Resolution-Aware Fitting of Active Appearance Models to Low-Resolution Images

Problem

Active Appearance Models (AAM) are compact representations of the shape and appearance of objects. [Cootes et al, '98]. Fitting AAMs to images is a difficult, nonlinear optimization task. Traditional approaches can fit well to high-resolution images, but they degrade quickly at lower resolutions. We diagnose why this is the case, and propose a remedy.

Background

An AAM has two components: shape and appearance. Each is a linear, Principal Components model learned via training.

The shape is defined by a set of landmarks.

In addition to non-rigid shape modes, four special basis shapes account for similarity transforms.

The appearance component is shape-normalized.

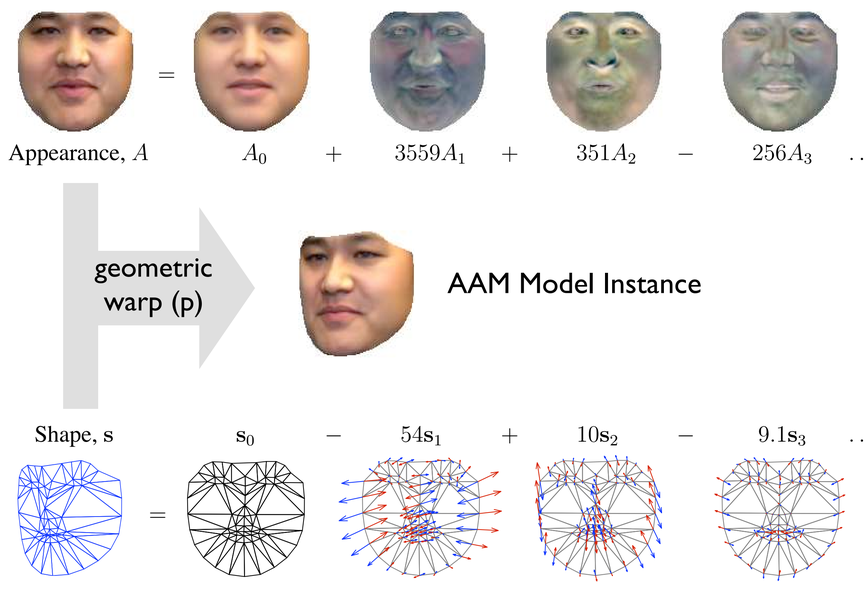
Fitting algorithms aim to recover those parameters which *best* explain a given image.

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \ldots, x_v, y_v)^{\mathrm{T}}$$

$$\mathbf{s}(\mathbf{p}) = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i \quad \text{where} \quad \mathbf{p} = (p_1, p_2, \ldots, p_n)$$

$$A(\mathbf{x}; \boldsymbol{\lambda}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) \qquad \forall\, \mathbf{x} \in \mathbf{s}_0$$

$$\text{where} \quad \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$$



from [Matthews and Baker, '04]

Traditional Formulation

Fit the AAM by minimizing the L2 norm error between the model instance and the input image warped onto the model coordinate frame.

Objective function:
AAMR-SIM

$$\sum_{\mathbf{x} \in \mathbf{s}_0} \Big[I\big(\mathbf{W}(\mathbf{x}; \mathbf{p})\big) - A(\mathbf{x}; \boldsymbol{\lambda}) \Big]^2$$

The "simultaneous inverse compositional" fitting algorithm [Gross et al., '05] iteratively solves for the shape and appearance updates.

Gauss-Newton gradient descent:

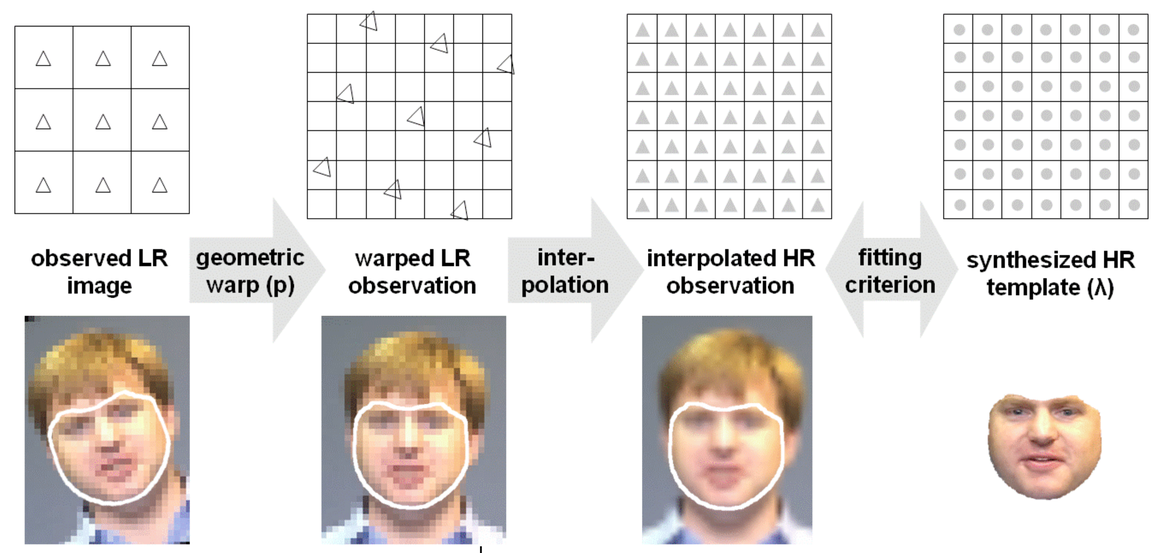
$$\begin{pmatrix} \Delta \mathbf{p} \\ \Delta \boldsymbol{\lambda} \end{pmatrix} = -H_{\mathrm{sim}}^{-1} \sum_{\mathbf{x}} \mathbf{SD}_{\mathrm{sim}}^{\mathrm{T}}(\mathbf{x}) \Big[I\big(\mathbf{W}(\mathbf{x}; \mathbf{p})\big) - A(\mathbf{x}; \boldsymbol{\lambda}) \Big]$$

$$\mathbf{SD}_{\mathrm{sim}}(\mathbf{x}) = \left(\nabla A \frac{\partial \mathbf{W}}{\partial p_1}, \ldots, \nabla A \frac{\partial \mathbf{W}}{\partial p_n}, A_1(\mathbf{x}), \ldots, A_m(\mathbf{x}) \right)$$

$$H_{\mathrm{sim}} = \sum_{\mathbf{x}} \mathbf{SD}_{\mathrm{sim}}^{\mathrm{T}}(\mathbf{x})\, \mathbf{SD}_{\mathrm{sim}}(\mathbf{x})$$

Inverse-compositional warp update:

$$\mathbf{W}(\mathbf{x}; \mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x}; \mathbf{p}) \circ \mathbf{W}(\mathbf{x}; \Delta \mathbf{p})^{-1}$$



From left to right, observed images get warped, interpolated, and finally compared against the synthesized model instance. When the input image is low in resolution, significant interpolation is needed to warp it onto the model coordinate frame.

Resolution-Aware Formulation

We explicitly account for the finite size sensing elements of digital cameras, and simultaneously model the processes of object appearance variation, geometric deformation, and image formation.

Objective function:
RAF

$$\sum_{\mathbf{u} \in I} \big[I(\mathbf{u}) - B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p}); \boldsymbol{\lambda})\big) \big]^2$$

The blur model can incorporate arbitrary cameras and point spread functions (PSF). We use the rectangular PSF:

$$B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p}); \boldsymbol{\lambda})\big) = \frac{1}{area(\mathbf{u})} \int_{\mathbf{u}' \in bin(\mathbf{u})} A\big(\mathbf{W}^{-1}(\mathbf{u}'; \mathbf{p}); \boldsymbol{\lambda}\big)\, d\mathbf{u}'$$

Discrete implementation:

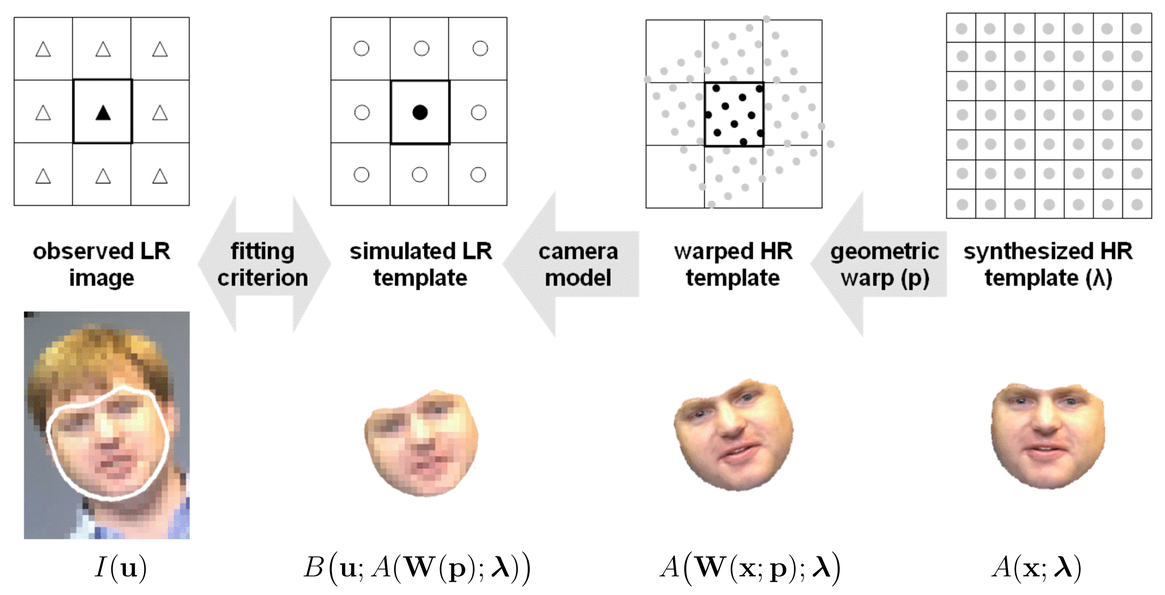
$$B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p}); \boldsymbol{\lambda})\big) = \frac{1}{area(\mathbf{u})} \sum_{\substack{\mathbf{x} \in s_0 \text{ s.t.} \\ \mathbf{u} - \left[\begin{smallmatrix} .5 \\ .5 \end{smallmatrix}\right] < \mathbf{W}(\mathbf{x}; \mathbf{p}) < \mathbf{u} + \left[\begin{smallmatrix} .5 \\ .5 \end{smallmatrix}\right]}} A(\mathbf{x}; \boldsymbol{\lambda}) \big| J\big(\mathbf{W}(\mathbf{p})\big) \big|$$

Gauss-Newton gradient descent:

$$\begin{pmatrix} \Delta \mathbf{p} \\ \Delta \boldsymbol{\lambda} \end{pmatrix} = -H_{sim}^{-1} \sum_{\mathbf{u} \in I} B\big(\mathbf{u}; \mathbf{SD}_{sim}^{\mathrm{T}}\big) \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0\big(\mathbf{W}(\mathbf{p})\big)\Big) + \sum_{i=1}^{m} \lambda_i B\Big(\mathbf{u}; A_i\big(\mathbf{W}(\mathbf{p})\big)\Big) \right]$$

$$\mathbf{SD}_{sim} = \left[\Big(\nabla A_0 + \sum_{i=1}^{m} \lambda_i \nabla A_i\Big) \frac{\partial \mathbf{W}}{\partial p_1}, \ldots, \Big(\nabla A_0 + \sum_{i=1}^{m} \lambda_i \nabla A_i\Big) \frac{\partial \mathbf{W}}{\partial p_n}, A_1\big(\mathbf{W}(\mathbf{p})\big), \ldots, A_m\big(\mathbf{W}(\mathbf{p})\big) \right]$$

$$H_{sim} = \sum_{\mathbf{u} \in I} B\big(\mathbf{u}; \mathbf{SD}_{sim}^{\mathrm{T}}\big) B\big(\mathbf{u}; \mathbf{SD}_{sim}\big)$$



The Resolution-Aware Fitting algorithm simulates the formation of low resolution images in a digital camera. In contrast to the traditional formulation, the fitting criterion is defined between observed and simulated image pixels.

Computational Comparison

- AAMR-SIM can pre-compute steepest-descent images and Hessians, leading to very efficient tracker implementations.

- RAF gives up the computational efficiency of AAMR-SIM in exchange for a more accurate/unbiased estimate of the parameters. Gradient-descent parameters need to be recomputed at every iteration.

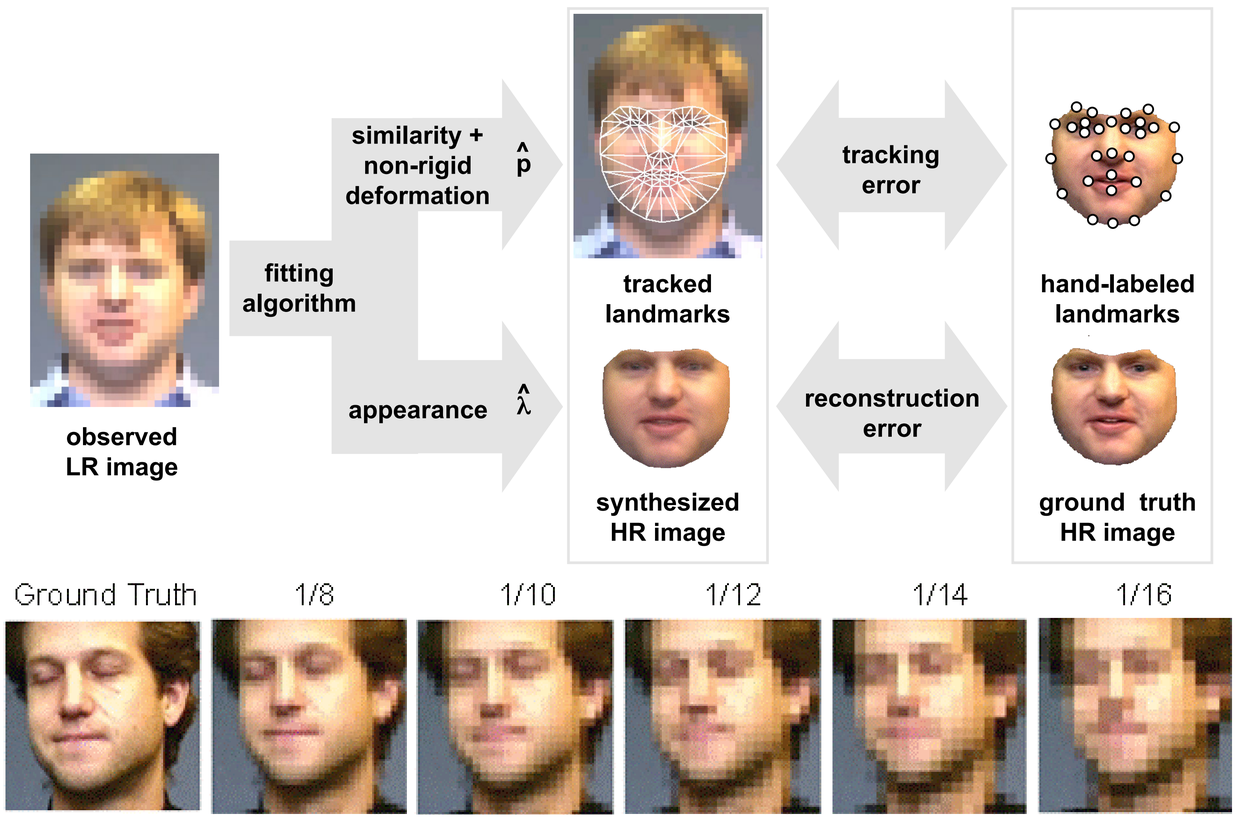
Göksel Dedeoğlu
Simon Baker
Takeo Kanade

Carnegie Mellon
THE ROBOTICS INSTITUTE

Quantitative Results

Algorithms and Accuracy Metrics

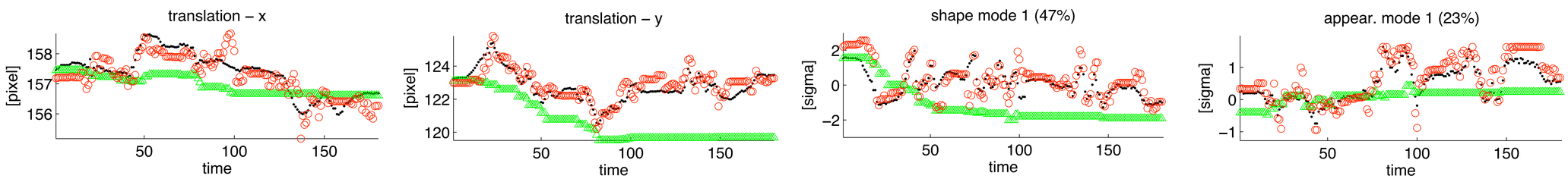
- AAMR-SIM: Traditional formulation.
- RAF: Resolution-Aware formulation with pill-box camera PSF.
- Both algorithms perform Gauss-Newton gradient-descent to estimate the shape and appearance parameters simultaneously.



Experimental setup

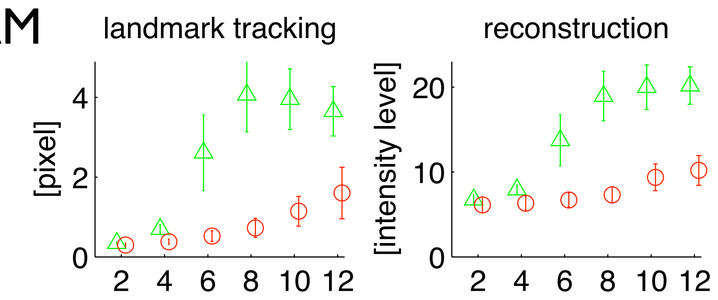
- Synthetically downsample input images at various scales.
- Initialize at t_0 with the high-resolution "ground truth" fit.
- Video sequence tracked frame-to-frame.
- Compare fitting results against "ground truth" fits.

Example parameter trajectories at scaling factor 1/10. Colors:
Ground Truth
RAF AAMR-SIM



Single-Person AAM

training set: 31 images
test set: 180 images
shape modes: 11
appearance modes: 23
face size: 100x104



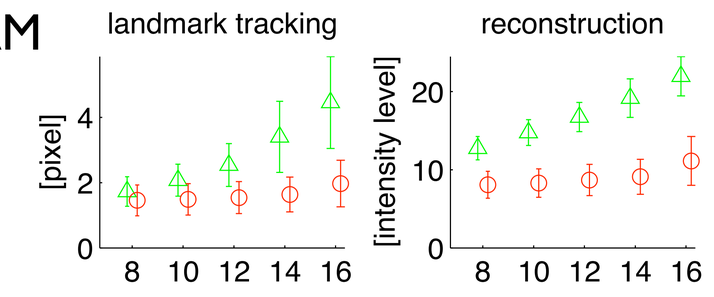
Axis Labels

X – zoom factor
Y – RMS error

△ AAMR-SIM
○ RAF

Multi-Person AAM

training set: 113 images
test set: 900 images
shape modes: 12
appearance modes: 40
face size: 190x193x3

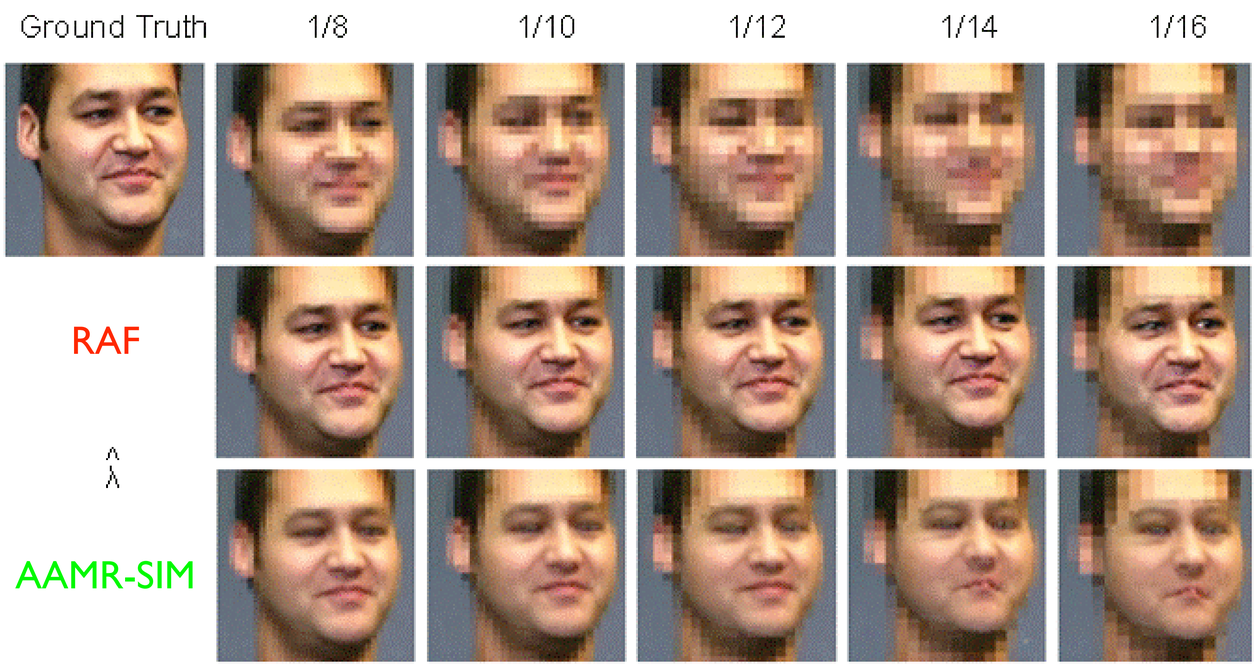


We observe substantial accuracy improvements across all metrics and variables.

Qualitative Results

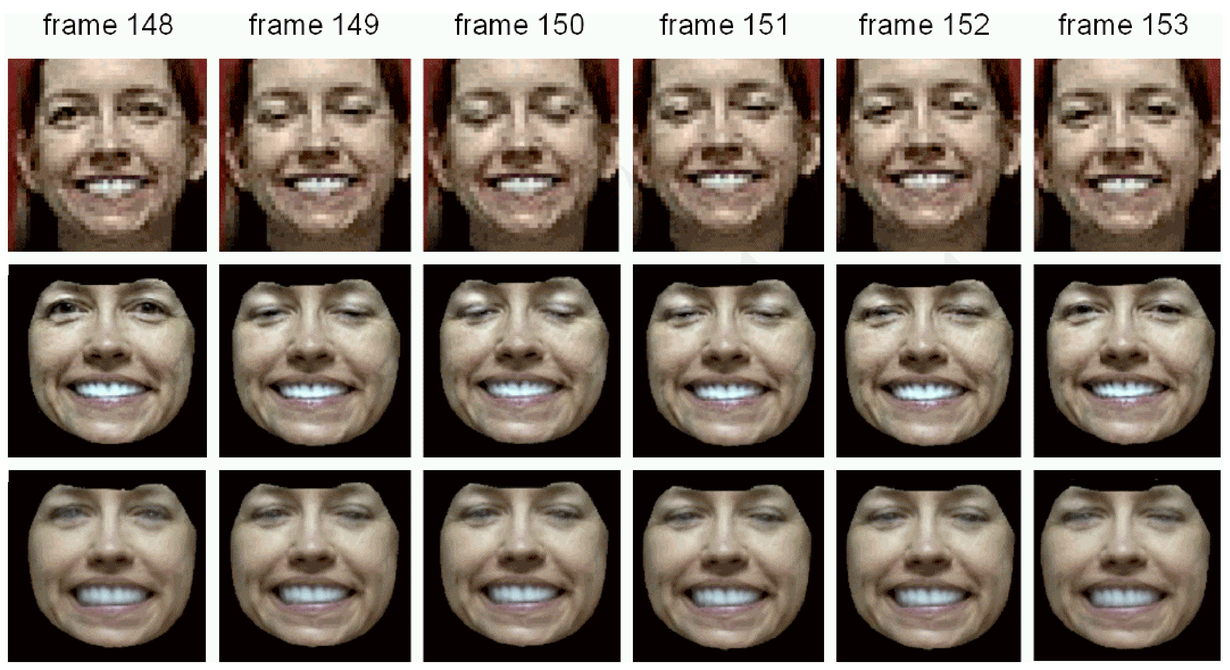
videos: http://www.cs.cmu.edu/~dedeoglu/eccv06

Example fits using synthetic low-resolution data.



Example fits using DV-compressed real low-resolution data

- Single-person AAM built at 110x114 pixel resolution.
- Shown below: eye-blink sequence, scaling ~1/3.
- RAF correctly recovers the eye-blink, whereas AAMR-SIM does not.



Resolution-Aware Fitting results remain visually detailed and realistic despite input resolution degradation.

Conclusion

- In low-resolution regimes, traditional AAM fitting algorithms suffer from an artifact of formulation. Since they rely on interpolation, their accuracy degrades quickly. The proposed Resolution-Aware Formulation avoids interpolation, and uses a camera (blur) model instead. This leads to significant accuracy improvements.

- In [Dedeoglu et al., '06], we argue that image-based warp estimation is an asymmetric problem. In the presence of relative scaling, the warp direction ought to be chosen such that the higher-resolution image gets pre-blurred and warped onto the lower-resolution one. As such, the Resolution-Aware Fitting algorithm is an application of this principle.

References

T.F. Cootes, G.J. Edwards, and C.J. Taylor: *Active Appearance Models*. Proc. of the European Conference on Computer Vision, Vol. 2, 1998, pp. 484-498.

I. Matthews and S. Baker: *Active Appearance Models Revisited*. International Journal of Computer Vision, Vol. 60, No. 2, Nov., 2004, pp. 135-164.

R. Gross, I. Matthews, and S. Baker: *Generic vs. Person-Specific Active Appearance Models*. Image and Vision Computing, Vol. 23, No. 11, Nov., 2005, pp. 1080-1093.

G. Dedeoglu, T. Kanade, and S. Baker: *The Asymmetry of Image Registration and its Application to Face Tracking*. submitted to IEEE Trans. on PAMI, under review, available as Technical Report CMU-RI-TR-06-06.

Acknowledgments

- AAM Toolbox: I. Matthews
- Discussions & comments: J. August, CMU misc-reading group, Y. Caspi
- Data collection: P. Bogdan, H. Yalçın, R. Patil, and S. Rossbach