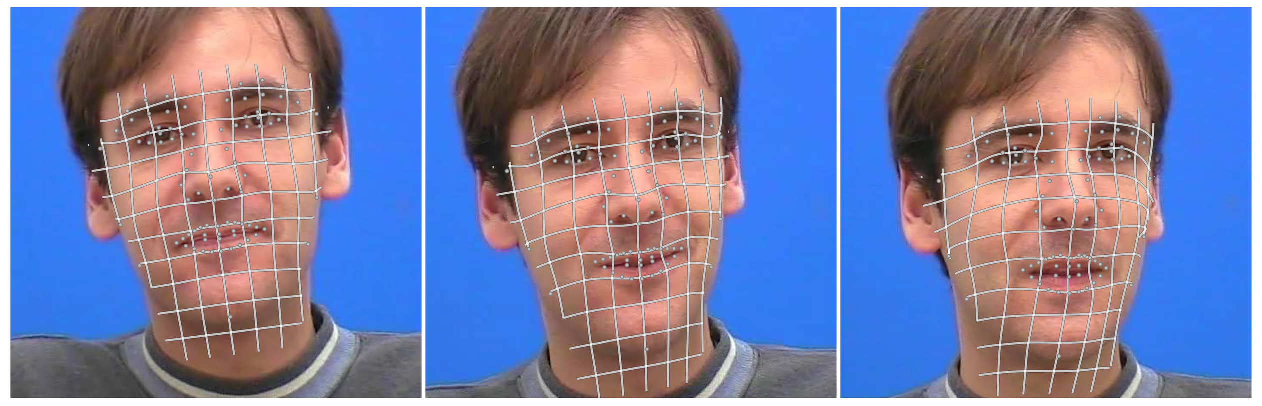


On Compositional Image Alignment

Brian Amberg, Andrew Blake, and Thomas Vetter

Brian.Amberg@unibas.ch, ab@microsoft.com, Thomas.Vetter@unibas.ch

UNI BASEL

Microsoft Research

Contribution: Fast *and* reliable Face Alignment

Inverse compositional image alignment (ICIA) is fast, but not reliable. We explain ICIA from a different perspective which leads naturally to two new algorithms with a better capture range and comparable speed.

There is no *inverse* in *ICIA*

Image aligment minimizes

$$F(\boldsymbol{q}) \triangleq \|f(\boldsymbol{q},\boldsymbol{\beta})\|^2_{\mathcal{D}}, \qquad (1)$$

$$\text{with } f(\boldsymbol{q}) \triangleq a - I \circ W(\boldsymbol{q})$$

composition with an incremental warp V approximates F around $\boldsymbol{q}_0$ as

$$F(C^{\circ}(\boldsymbol{q}_0,\boldsymbol{p})) \approx \tilde{F}(\boldsymbol{q}_0,\boldsymbol{p}) \triangleq \left\|\tilde{f}(\boldsymbol{q}_0,\boldsymbol{p})\right\|^2_{\mathcal{D}} \qquad (2)$$

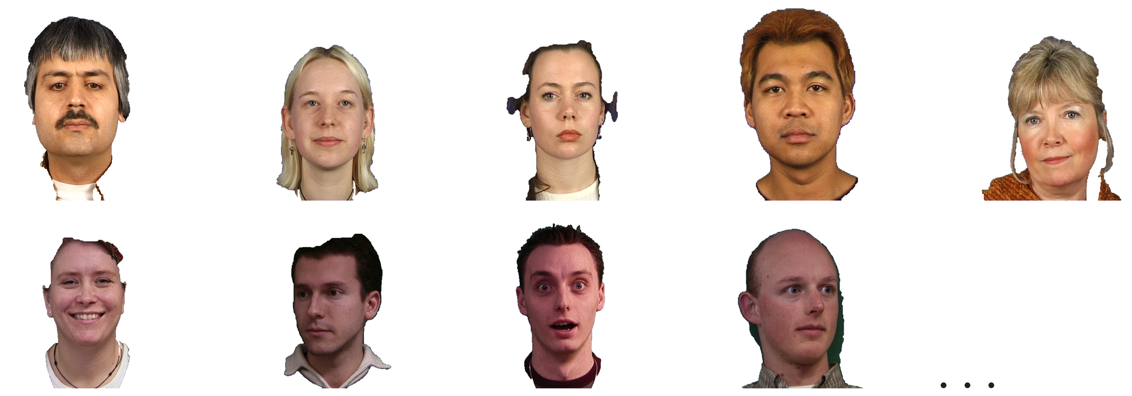
$$\text{with } \tilde{f}(\boldsymbol{q}_0,\boldsymbol{p}) \triangleq P(a - I \circ W(\boldsymbol{q}_0) \circ V(\boldsymbol{p}))$$

The gradient descent or Gauss-Newton update rule then gives an estimate of the incremental warp, which drives the model warp.

ICIA can be derived by substituting the current backwarped image with the model appearance after taking the derivative. The substitution can be used to get an approximate gradient and/or Hessian, leading to a family of algorithms.

Additionally we replace the incremental warp V with an orthonormalized warp and regularize in the composition step. The result is a vast improvement in robustness without sacrificing speed.

Training + Testing Data



The model was trained from 456 images from the IMM and XM2VTS datasets using 120 landmarks. Get the landmarks, model, and source code at:

`www.cs.unibas.ch/personen/amberg_brian/aam/`

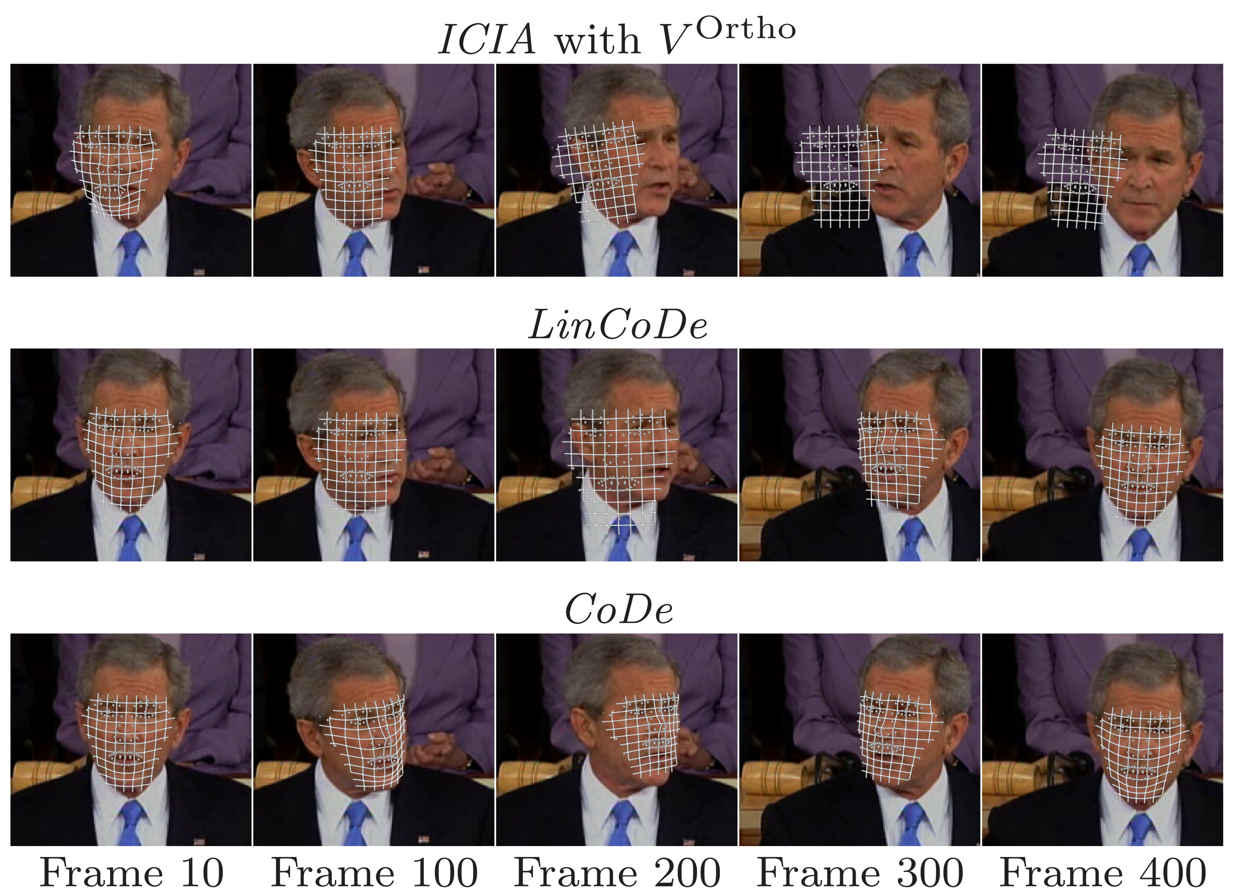
References

[1] B. Amberg, A. Blake, T. Vetter On Compositional Image Alignment with an Application to Activce Appearance Models In *CVPR'09*, 2009.

Compositional Alignment

```
for Blur and regularisation values do
    Initialize q, q_best and κ
    repeat
        Calculate ∇_p F̃(q, 0), F(q)
        if F(q) < F(q_best) then
            q_best ← q
            Increase κ
        else
            if κ smaller than threshold then
                return
            decrease κ
        Calculate p from ∇_p F̃(q_best, p) and κ
        q ← C°(q, p)
    until converged
```

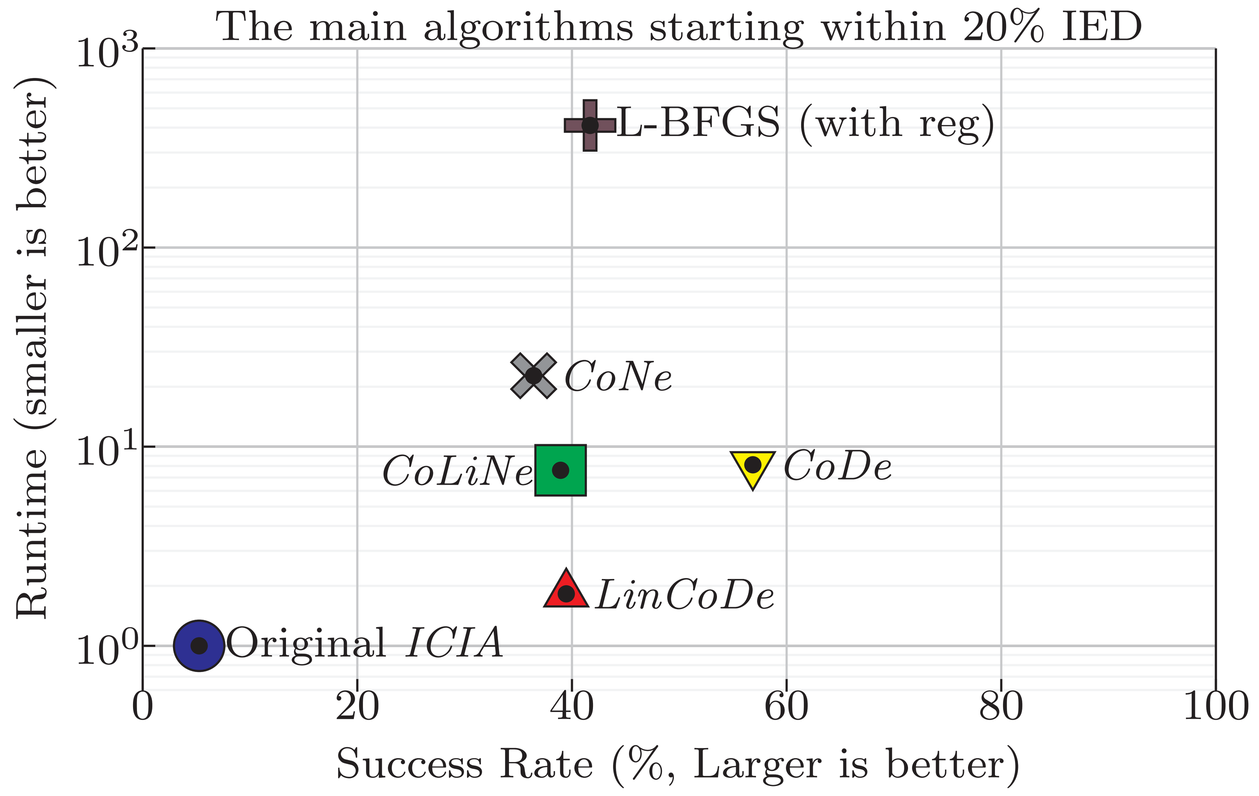
Low Res Tracking

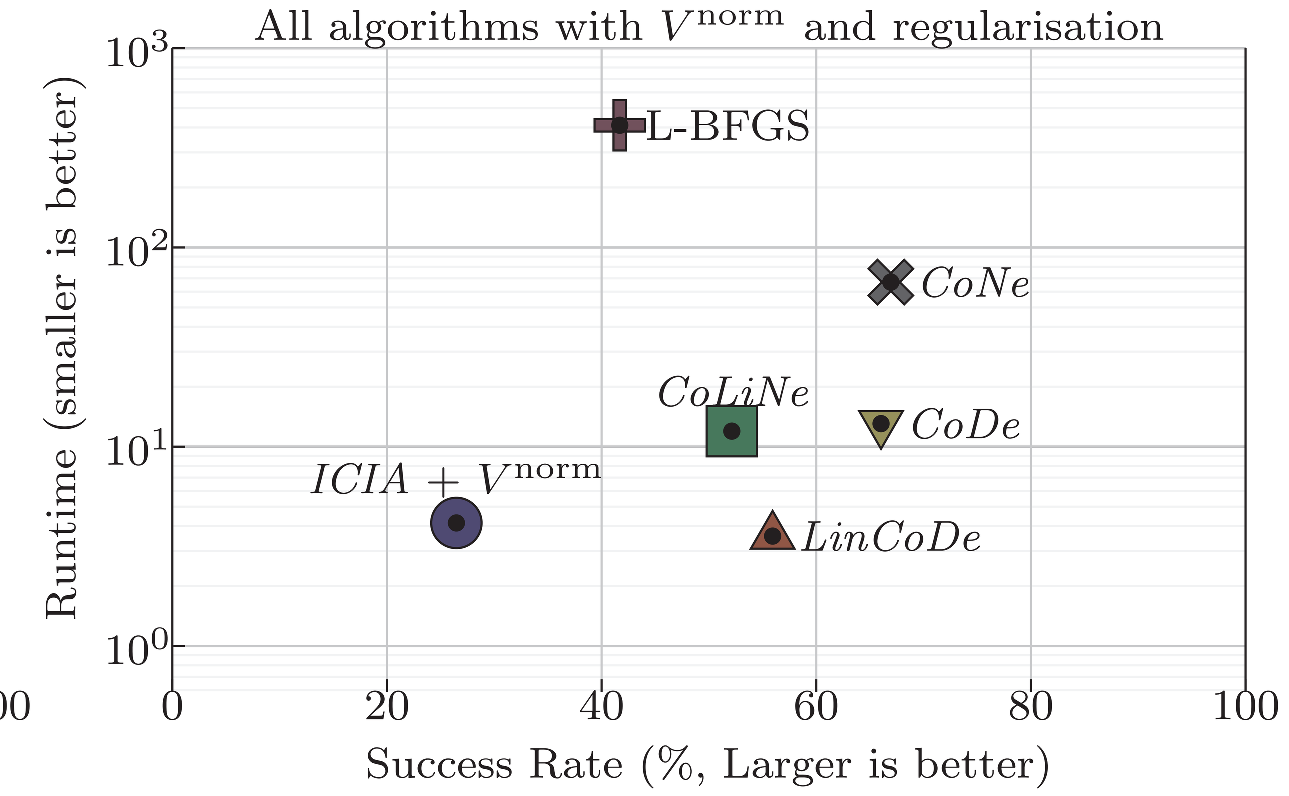


Tracking a low resolution video with large head motions succeeds with *CoDe*, where *ICIA* fails.

All methods used the orthonormal incremental warp, and relatively strong regularisation. *ICIA* starts to drift in the early frames, while *CoDe* tracks the full sequence. The approximate gradient method *LinCoDe* also suceeds, but looses track of the details for about 100 frames.

Our methods are at the performance/speed sweet point

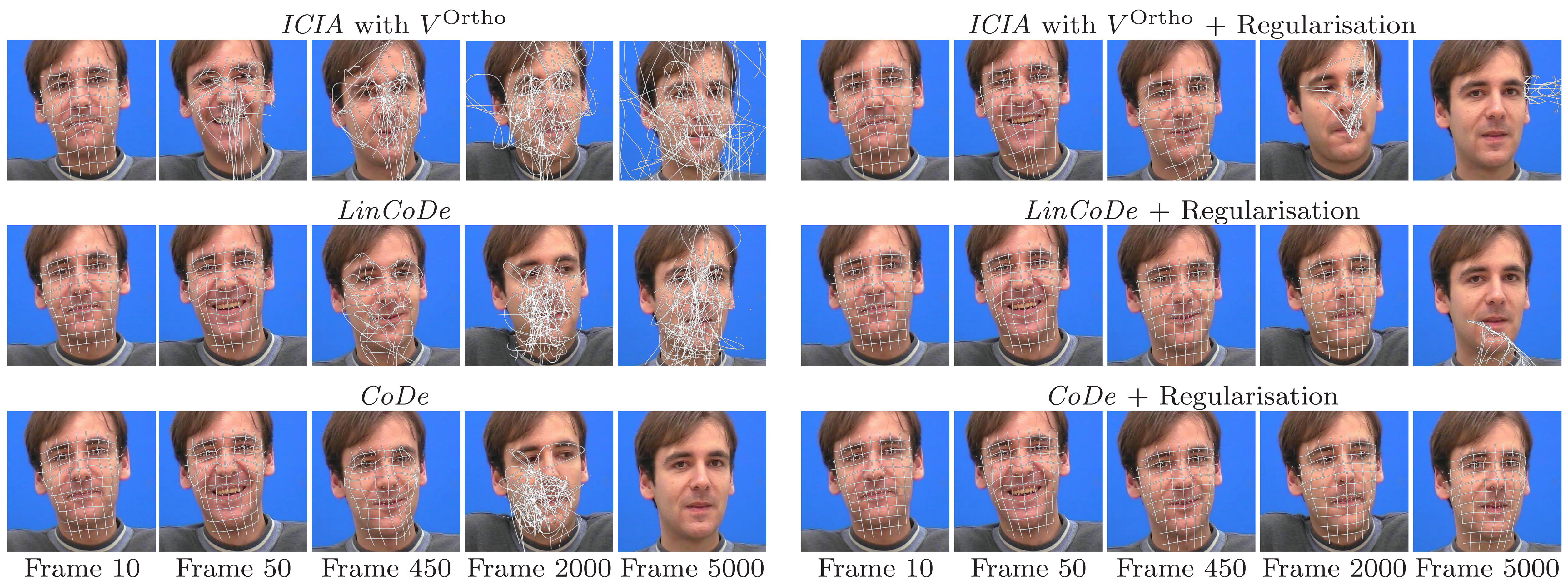




Fitting a multiperson AAM. The best speed–performance tradeoffs come from the two new algorithms *CoDe* and *LinCoDe*. Note that *ICIA* is practically useless on this difficult multi-person dataset with a success rate near zero (left). It can be improved (right) by using the orthonormal incremental warp and regularisation. The *CoDe* algorithm with regularisation (right) is as accurate as the slow, approximation-free, compositional Gauss-Newton *CoNe* method but is seven times more efficient.

The experiments were performed with leave one identity out on a mixture of two databases (XM2VTS and IMM).

Tracking 5000 frames with a general model



Our algorithm makes fast and robust tracking possible. We compare face tracking under natural motion, using *ICIA*, *LinCoDe* and *CoDe*. The original *ICIA* fails immediately with this large model and new face data. Substituting the orthonormal incremental warp for the original *ICIA* warp, the algorithm still loses track very early, whereas *LinCoDe* and *CoDe* can track much further. Finally, adding regularisation to all algorithms, *ICIA* still loses track completely after approximately 500 frames and does not recover the local deformations accurately. In contrast *CoDe* now tracks the full 5000 frame sequence without reinitialization, and *LinCoDe* tracks for 2500 frames.

The same training dataset was used for both tracking experiments. The training data was aquired with different camera and light settings from different subjects.