On Improvements of Active Appearance Models

Teodor Mandziy

Department of Computational Methods and Systems for Information Transformation
Lviv Phisico-Mechanical Institute of The National Academy of Science of Ukraine
teodor_mandziy@ipm.lviv.ua

Abstract

In this paper a nonlinear optimization technique for active appearance models is proposed. This approach allows one to reduce the number of optimization procedure iteration steps and to slightly decrease computational complexity of singular optimization step. Experimental result are obtained and compared to performance of conventional active appearance models.

Introduction

Interpretation of images containing objects whose shape and texture can vary is a very necessary and challenging task. As the examples of such objects can be images of human faces or MR brain sections. Models called to deal with such type of object images have very wide practical application. They are able to handle various types of tasks such as object appearance modeling, objects tracking, feature localization, image segmentation etc. They also can be used for objects state recognition like recognition of human emotions. Those models are irreplaceable for tasks of image segmentation where correct segmentation is impossible without relying on shape information due to its low signal to noise ratio or texture inconsistency. Such models application is very important and successful in medicine and industry for radiographic, magnetic resonance and ultrasonic images segmentation.

There were proposed many models to handle aforementioned tasks. Probably the first attempt to deal with images of varying objects was active contours [1]. They are able to track changes in shape but they are shape free models and initially have not been able to respond to particular shapes. There exist some improvements of the model to make it sensitive to particular user defined shapes [2, 3].

Proposed by T. Cootes Active Shape Models (ASM) (also known as smart snakes) [3] was design especially for handling shape variations of varying objects images. The ASM is relying on statistical model of shape variation. Shape in this model is represented as a set of landmarks, relative variation of which are constrained by a Point Distribution Models (PDM) captured from a training set of labeled images. Matching model to image is made by iterative technique which tries to obtain a new landmark point location by nearby search around current landmark point location aiming to find best texture model match expected at the landmark position. After new landmark points locations are found parameters of a model are adjusted to the best match of these new locations to model generated ones. Since T. Cootes original paper there were made a lot of effort to improve the ASM: double contours ASM [4], ASM with bifurcation contours handling [5], non-linear multi-view ASM [6] etc.

Base concepts of an ASM were developed by T. Cootes into Active Appearance Models (AAM) [7]. The AAM deals with the whole model of appearance, which includes shape and texture variations modeling. It allows one to generate the whole image of modeled object and its variations based on different model parameters.

The main difference between active shape models and active appearance models is that first tries to minimize distance between model generated set of landmarks and points founded on the image which best match texture patterns of the model and the second minimize difference between the given image and synthetic image generated by the model. From original Cootes paper on AAM there were made many improvements and adaptations to conventional AAM: Shape-AAM [8], Inverse-Compositional AAM [9], non-linear texture features AAM [10], AAM with occlusion [11], dense AAM [12].

The following describes the basics of the active appearance models, gives an introduction to existing optimization techniques and compares them to proposed nonlinear optimization technique based on artificial neural networks. Performance comparison is based on accuracy of models outcomes and time taken by models to tune model parameters and to converge.

1. Introduction to Active Appearance Models

1.1. Active Appearance Model formulation

Active Appearance Models are powerful tool for matching statistical models of appearance to new images. Statistical model of object appearance can be constructed using previously properly annotated set of training images. Image annotation implies locating significant feature points on each given training image of modeled object and marking these points by a set of landmark points. Given a set of landmark points $X = \{X_i\}$, where:

$$X_i = [x_1^i, \dots x_n^i, y_1^i, \dots y_n^i]$$

and set of training texture vectors $G = \{G_i\}$ obtained based on shape information from the training images one can construct statistical models of both shape and texture variations by applying Principal Component Analysis (PCA). The mathematical formulation of the texture and shape models, which use very few parameters, can be written as follows:

$$x = \bar{x} + P_s c_s$$
$$g = \bar{g} + P_g c_g \tag{1}$$

where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean texture, P_s and P_g are models of shape and texture variations respectively and c_s , c_g are parameters of generated shape and texture respectively.

Due to the possible correlation between shape and texture variations we should apply further PCA to form final combined model of appearance. PCA is applied to new training set formed by shape and texture parameters of annotated training images:

$$b = \begin{pmatrix} W_s c_s \\ c_g \end{pmatrix}, \tag{2}$$

where W_s is the normalization matrix required for handling of difference in units of shape and texture models. Resulting statistical models of shape x and texture g with a set of parameters c can be written as follows:

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \tag{3}$$

where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean texture and Q_s , Q_g are the matrixes which models shape and texture variations respectively.

1.2 Model optimization to a new data

To match a new image with previously built model one should find a proper model parameters c for models of shape and texture and additionally apply suitable global transformations such as rotation, scale and translation to model shape x .

To find a suitable set of the model parameters AMM use an iterative optimization procedure aiming to minimize difference between given image and synthetic image generated by the model

$$E(p) = |r(p)|^2 = r^T r \ , \tag{4}$$

where $E(p)$ is the function of error, p contains complete set of parameters (combined model parameters c and global pose parameters) and $r = r(p)$ is a function given a overall parameters vector p returning residual of difference between model synthesized image and actual image covered by a synthesized image.

It was proposed by T. Cootes to search for optimal model's parameters using only the information contained in the residual $r = r(p)$ in the following form:

$$dp = Rr \ . \tag{5}$$

To derive matrix R we perform first order approximation (based on a Taylor series) of $r = r(p)$ near p :

$$r(p + dp) = r + Jdp \ , \tag{6}$$

where J is the Jacobian of $r(p)$. Substituting p for $p + dp$ in (4) we obtain the next:

$$\begin{aligned} E(p + dp) &= (Jdp + r)^T (Jdp + r) = \\ &= dp^T J^T Jdp + 2dp^T J^T r + C \end{aligned} \tag{7}$$

where C is a constant. Differentiation of (7) with respect to p gives a gradient:

$$\frac{\partial E}{\partial p} = 2(J^T Jdp + J^T r) \ . \tag{8}$$

To find an extremum of E the $\frac{\partial E}{\partial p}$ should be equated to zero which gives the following equation:

$$J^T Jdp = -J^T r) \ . \tag{9}$$

By solving above equation we can derive the displacement dp directing the movement of overall model parameters p to achieve the local minimum of the error function $E(p)$. To make AAM computationally efficient the assumption about constant Jacobian $J = J_0$, which can be estimated from a training set, is made. The final model parameters update rule is the following:

$$dp = -Rr \ , \tag{10}$$

where R is the pseudo inverse of J_0 :

$$R = (J_0^T J_0)^{-1} J_0 \ . \tag{11}$$

For the purpose of linear transformation R estimations also a multivariate linear regression model can be used [13].

2. Nonlinear optimization of an Active Appearance Models

Described above method for model optimization to a new data is linear and also an assumption about constant matrix R is made which implies a constant Jacobian, thus it can have a poor performance caused by fixed linear approximation of nonlinear nature of given optimization problem.

Bataur and Hayes in [14] showed that fixed Jacobian could be inefficient approximation for some cases and proposed to use adaptive Jacobian that is a linear function of current model parameters. Proposed by them linear dependence leads to more robust and accurate performance, especially for objects images with highly varying textures (e.g. changing lightning condition for face images). Following [14] Cootes and Taylor in [15] estimate Jacobian based on set of residuals estimated during optimization steps. Such an improvement showed better performance with respect to errors in model fitting to new images compared to conventional AMM.

2.1 Optimization in model parameters feature space

Recall that conventional AMM aims to minimize the difference between given input image and image synthesized by previously trained model:

$$\delta g = g_m - g_o \,, \tag{12}$$

where I_m is the texture generated by obtained model and I_o is the texture of given image. The obtained residual is used by linear model (5) to tune model parameters for better image-model matching.

Unlike a conventional AAM here aim is to minimize not the difference δg between the model synthesized image and input image but δc_g, which is a residual of difference between model parameters of texture and texture parameters of that model computed from projection of original image into a model parameter feature space:

$$\delta c_g = c_m - c_o = r \,, \tag{13}$$

where c_m and c_o are the texture parameters of the model and parameters computed by projecting original image into feature space.

2.2 Optimization in model parameters feature space

To enhance matching capabilities of an AMM instead of the linear model (5) we use nonlinear relationship to predict step dp for overall models parameters and residual r, which in our case looks like (14):

$$dp = T(r) \,. \tag{14}$$

In this work nonlinear relation (14) is modeled by means of a multi-layer perceptron. Neural networks recommended themselves as powerful tool for data classification and function approximation. As know from [16] multilayer feed forward network is an universal approximator. Feeding δc_g to neural network one should not worry about any preprocessing steps (e.g. usually raw data are processed by principal component analysis to decorrelate components of input vector) on data that usually forego neural network training. It is because δc_g already contains decorrelated components due to it was composed by difference of two texture vectors projected onto principal component axes of texture model.

In conventional AMM in order to compute residual r the synthetic image should be generated by the model using current model parameters. Utilization of residuals, computed in feature space, for optimization of the model to new data replace generation of model image by operation of original image spanned by the model in current state projection to model feature space:

$$c_0 = P_g^T\left(g_0 - \overline{g}\right). \tag{15}$$

This approach allows one to reduce drastically dimension of r term in (5) and consequently reduce the computational cost of the model.

3. Experimental results

The proposed improvements of Active appearance models were tested on IMM Face Database that contains 240 annotated images of human faces.

After training of an AAM, with proposed improvement, result was compared to results produced by conventional AAM. Comparison showed that in case where proposed and conventional models are converged in proper way accuracy of proposed model after convergence in general is comparable with conventional one. An advantage of the proposed model consists in better response to bad initialization of the model. It shows more consistent results when initial position of the AAM on a given image is considerably displaced or rotated with respect to true position of an object on the given image. It can be seen from results, given on figure, that optimization algorithm of conventional AAM sticks somewhere in local minimum and consequently fails to converge to desirable model parameters that closely respond to true object image. As shown on figure, in contrast to conventional AAM, proposed nonlinear optimization technique better respond to grater displacement of initial model and as a result it can converge in cases where conventional AAM fails.

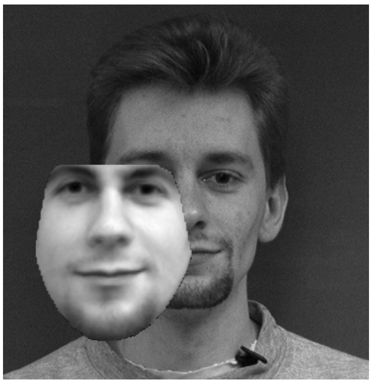
To achieve better response to greater displacements the proposed model require more training examples in training set, which represent such disturbance, compared to conventional AAM. This is caused by much poorer representation of image differences provided by projection of them into model parameter feature space.

Dimensionality of texture vectors, the model operates with, depends on modeled object, complexity of an object texture and required precision of an outcome model and can be high which influence the computational cost of an AAM model optimization during matching it with new image data. By using residual of image difference in feature space one can reduce the dimension of this residual sufficiently due to much lower dimension of feature space compared to original space. For comparatively small dimensions of texture vectors conventional model can be more computationally efficient compared to proposed model. But for comparatively high dimensions of texture vectors proposed model becomes less computationally expensive. It is due to handling high dimensional residuals can take more operation compared to operation required for neural network computation.

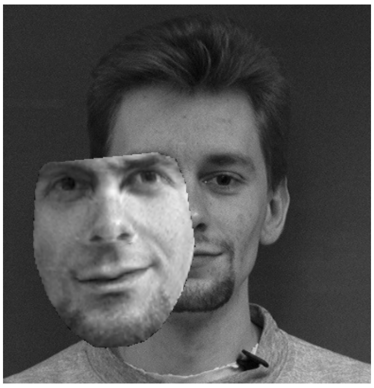
4. Conclusions

As a result of proposed here improvements of an AAM the model is able to better respond to bigger displacements of an initial AAM position compared to conventional AAM. But such an advantage is made to the prejudice of computation complexity of training procedure. Though in practice we mostly interested in complexity of final model, which is improved in this case.

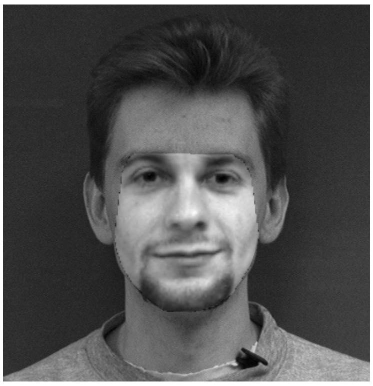
Working with images difference residuals one potentially can reduce computational cost. This dimensionality reduction induce much poorer representation of image differences, thus it demands more complex models for the prediction of model displacement to be made.



a)



b)



c)

Figure: Comparative results of a conventional AAM and AAM with proposed improvements: *a)* initialization of the model; b) result of conventional AAM performance; c) performance result of improved AAM.

5. References

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models." *International Journal of Computer* Vision., *Vol. 1,Num. 4, pp. 321-331, 1987.*

[2] Junmo Kim, Müjdat Çetin, Alan S. Willsky: Nonparametric shape priors for active contour-based image segmentation. *Signal Processing* 87(12): 3021-3044 (2007)

[3] F. Lecellier, S. Jehan-Besson , M.J. Fadili , G. Aubert, M. Revenu , E. Saloux, "*Region-based active contours with noise and shape priors*" , IEEE ICIP, pp. 1649-1652, Atlanta, USA, 2006

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.

[4] M. Seise, S. J. McKenna, I.W. Ricketts, and C. A.Wigderowitz, “Double contour active shape models,” in *British Machine Vision Conference*, vol. 2, 2005, pp. 159–168.

[5] M. Seise, Stephen J. McKenna, Ian W. Ricketts, C. A. Wigderowitz: Learning *Active Shape Models for Bifurcating Contours*. IEEE Trans. Med. Imaging 26(5): 666-677 (2007)

[6] S. Romdhani, S. Gong, and A. Psarrou, *“A Multi-View Nonlinear Active Shape Model Using Kernel PCA,” Proc. British Machine Vision Conf.,* pp. 483-492, 1999.

[7] T. Cootes, G. J. Edwards, and C. J. Taylor. *Active appearance models.* In H.Burkhardt and B. Neumann, editors, 5th *European Conference on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. *A comparative evaluation of active appearance model algorithms.* In P. Lewis and M. Nixon, editors, *9th British Machine Vision Conference*, volume 2, pages 680–689, Southampton, UK, Sept. 1998. BMVA Press.

[9] Matthews and S. Baker. *Active appearance models revisited*. International Journal of Computer Vision, 60(2):135 – 164, November 2004.

[10] I.M.Scott, T.F.Cootes, and C.J.Taylor. *Improving appearance model matching using local image structure.* In 18*th Conference on Information Processing in Medical Imaging*, pages 258–269. Springer-Verlag, 2003.

[11] R. Gross, I. Matthews, and S. Baker. *Constructing and fitting active appearance models with occlusion*. In *Proceedings of the IEEE Workshop on Face Processing in Video*, June 2004.

[12] K. Ramnath, S. Baker, I. Matthews, and D. Ramanan, *"Increasing the Density of Active Appearance Models",* in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[13]A. Bataur and M. Hayes. *Adaptive active appearance models*. IEEE Trans. Imaging Processing, 14:1707–21, 2005.

[13] T. F. Cootes and C. J Taylor. *Statistical Models of Appearance for Computer Vision*. Tech. Report, University of Manchester, Feb. 2000

[14] A. Bataur and M. Hayes. *Adaptive active appearance models.* IEEE Trans. Imaging Processing, 14:1707–21, 2005.

[15] T.F.Cootes and C.J.Taylor, *"An Algorithm for Tuning an Active Appearance Model to New Data"*, Proc. British Machine Vision Conference, Vol. 3, pp.919-928, 2006

[16] Hornik, K., Stinchcombe, M., & White, H. (1989). *Multilayer feedforward networks are universal approximators*. Neural Networks, 2, 359-366