each training face is so represented by shape and texture model parameters $\mathbf{b}_s$, $\mathbf{b}_g$:

$$\mathbf{s} = \mathbf{s}_m + \phi_s \mathbf{b}_s \qquad \mathbf{g} = \mathbf{g}_m + \phi_g \mathbf{b}_g$$

where $\mathbf{s}_m$, $\mathbf{g}_m$ are respectively the mean shape and texture (suitably normalized), ϕ_s, ϕ_g are the eigenvectors of shape and shape-free texture covariance matrices. A third PCA is then performed on a concatenated shape and texture parameters $\mathbf{b}$, to obtain a combined model vector $\mathbf{c}$ given by $\mathbf{b} = \phi_c \mathbf{c}$, where:

$$\mathbf{b} = \begin{pmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix}$$

$\mathbf{W}_s$ is a diagonal scaling matrix between shape and texture and ϕ_c is a set of eigenvectors.

From the combined appearance model vector $\mathbf{c}$, a new instance of shape and shape-free texture can be generated:

$$\mathbf{s}_{model}(\mathbf{c}) = \mathbf{s}_m + \mathbf{Q}_s \mathbf{c} \qquad \mathbf{g}_{model}(\mathbf{c}) = \mathbf{g}_m + \mathbf{Q}_g \mathbf{c}.$$

where matrices $\mathbf{Q}_s$ and $\mathbf{Q}_g$ describe the modes of variations derived from the training set.

In order to match a target face in a given image, the shape and texture instances have to be translated, scaled and rotated. This similarity transformation can be represented by a vector of four parameters $\mathbf{p} = (t_x, t_y, \alpha, \theta)$. Those parameters denote respectively the x and y translation of the shape, and the scaling factor and inplane rotation relatively to the learnt mean shape.

The AAM paradigm provides an iterative simplified Gauss-Newton search technique in order to compute automatically the pose and appearance parameters $(\hat{\mathbf{p}}, \hat{\mathbf{c}})$ that best approximate the target face in the image [3]. The minimized criterion is the quadratic norm $\mathbf{E}$ of the residual vector

$$\mathbf{r}(\mathbf{p}, \mathbf{c}) = \mathbf{g}_{model}(\mathbf{c}) - \mathbf{g}_{im}(\mathbf{p}, \mathbf{c}) \qquad (1)$$

where $\mathbf{g}_{im}(\mathbf{p}, \mathbf{c})$ denotes the image texture sampled at the hypothesized pose $\mathbf{p}$ and shape $\mathbf{s}_{model}(\mathbf{c})$. Starting from an initial guess $(\check{\mathbf{p}}, \check{\mathbf{c}})$, it is shown for example in [5] that the corrections $(\delta\mathbf{p}, \delta\mathbf{c})$ to apply to the parameters so as to minimize the related norm of the residual vector are given by:

$$\delta\mathbf{p} = \mathbf{R}_\mathbf{p}.\mathbf{r}(\check{\mathbf{p}}, \check{\mathbf{c}}) \qquad \delta\mathbf{c} = \mathbf{R}_\mathbf{c}.\mathbf{r}(\check{\mathbf{p}}, \check{\mathbf{c}}). \qquad (2)$$

where the two constant matrices (with † denoting the pseudoinverse):

$$\mathbf{R}_\mathbf{p} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{p}}\right)^\dagger \qquad \mathbf{R}_\mathbf{c} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right)^\dagger \qquad (3)$$

are precomputed offline by numeric differentiation from training data, systematically displacing each parameter from the known optimal value on typical images and averaging over the training set. Based on equations (2), the iterative model refinement procedure is described in [3]. Each iteration updates the AAM parameters by gradually projecting the current texture residual vector onto the constant matrices $\mathbf{R}_\mathbf{p}$ and $\mathbf{R}_\mathbf{c}$ until the norm of the residual vector does not decrease anymore.

2.2 Derivative-free optimization

We consider one of the best state-of-the-art derivative free optimization methods, the so called NEWUOA proposed by M.J.D. Powell in 2004. This trust-based method is fully described in [14]. We summarize here the main lines of the algorithm.

Let us denote a scalar valued objective function $\mathbf{E}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$, and consider the following unconstrained optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{E}(\mathbf{x})$$

To solve this problem, the algorithm interpolates m points of $\mathbf{E}(\mathbf{x})$ by means of a quadratic model $\mathbf{Q}$ within trust-regions in an iterative way and finds its minimum. Each iteration k affects at most one of the interpolation points. Let consider the set Y_k composed of m interpolation points $\mathbf{y}_k^j$, $(j = 1, \ldots, m)$ and that, at the beginning of the k^{th} iteration, no point in Y_k has a lower objective function value than $\mathbf{x}_k$ so far, and $\mathbf{x}_k \in Y_k$. The quadratic model at this time is given by:

$$\mathbf{Q}_k(\mathbf{x}_k + \mathbf{d}) = \mathbf{E}(\mathbf{x}_k) + \mathbf{d}^T \mathbf{g}_k + \frac{1}{2}\mathbf{d}^T \mathbf{H}_k \mathbf{d}, \quad \mathbf{d} \in \mathbb{R}^n$$

where $\mathbf{g}_k$ is a real vector and $\mathbf{H}_k = \nabla^2 \mathbf{Q}_k$ is a $n \times n$ symmetric matrix. $\mathbf{g}_k$ and $\mathbf{H}_k$ are determined by imposing the m following interpolation conditions:

$$\mathbf{Q}_k(\mathbf{y}_k^j) = \mathbf{E}(\mathbf{y}_k^j), \quad j = 1, \ldots, m$$

The points in Y_k must satisfy some geometric properties to remain sufficiently linearly independent, in order to avoid degeneracies [14, 8]. Once $\mathbf{Q}_k$ has been identified, a step $\mathbf{d}_k$ has to be added to $\mathbf{x}_k$ by solving the trust-region subproblem:

$$\min_{\mathbf{d}} \mathbf{Q}_k(\mathbf{x}_k + \mathbf{d}), \quad \text{subject to } \|\mathbf{d}\| \leq \Delta_k$$

for some trust-region radius $\Delta_k \geq \rho_k > 0$. Δ_k will be further reduced at most iterations until it reaches ρ_k (see [15] for the details). The parameter ρ_k, $k = 1, 2, 3, \ldots$, will decrease gradually from ρ_{ini} to $\rho_{fin} > \rho_{ini}$ only when the constraint $\Delta_k \geq \rho_k$ prevents further reductions in $\mathbf{E}$. ρ_{ini} and ρ_{fin} are respectively a user-defined initial and final radius. The algorithm finishes when the trust-region radius reaches the lower bound ρ_{fin} that fixes the final accuracy.
Furthermore, Powell recommends $m = 2n + 1$ (instead of the general $m = \frac{1}{2}(n+1)(n+2)$), which only increases as n, the number of variables. This is made possible, when updating the quadratic model, by minimizing the Frobenius norm of the change in $\mathbf{H}$ [15].

3 A new AAM search

Our approach follows the basic AAM *training* procedure [4]. The basic AAM *matching* is done in an iterative way until convergence, by projecting the current texture residual image onto a constant matrix in order to update the displacements in the parameters. Because the residual image is measured in a normalized shape-free domain, the error surface around the true minimum is supposed to be reasonably well approximated by a quadratic function. Instead of using a simplified Gauss-Newton optimization, such property motivates us to use a quadratic interpolation based optimization algorithm to match the AAM, starting from a pose $\mathbf{p}_0$ within an initial trust region around the target, and with $\mathbf{c}_0 = 0$. For this purpose, we use the NEWUOA algorithm to minimize the norm of the residual image (the objective function $\mathbf{E}$) w.r.t. the AAM parameters $\mathbf{p}$ and $\mathbf{c}$ without derivatives.

4 Experiments

In this section, we compare the NEWUOA matching method performances against the standard one based on a fixed Jacobian matrix learned from a training set.

Table 1: Three AAM models.

	Model 1	*Model 2*	*Model 3*
Training database	C-K	IMM	IMM
# Training images	338	105	200
Shape size	49	49	58
Shape-free texture size	13377	30884	31596
% of Shape variation	95	95	95
# Shape modes	27	14	19
% of Texture variation	95	95	95
# Texture modes	113	68	117
% of Combined variation	95	80	95
# Combined modes	54	12	63
Test database	IMM	IMM	FGnet
# Test images	37	5	879

We use the Cohn-Kanade (C-K) facial expression [10] and IMM [12] databases to train three models summarized in table 1. We select annotated monocular images of 40 different human faces in the IMM database, with different poses and expressions and precluding the images with a spot light added at the person's left side. In the C-K database, we select monocular images of 46 different human faces depicting one to six basic low- and high-magnitude facial expressions. Experimentations with the NEWUOA algorithm are performed using the Fortran code provided by Powell [14].

Figure 1: Search and reconstruction results on one test image. Top line, from left to right: target face, basic AAM matching, AAM shape. Bottom line: NEWUOA matching, NEWUOA shape.

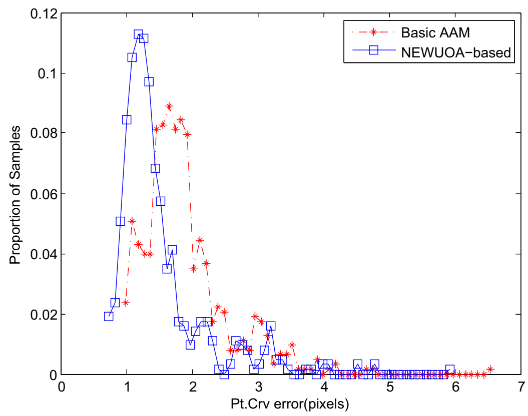

Figure 2: Distribution of boundary errors averaged on 37 test images: convergence accuracy histogram.

First, we use the *Model 1* (Table 1) and 37 images of IMM to test the accuracy. For each image we displace the model from the known best position by a range of displacements in x, then do the AAM searches. The distribution of the point-to-curve errors is shown in Figure 2. Our approach exhibits larger proportions of samples at the lower end of the error range. Figure 1 shows the reconstructions of one face for an initial displacement in x of 20% of the shape width. In terms of complexity, in derivative-free optimization, the performance of the algorithms is usually measured by the number of objective function evaluations until convergence. The complexity of one iteration of the NEWUOA algorithm is attractive compared to others and is of order of only $O(n^3)$ (only in the worst case, otherwise $O(n^2)$) when m is linear in n.
Both the basic AAM and the NEWUOA matching algorithms are iterative, and several tries are done during

each iteration. In the case of Figure 2, 15 (resp. 451) objective function evaluations are performed during the AAM (resp. NEWUOA) searches (results averaged on 37 images). Future work will consist in considering adaptive trust-region radius to reduce the complexity of the NEWUOA search. For testing the performance of complex displacement, we use the *Model 2* (Table 1) and displace the model by a range of displacement in both x and y directions. The boundary error is shown in Figure 3. The NEWUOA matching method leads to a much more stable result.

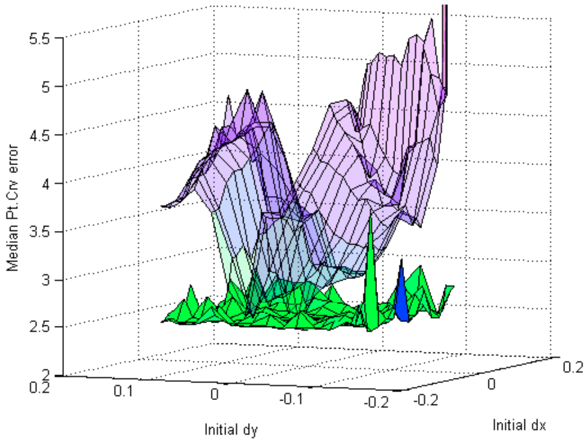

Figure 3: Median boundary error vs. initial displacement in x and y directions in the range of $\pm 12\%$ of the face shape width. The upper (resp. lower) surface shows the error of basic AAM (resp. NEWUOA) matching.

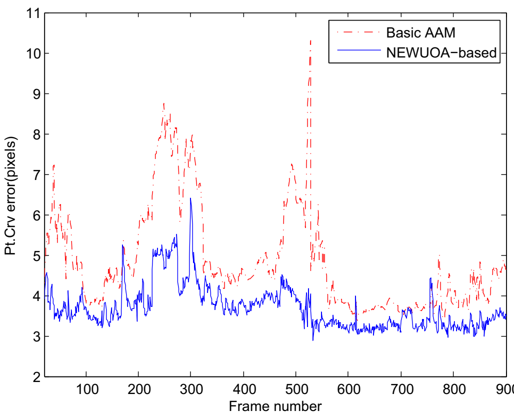

Figure 4: Boundary error against frame number fitted on a talking face.

Model 3 (Table 1) is tested on the FGnet video sequence[1], depicting a person engaged in a natural conversation. On each successive frame, the basic AAM and NEWUOA matching algorithms were initialized at the previously detected pose and respectively at $\mathbf{c} = \check{\mathbf{c}}$ (the best appearance in the previous frame) and $\mathbf{c} = 0$. Improvements in terms of boundary errors between the groundtruth and reconstructed shapes are shown in Figure 4.

5 Conclusions

We have presented an accurate AAM matching method, using a recent state-of-the-art unconstrained derivative-free optimizer. Results have been compared to the standard approach based on a simplified Gauss-Newton optimization, and lead to more accurate estimates of the AAM parameters. Future work will address the complexity of the algorithm w.r.t. trust-region radius, and its relevance to build accurate and high-resolution person-specific appearance models.

[1] www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html

References

[1] Simon Baker and Iain Matthews. Equivalence and efficiency of image alignment algorithms. pages 1090–1097, December 2001.

[2] A. Batur and M. Hayes. Adaptive active appearance models. *IEEE Transactions on Image Processing*, 14(11), 2005.

[3] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Anylisis and Machine intelligence*, 23(6), 2001.

[4] T.F. Cootes and P. Kittipanya-ngam. Comparing variations on the active appearance model algorithm. In *BMVC*, pages 837–846, 2002.

[5] T.F. Cootes and C.J. Taylor. An algorithm for tuning an active appearance model to new data. In *BMVC*, pages 919–928, 2006.

[6] René Donner, Michael Reiter, Langs Georg, Philipp Peloschek, and Horst Bischof. Fast active appearance model search using canonical correlation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), October 2005.

[7] G.J. Edwards, C.J. Taylor, and T.F. Cootes. Interpreting face images using active appearance models. In *Intl. Conference on Automatic Face & Gesture Recognition*, pages 300–305, 1998.

[8] G. Fasano, J.L. Morales, and Jorge Nocedal. On the geometry phase in model-based algorithms for derivative-free optimization. *Optimization Methods and Software*, 24(1), February 2009.

[9] C. Hack and C. Taylor. Modelling 'talking head' behaviour. In *BMVC*, pages 609–618, 2003.

[10] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Intl. Conference on Automatic Face & Gesture Recognition*, pages 46–53, 2000.

[11] Jorge J. Moré and Stefan M. Wild. Benchmarking derivative-free optimization algorithms. Technical Report ANL/MCS-P1471-1207, Argonne National Laboratory, Illinois, U.S.A., April 2008.

[12] 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 University of Denmark, may 2004.

[13] R. Oeuvray. *Trust-Region Methods Based on Radial Basis Functions with Application to Biomedical Imaging*. PhD thesis, EPFL, Lausanne, Switzerland, 2005.

[14] M.J.D. Powell. *Large-Scale Nonlinear Optimization*, chapter The NEWUOA software for unconstrained optimization without derivatives, pages 255–297. Springer, G. Di Pillo and M. Roma (eds.), 2006.

[15] M.J.D. Powell. Developments of NEWUOA for minimization without derivatives. *IMA Journal of Numerical Analysis*, 28(4):649–664, 2008.

[16] Jason Saragih and Roland Goecke. A nonlinear discriminative approach to AAM fitting. In *ICCV*, pages 1733–1738, 2007.

[17] Ralph Gross Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Pittsburgh, PA, November 2003.

[18] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME - a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319–1331, 2003.

[19] N. Wiest-Daesslé, S. Prima, S.P. Morrissey, and C. Barillot. Validation of a new optimisation algorithm for registration tasks in medical imaging. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Washington, USA, April 2007.

[20] H. Zhang Q. Cheng X. Hou, S.Z. Li. Direct appearance models. In *CVPR*, pages 828–833, 2001.