Learning to look at humans

Rolf P. Würtz and Thomas Walther
(Institut für Neuroinformatik, Ruhr-Universität Bochum)

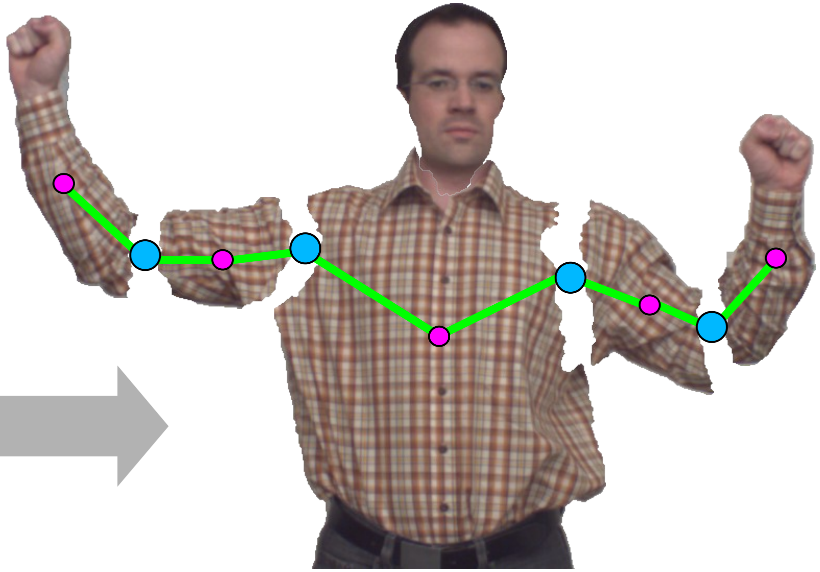
Outline

- Recalling former results
- Shape prototype formulation
- Color prototype formulation
- Integrating angular constraints
- Extended generalization tests
- Conclusion and outlook

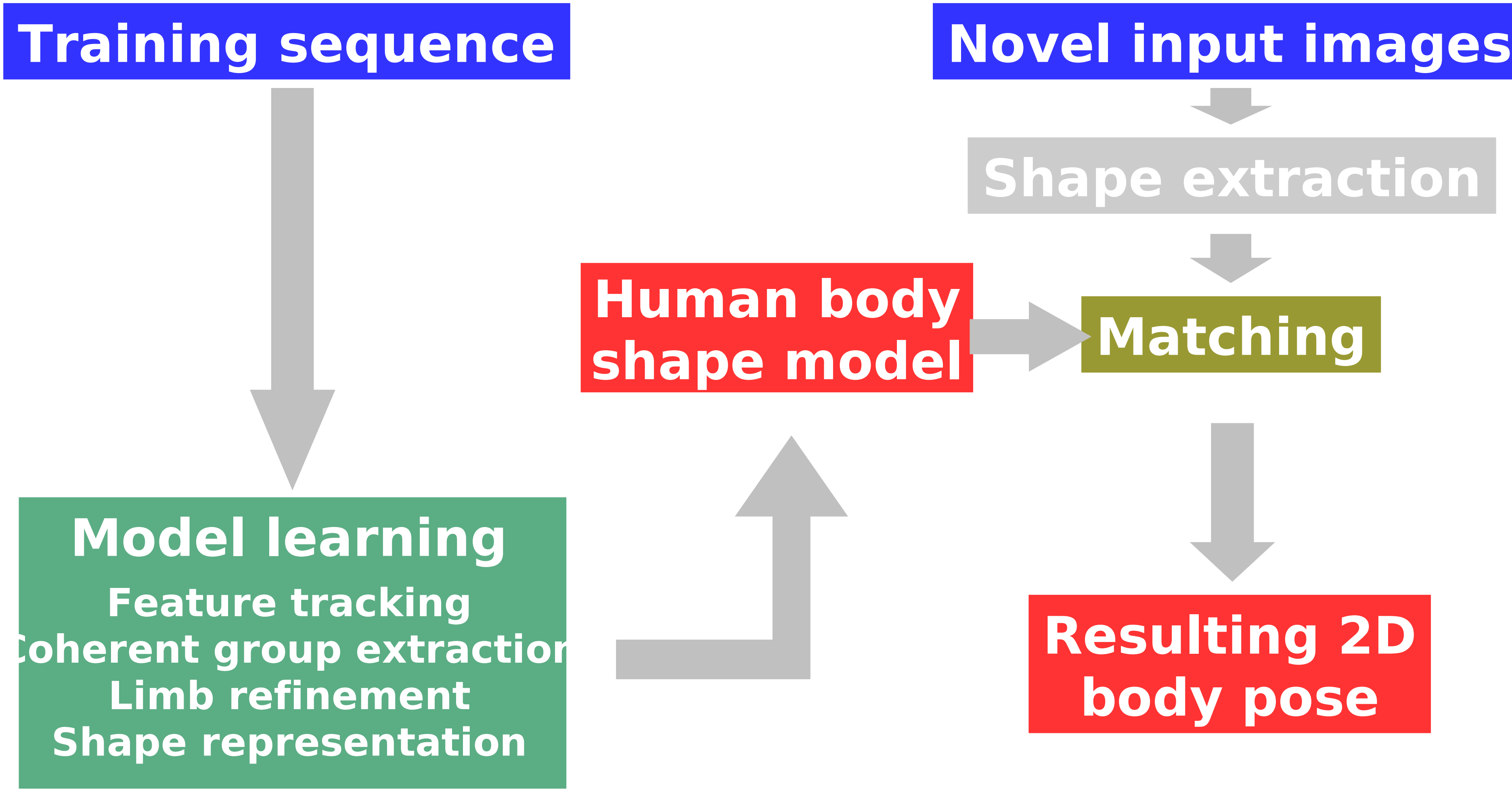
Former results (pictorial structure paradigm [Fel05])

Autonomous limb learning [WW08]

Pictorial Structure Model



Pose estimation cycle
(still images, single model)



Former results

- Using a single HBM for still image analysis gives encouraging results, <u>however</u>:
  - *shape generality* of the employed model is necessarily weak
  - *color generality* is even worse, and had therefore been neglected so far

Solution approach

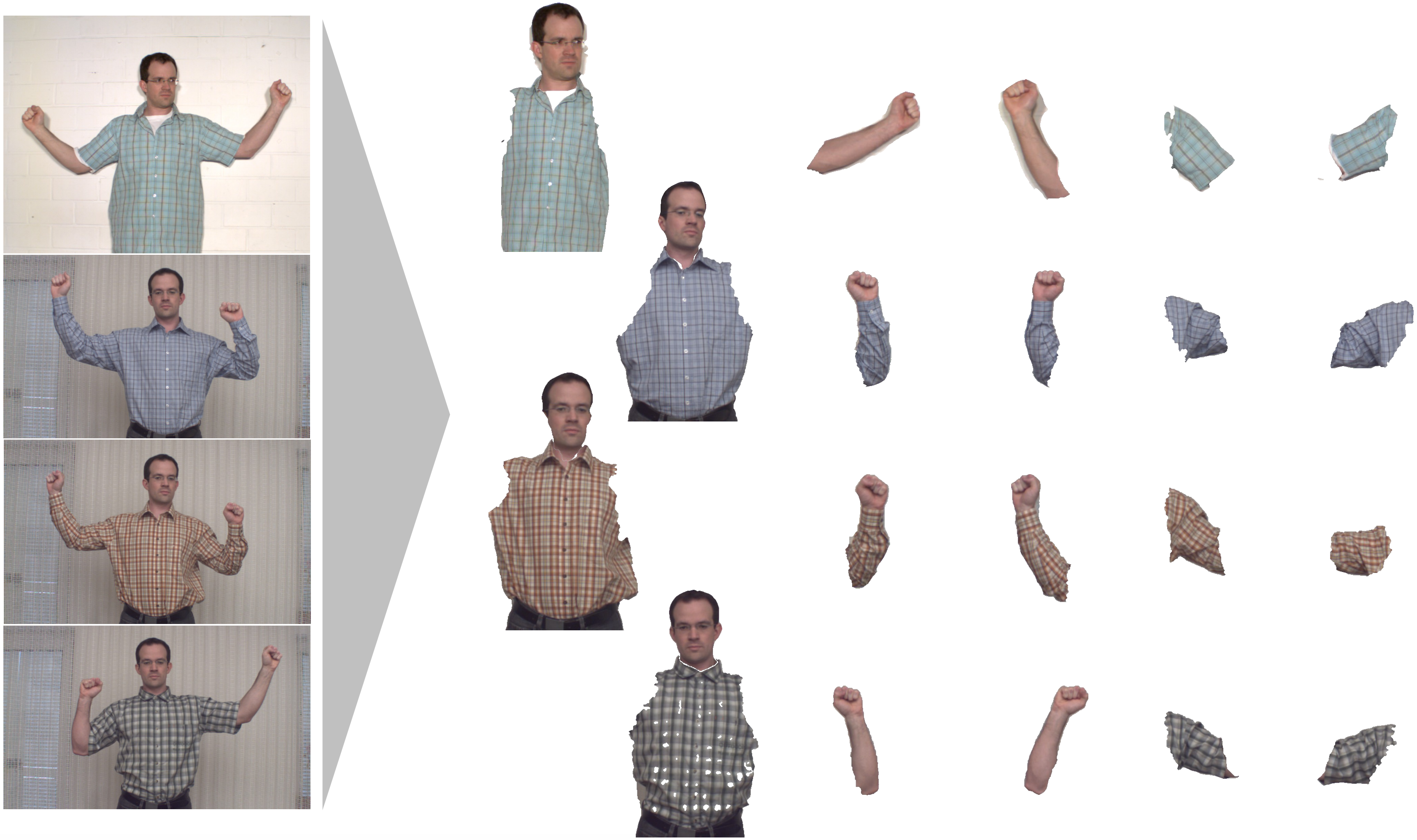
- Combine multiple models, learned from different sequences, to better generalize over shape and color information

- Additionally, learn angular constraints of body joints from a couple of sequences, enhancing avoidance of unlikely poses

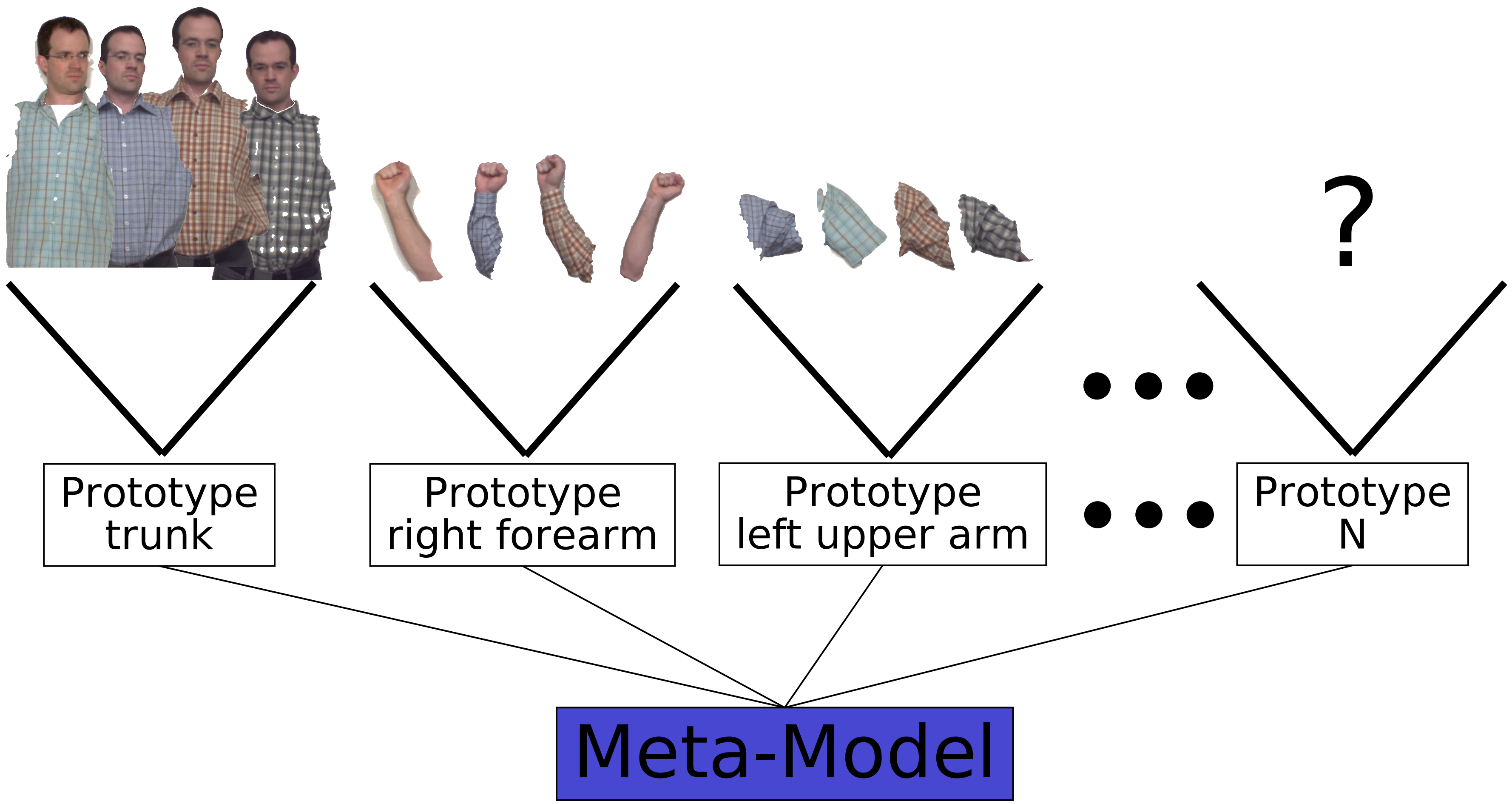
Knowledge combination - two approaches (cf. [Mur04])

- *Exemplar* view
  - blandly store each incoming information fragment as a single exemplar of the observed limb
  - fast, simple storage routines can often be used
  - information recall often requires complex mechanisms or becomes unbearably slow, memory requirements quickly become unwieldy
- *Protoype* view
  - incoming information fragments are used to form prototypical representations of the single limbs
  - information recall is generally fast and simple, memory requirements are normally low
  - learning (i. e. merging information into the prototypes) is computationally expensive

Prototype approach (learning limbs from multiple sequences)



Prototype approach
(meta model generation)



Shape prototype generation

- How to select the most promising prototype candidate frame from each input sequence?
- Cloth induces significant shape changes of the observed limbs over time, making selection difficult and necessarily non-generic
- Idea: build intra-sequence limb shape prototypes first (mean limb shapes) and combine these to eventually formulate a meta limb

Learning intra-sequence mean shapes

- Standard approach: use human domain knowledge to annotate input data with unique landmarks (in simple cases, this step might be automatized)

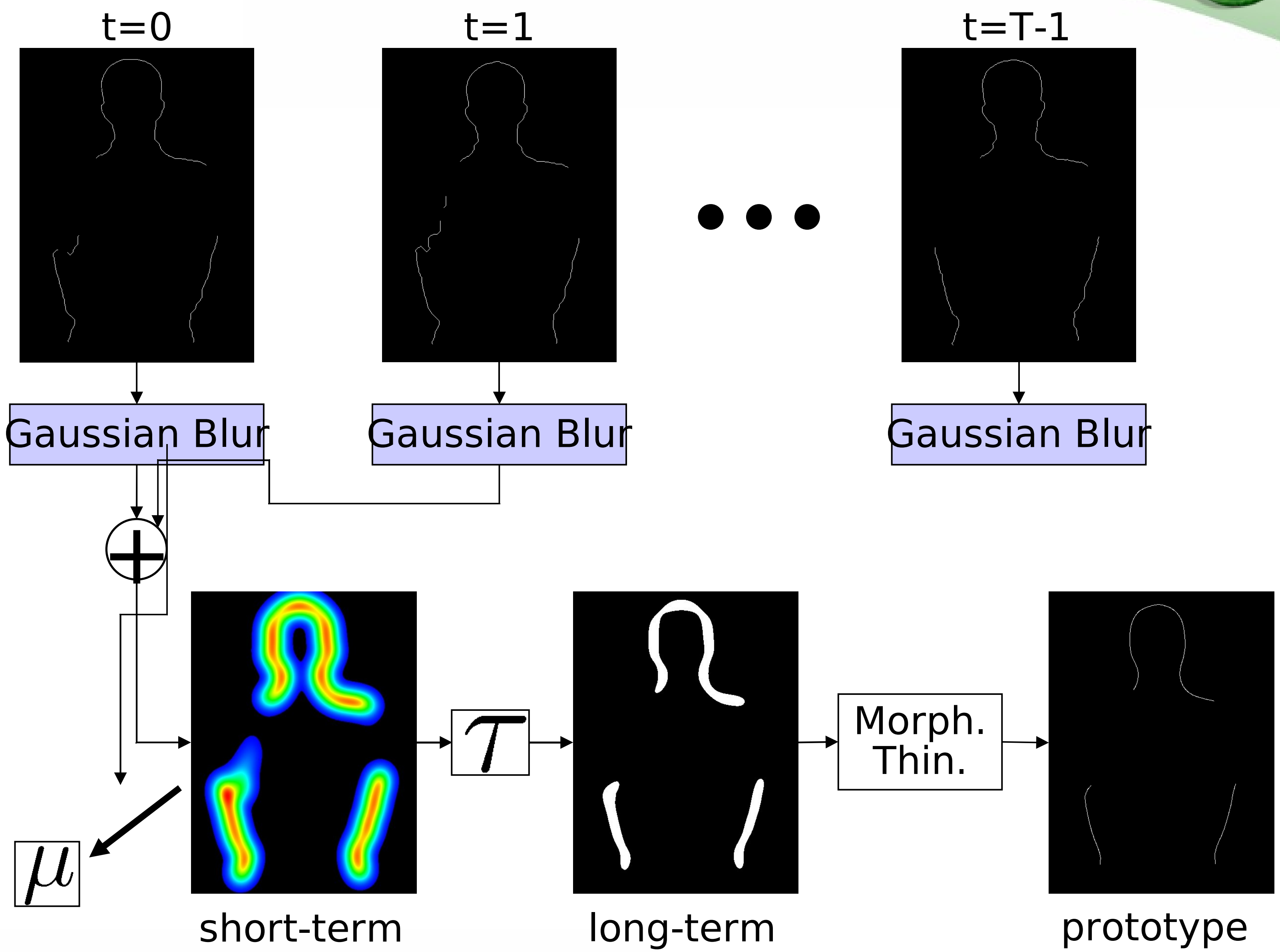
- Register the retrieved landmarks

- Average the registered landmark points to find mean shape,
  possibly extend to point distribution models (cf. [Coo04])

- However: no manual annotation availiable here; automatic annotation is far too unreliable in the current context

Learning intra-sequence mean shapes

- Our approach (emulating short-/long-term memory):
  - though dedicated landmarks cannot be found, whole limb registration is easily possible (the pose of each limb is known in all frames)
  - shapes from frames >0 are mapped back to frame 0 and are blurred by Gaussian with large σ
  - the ‚Gaussianized‘ shapes vote for the final mean shape by being added to an accumulator image; to reduce the effect of outliers, votes evaporate over time with an empirical rate μ
  - if votes for a certain pixel exceed a relative threshold τ, the pixel irreversibly becomes part of the sought-after mean shape
  - the resulting, blurry mean shape is refined by thinning
  - Vote scheme partially inspired by Graumann/Lee: Shape Discovery from Unlabeled Image Collections (to appear)


t=0
t=1
t=T-1
Gaussian Blur
Gaussian Blur
Gaussian Blur
τ
Morph. Thin.
μ
short-term
long-term
prototype

Learning global mean shapes

- Initial meta model limb shape prototypes (MMLSP) equal the intra-sequence shape prototypes derived from the first input sequence

- Given a new input sequence, limb correspondences are retrieved by
  - registering the new limb shape prototypes to the MMLSP (using a fully-articulated model matching cycle->stability of results)
  - solving the arising assignment problem becomes trivial (and is here based on COG distances of the single limbs)

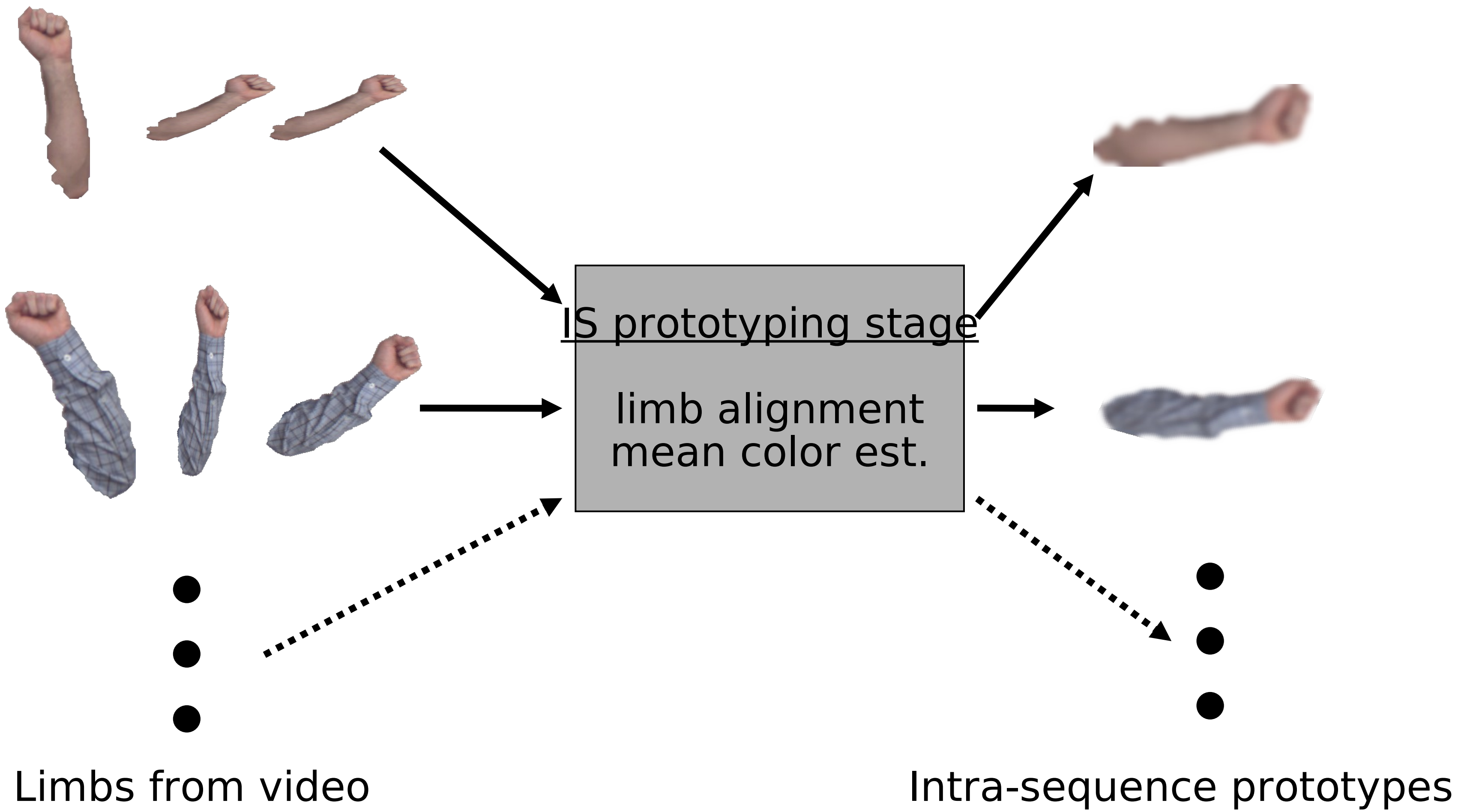
Learning global mean shapes

- Using the correspondences, characteristics (limb/joint orientation and labeling) of the new model are fitted to the existing meta model
- Residual distances between the registered limbs (new
  model/meta model) are largely annihilated by local ‚Iterative Closest Point' (ICP)-Techniques [Zha94]
- Global mean shapes are again learned by simplified Gaussian accumulation of the incoming intra-sequence prototypes; followed by simple thresholding and thinning

Learning global color prototypes

- For each sequence: find pixelwise color mean for each limb by integrating over all input frames->intra sequence color prototypes (ISCP)
- Initialize meta model with the first encountered ISCP
- Build limb correlation masks by windowed, thresholded, pixelwise cross-correlation between aligned incoming ISCP and existing meta model color prototypes
- Inside correlation masks, find pixel-wise color mean to establish global color prototypes

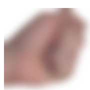
Learning global color prototypes



Learning global color prototypes

Global prototyping stage

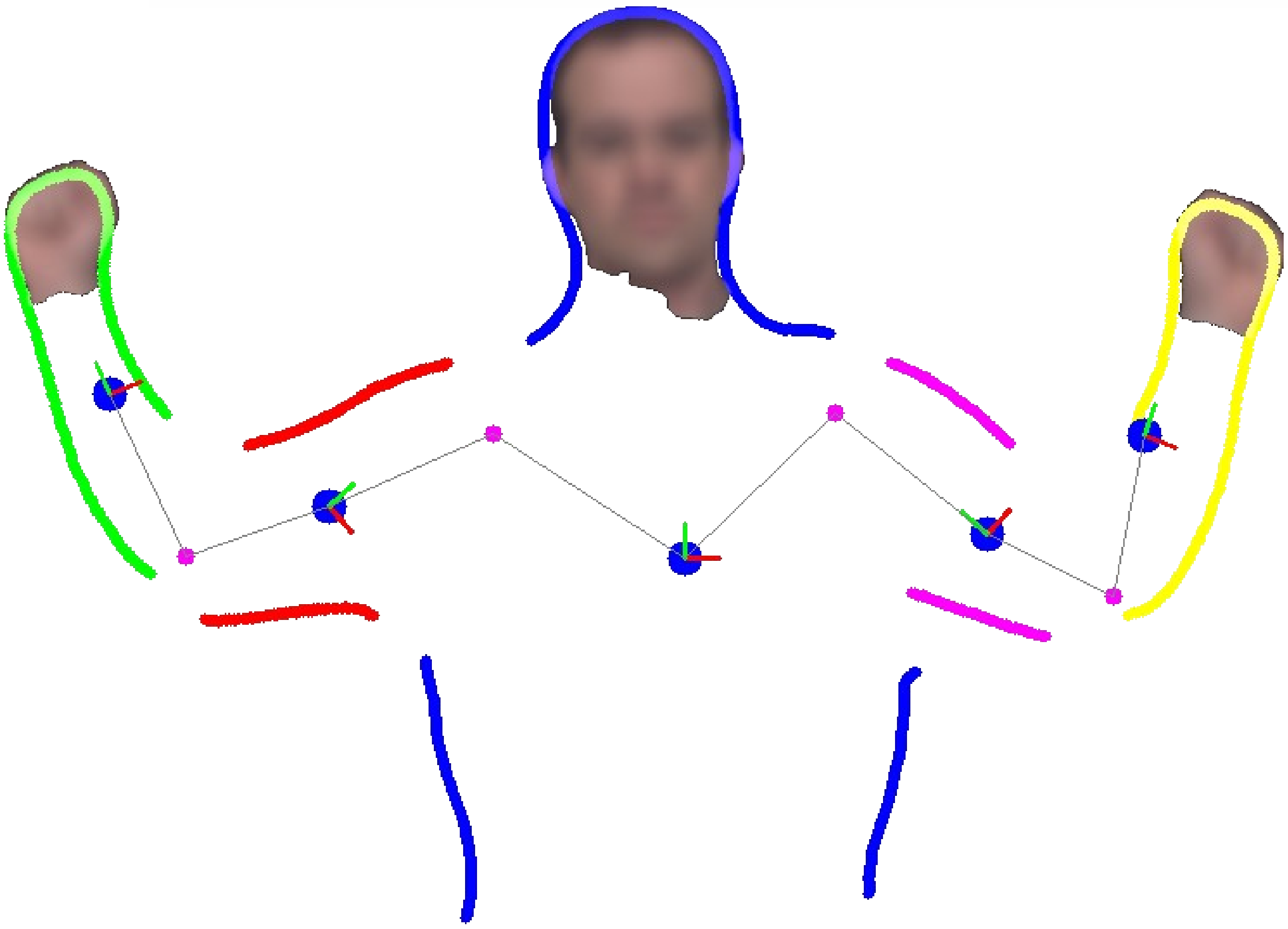
model alignment
ICP registration
windowed histogram extraction
windowed cross correlation



Globally persistent
color regions

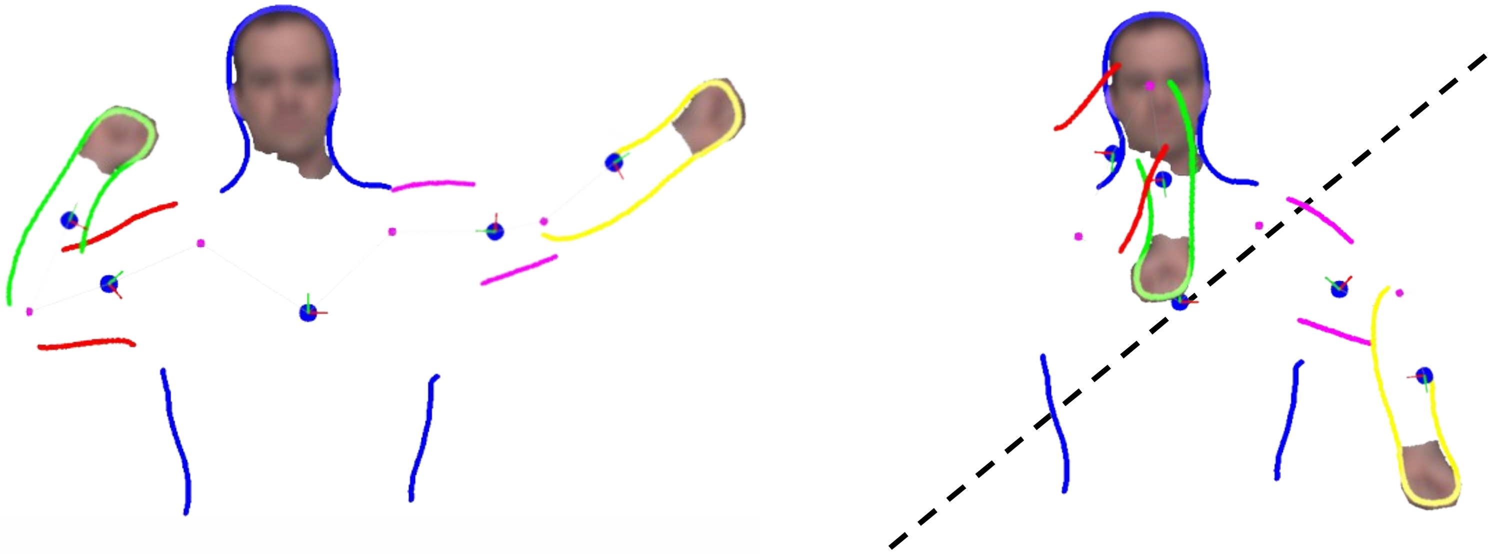
Intra-sequence prototypes

Resulting meta model



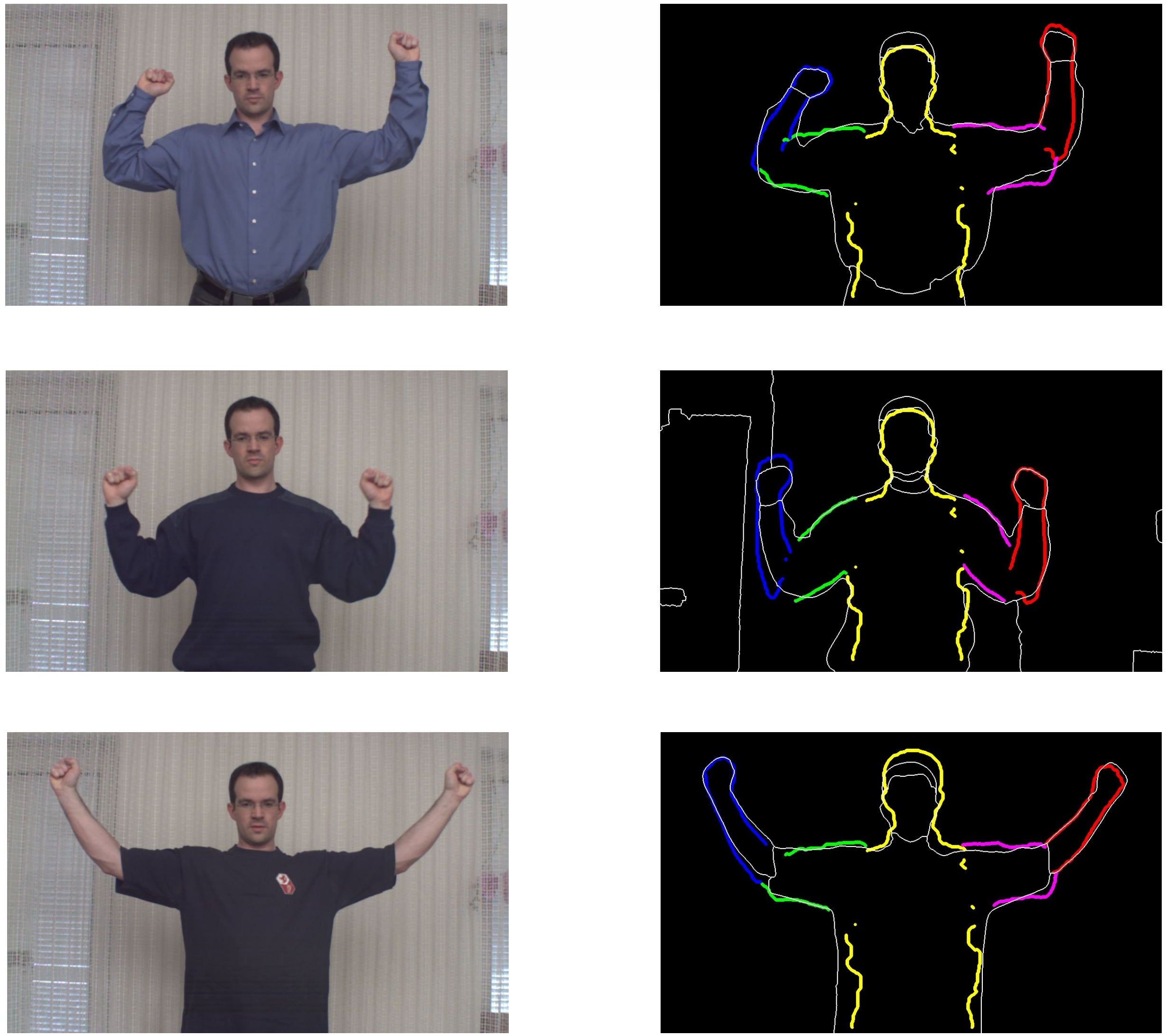
Learning joint angle constraints

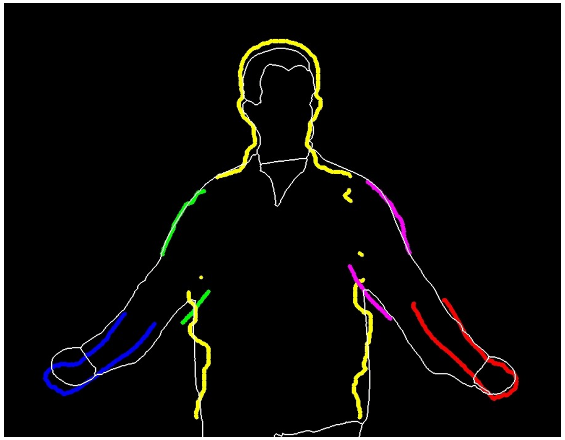
- Minimum/maximum joint angles (relative angles between limbs) come for free
- Integrating these angular constraints into the matching process is feasible and helps to avoid unlikely postures:

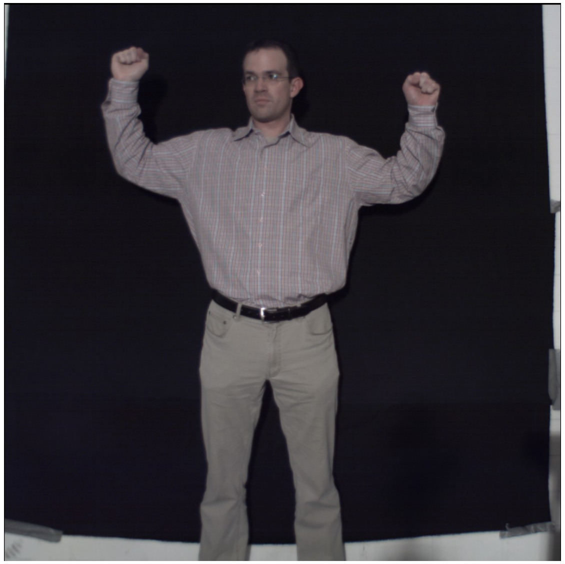


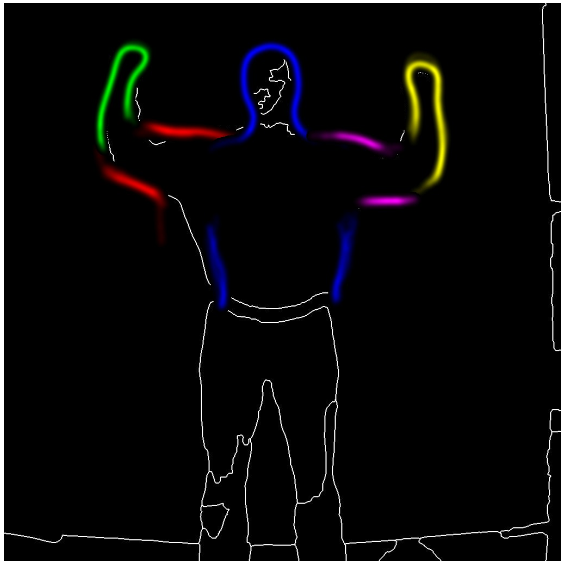
Generalization experiments [WW09]

- Meta model has been learned from a single individual, given a limited number of simple training scenarios, so generalization capabilities have to be probed across:
  - changing clothes
  - varying camera distances
  - previously unseen, more complex scenario backgrounds
  - varying camera devices
  - different individuals

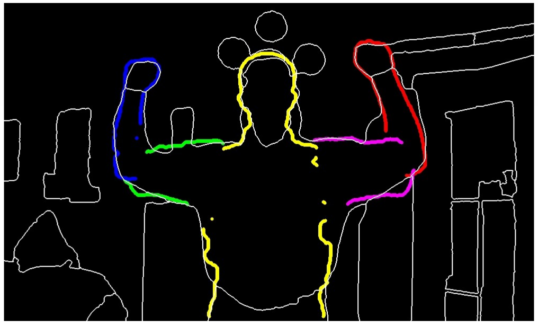
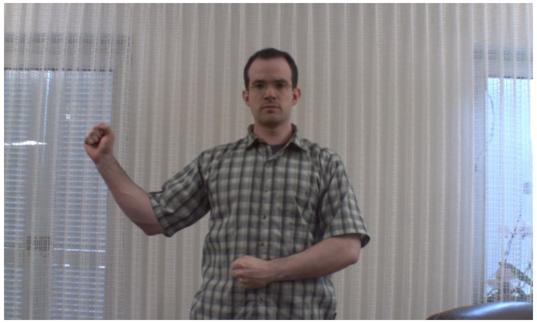
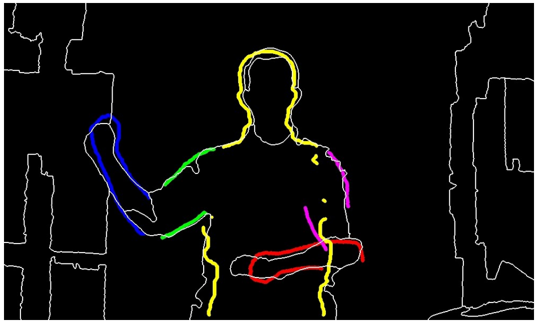


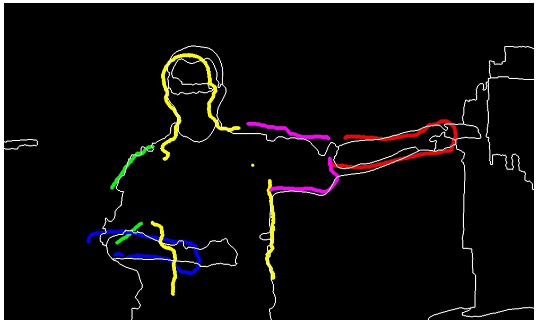


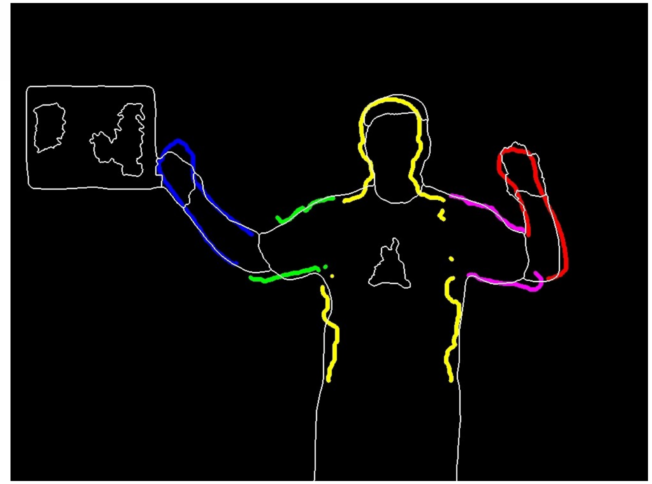
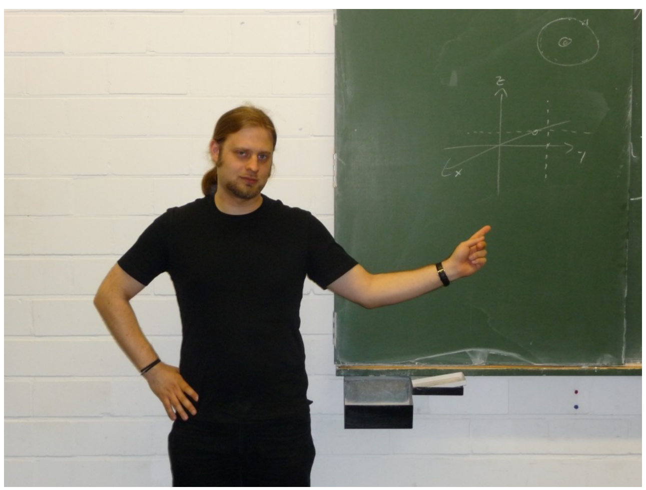
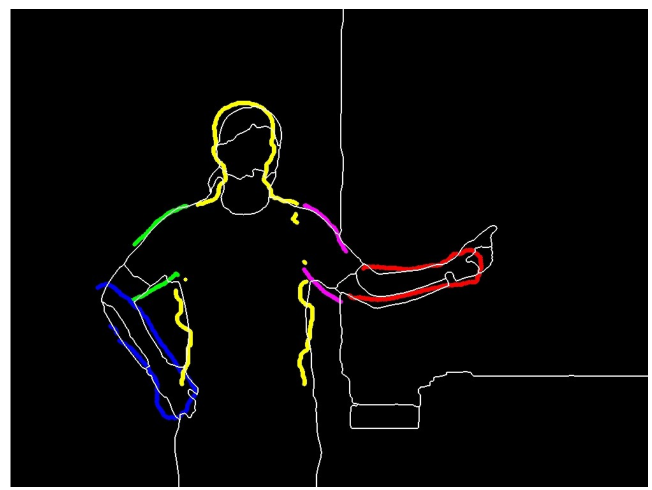




Different output routine!



Conclusion and outlook

- System learns abstract representations of the upper human body in a fully autonomous manner
- Pertinent features (e. g. mean limb shape, persistent color patches) are kept, while irrelevant information (cloth deformation, illumination, ...) is largely discarded
- Angular constraints prevent unlikely poses from being detected

Outlook

- Shape and angular constraints are already exploited, color cues will be integrated next
- Broadening the range of recognizable poses
- Tracking mechanisms could be employed to allow for continuous pose estimation in video streams

Outlook

- Transfer of methods onto a humanoid robot

- Transfer of methods onto multiple cameras (cooperation with Univ. Hannover)

Literature:

- [Coo04]: T. F. Cootes and C. J. Taylor: Statistical Models of Appearance for Computer Vision, report, University of Manchester, Image Science and Biomedical Engineering, 2004
- [Fel05]: Pedro F. Felzenszwalb and Daniel P. Huttenlocher: Pictorial Structures for Object Recognition, International Journal of Computer Vision, volume 61, 55-79, 2005
- [FE73]: Martin A. Fischler and Robert A. Elschlager: The Representation and Matching of Pictorial Structures, IEEE Transactions on Computers, volume c-22, 67-92, 1973
- [Mur04]: Gregory L. Murphy, The Big Book of Concepts, MIT Press, 2004
- [WW08]: T. Walther and R. P. Würtz, Learning to look at humans – what are the parts of a moving body?, in Proceedings of the Fifth Conference on Articulated Motion and Deformable Objects, Mallorca, Andratx, Juli 2008, Springer, 22-31
- [WW09]: T. Walther and R. P. Würtz, Unsupervised learning of human body parts from video footage, ICCV09, Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, Kyoto, Japan, Sept. 2009, in press
- [Zhang94]: Z. Zhang: Iterative point matching for registration of free-form curves and surfaces, International Journal of Computer Vision, volume 13, issue 2, pp. 119-152, 1994