Hierarchical Object Parsing from Structured Noisy Point Clouds

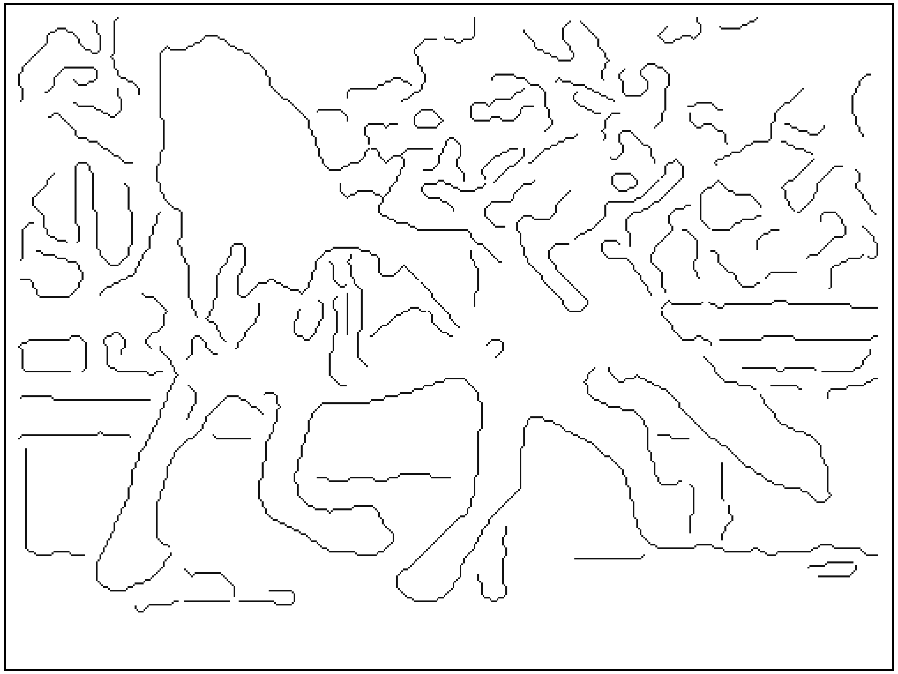
Adrian Barbu

Department of Statistics

Florida State University

Objective and Motivation

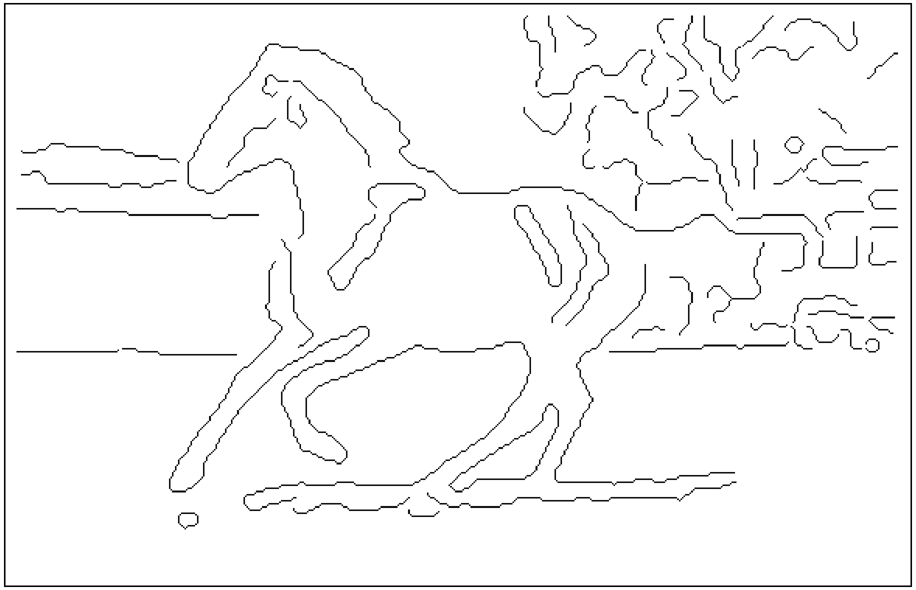
- Objective
  - Accurate object parsing: e.g. horse parsing
  - Find the object and delineate its boundary and parts
  - Input: noisy point cloud e.g. edge detection
  - Fast and robust



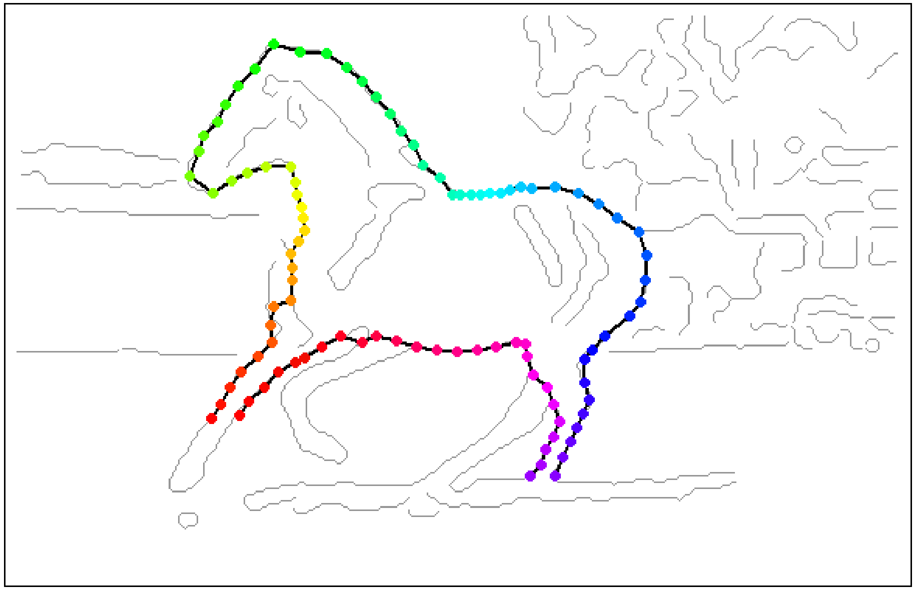
- Difficulties
  - Large amounts of missing data
  - Large amounts of noise
    - Points in the background
  - Large shape variability due to
    - Viewpoint
    - Articulations: position of head and legs

Object Parsing

- Find position of points of interest of the object
  - E.g. aligned boundary
- Fill-in missing data using the shape prior



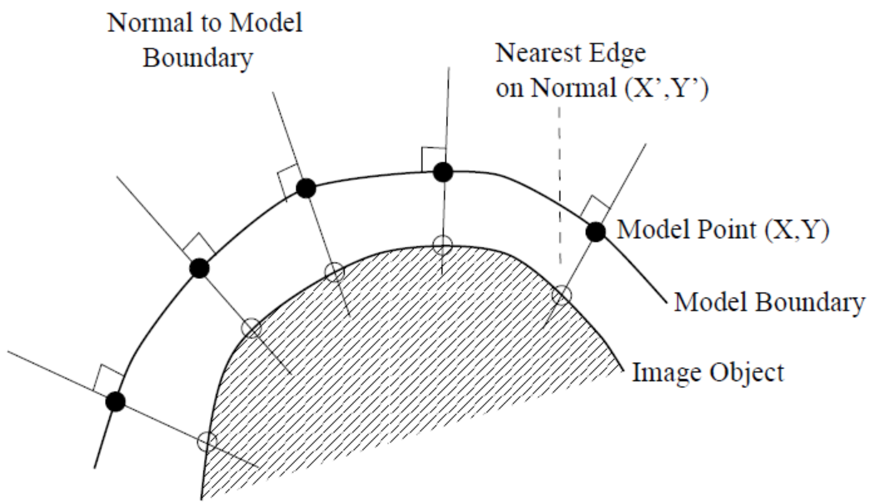
Edge detection



Parsing result

Active Shape Model

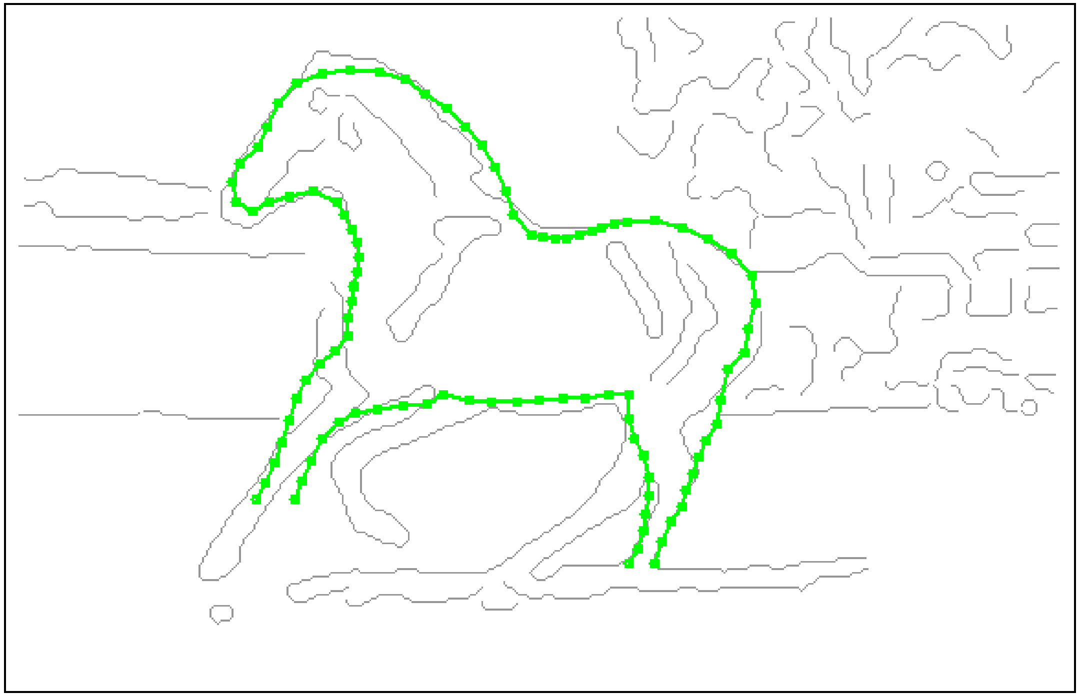
- Overview:
  1. Start with an initial shape (A,β)
  2. Find most probable boundary edges along each normal
     - Obtain a rough shape
  3. Project rough shape to PCA space
     - Obtain new shape (A,β)
  4. Repeat 2-3 until convergence



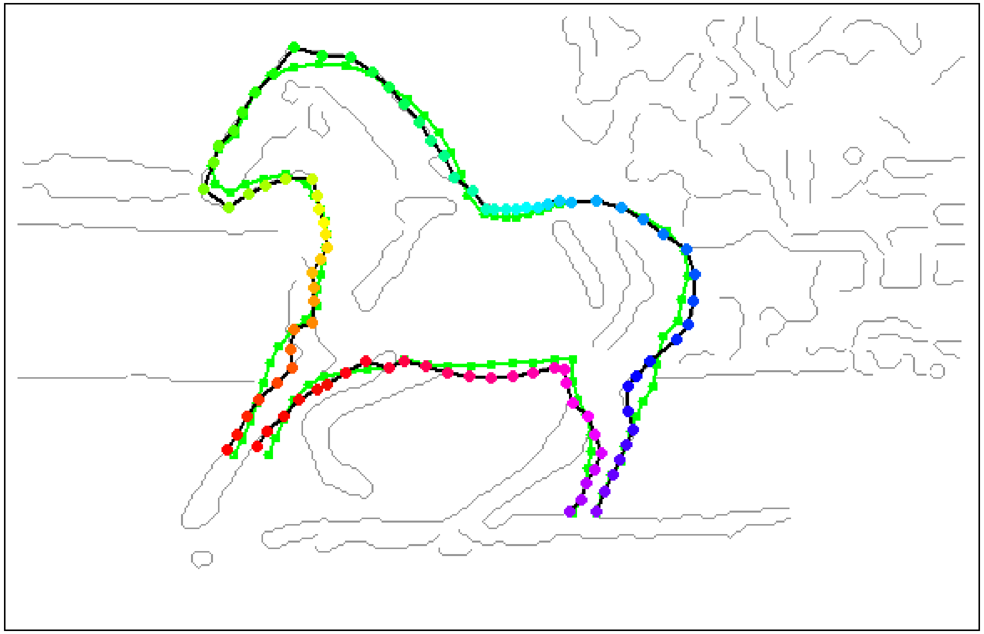
- ASM Advantages:
  - Fast
  - Good dimensionality reduction
  - Works well on clean data

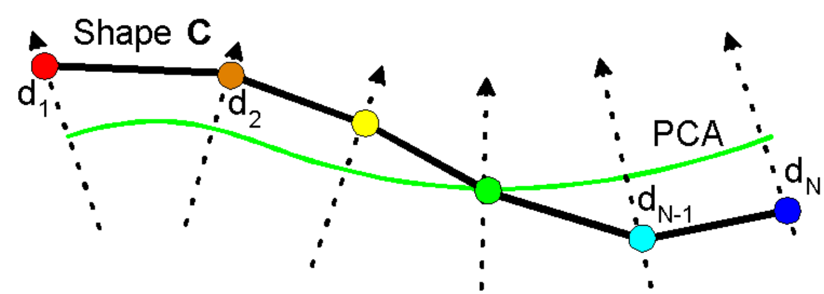
ASM Disadvantages

- Not accurate enough
  - Low dimensional representation cannot have high accuracy
  - Cannot be used for point clouds
- Depends on initialization
- Not clear what model it optimizes



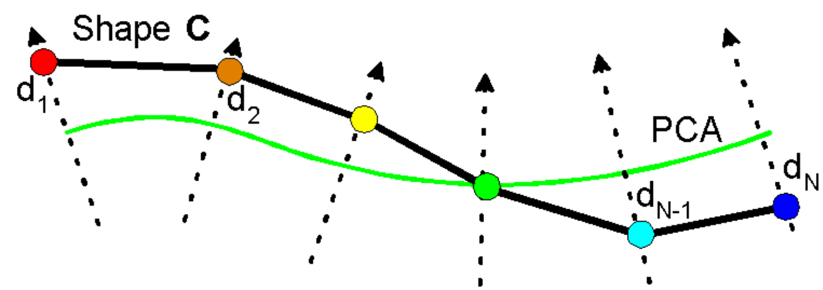
Proposed Hierarchical Model





- Bayesian model
- PCA model (A,β) to limit shape variability
  - Serves as a backbone
- MRF deformation from PCA along normals
- Data term based on edge continuation

Hierarchical Model



- Bayesian model:

$$E(C,A,\beta) = E_{data}(C) + E(C|A,\beta) + E_p(A) + E_p(\beta)$$

Prior Models $E_p(A)$, $E_p(\beta)$

PCA Model
(A,β)

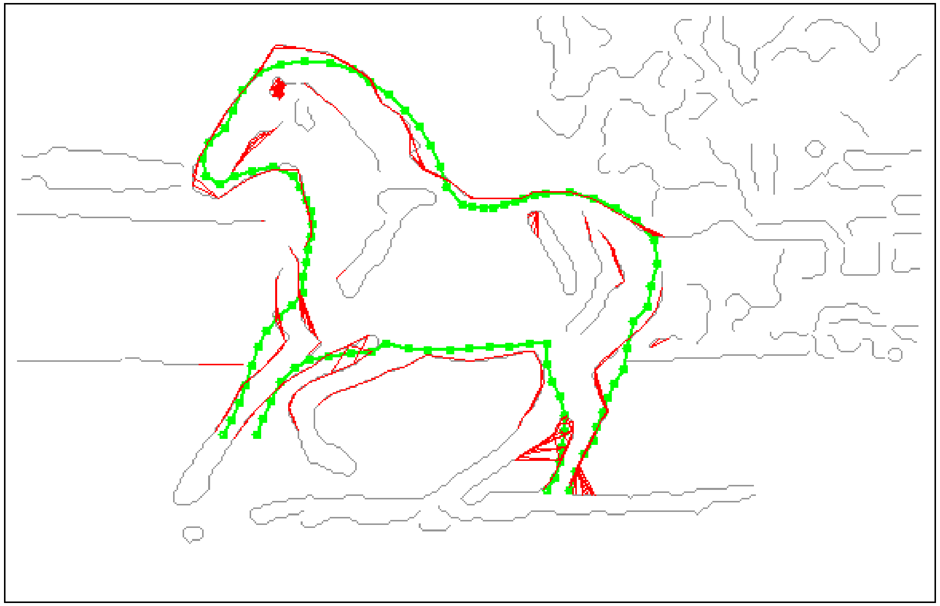
$$E(C|A,\beta) = \sum_i \alpha_i d_i^2 + \sum_i \gamma_i (d_i - d_{i-1})^2$$

Object Shape
$C=(C_1,\ldots,C_N)$

$$E_{data}(C) = \sum_{i=1}^{N-1} \varphi(C_i, C_{i+1})$$

Noisy Point Cloud
(edge detection)

Hierarchical Model



PCA shape and DP data edges

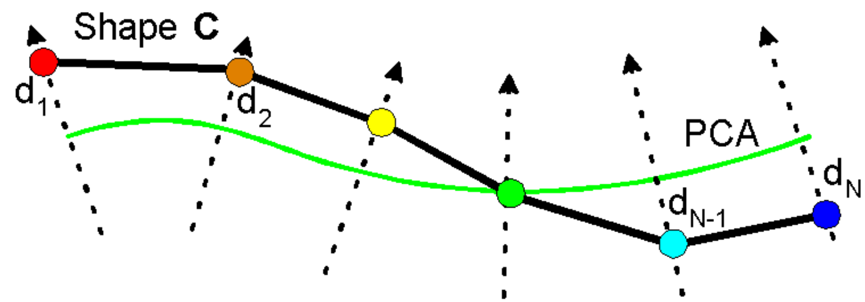
- Data Term:

$$E_{data}(C) = \sum_{i=1}^{N-1} \varphi(C_i, C_{i+1})$$

$$\varphi(P,Q) = \begin{cases} -1 & \text{if there is a edge connecting P and Q} \\ 0 & \text{otherwise} \end{cases}$$

- Encourages deformations that have edges (chains of points) connecting them

Hierarchical Model



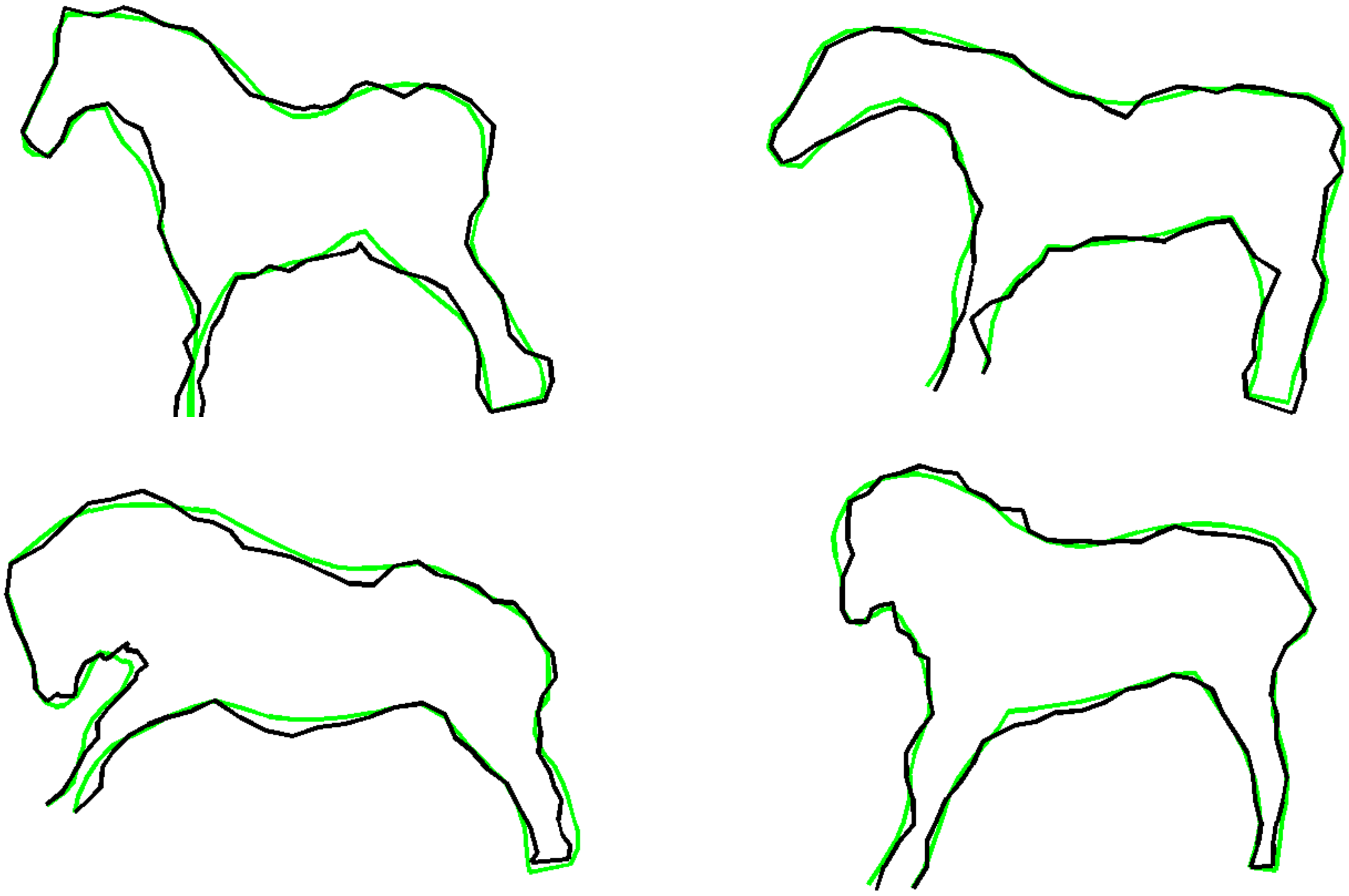
- Prior Term:

$$E(C|A,\beta) = \sum_i \alpha_i d_i^2 + \sum_i \gamma_i (d_i - d_{i-1})^2$$

- Encourages small deformations that are parallel to the PCA
- Prior E(A) allows a range of orientations and scales
- Prior E(β) is from a multivariate normal based on the PCA eigenvalues

$$E_p(\beta) = \rho \sum_{i=1}^{N} \frac{\beta_i^2}{\lambda_i}$$

Generative Model



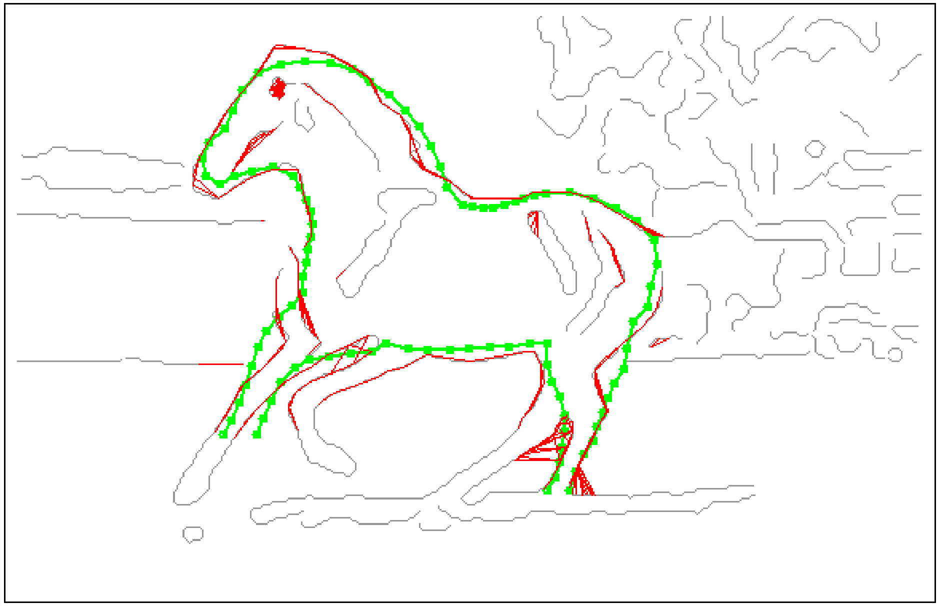
- Samples from the generative shape prior model

Advantages and Challenges of the Proposed Model

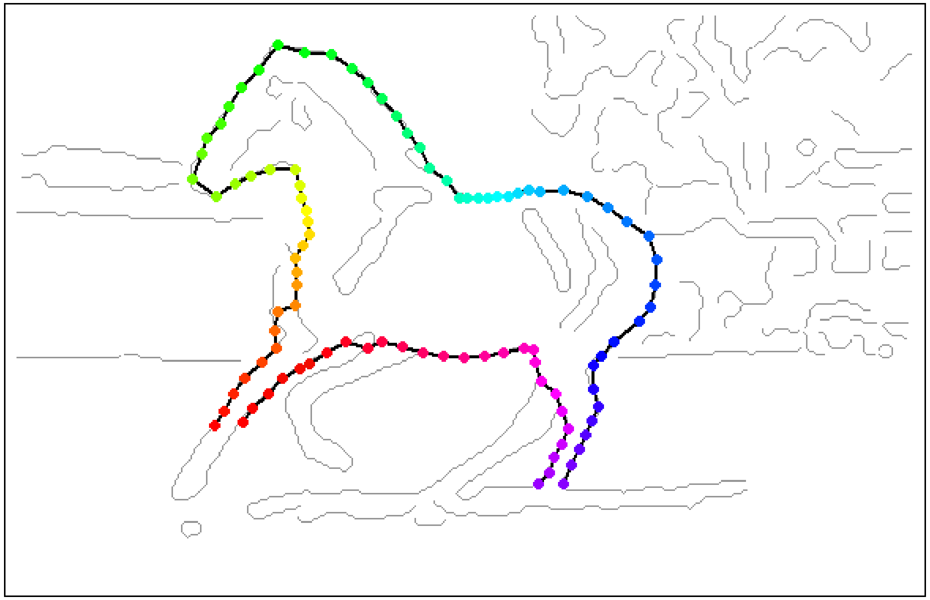
- Generative model
  - The shape model with prior can be sampled to get an idea on the shape variability
  - Small number of parameters means good generalization power
- Flexible yet not too flexible
  - Deformation term allows deviations from the PCA shapes
  - Can accurately follow the object boundary
  - PCA backbone limits the flexibility
- Challenges
  - Cannot use any existing fast inference algorithm
  - MCMC too slow

Towards an Inference Algorithm

- ASM inspired approach
  - Given the PCA shape (A,β), the segmentation can be found by dynamic programming
  - Given the segmentation *C*, the PCA shape can be found by least squares



PCA shape and DP data edges

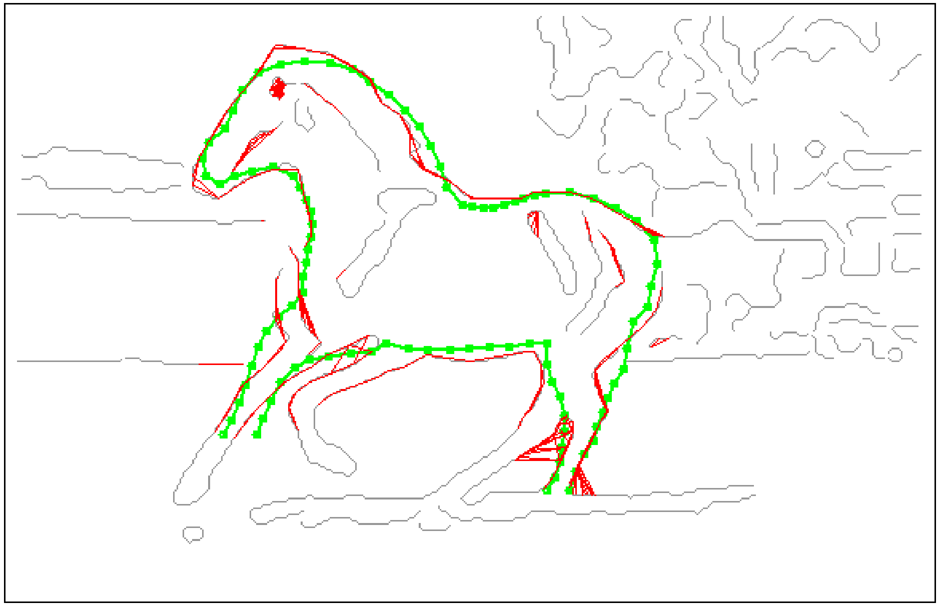


Parsing result

Towards an Inference Algorithm

- ASM-inspired Local Optimization
  1. Start with an initial PCA shape (A,β)
  2. Find segmentation C by Dynamic Programming
  3. Refine PCA shape (A,β) by least squares
  4. Iterate 2-3 until convergence

- Drawbacks:
  - Depends on initialization
  - Obtains local minimum



PCA shape and DP data edges

Inference Algorithm

- Proposed solution
  - Consider many initial candidates (A_i,β_i), $i=1,\ldots,N_{cand}$
  - Run local optimization for each candidate
  - Pick lowest energy solution (C,A,β) as the result

- Challenges:
  - How to choose the initial candidates?
  - How many candidates to use?
- Our solution:
  - Good data-driven (bottom-up) candidates
  - Non-max suppression to avoid repeated candidates
  - Number of candidates chosen based on training error

Related Work

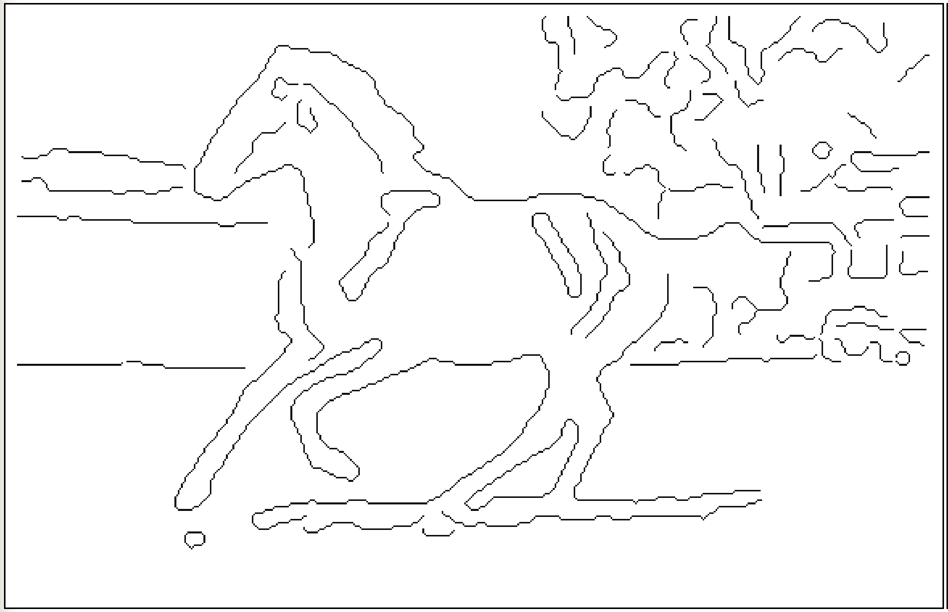
- Recursive Compositional Models, Zhu, Chen & Yuille, 2009
  - Represents shape hierarchically using triplets of parts, each part is a triplet of parts, etc.
  - Dynamic programming with pruning for inference
- Multi-view Car Alignment, Li, Gu &Kanade, CVPR 2009
  - Shape model by Probabilistic PCA
    - Deformation is i.i.d. Gaussian
  - Data term based on classifiers at the model points
    - Uses intensity information
    - Data less noisy than the edge detection
  - Shape proposals obtained by RANSAC
    - Assumes 40% outliers

Related Work

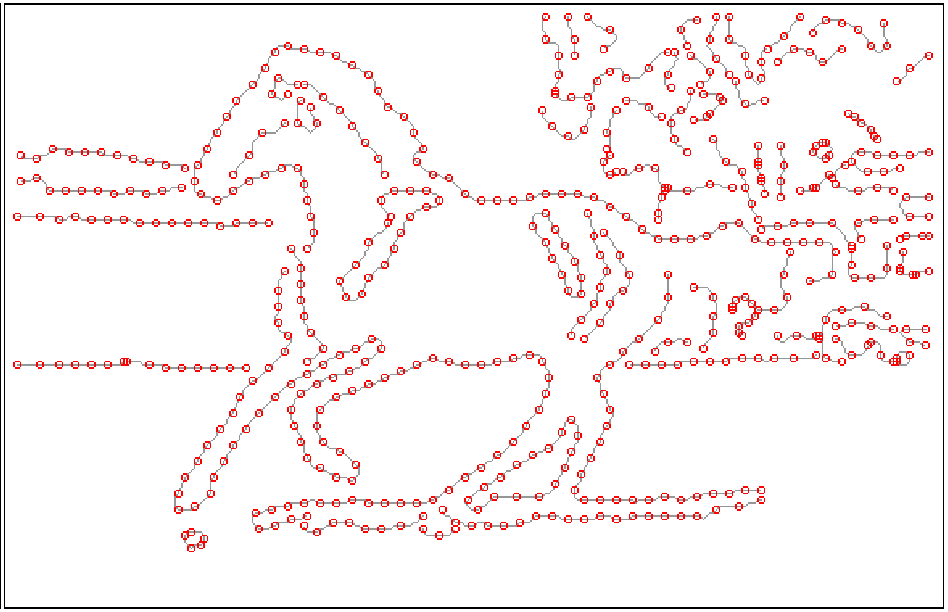
- Hierarchical Shape Matching, Felzenswald & Schwartz 2007
  - Shape model based on a tree
  - Focus on shape matching and retrieval
- Active Skeleton, Bai et al, ICCV 2009
  - Skeleton-based shape model
  - Used for object detection
- Knowledge based segmentation, Besbes et al, CVPR 2009
  - Shape prior based on pairwise cliques
  - Primal-dual algorithm for inference
- Active Basis Model, Wu et al, IJCV 2009
  - A template with local deformations
  - Not used for object parsing

Preprocessing: Segment Chains

- Detected edges are traced into pixel chains
- Pixel chains are cut into chains of short segments



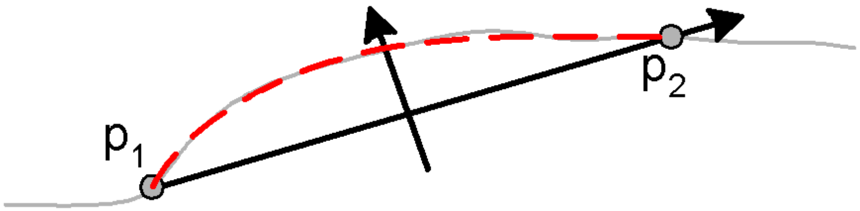
Edge detection

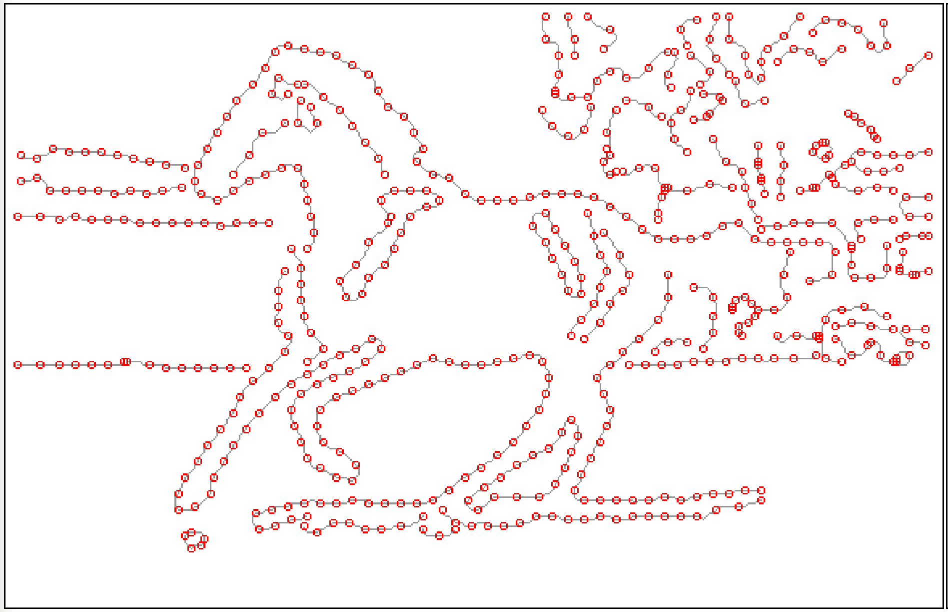


Segment chains

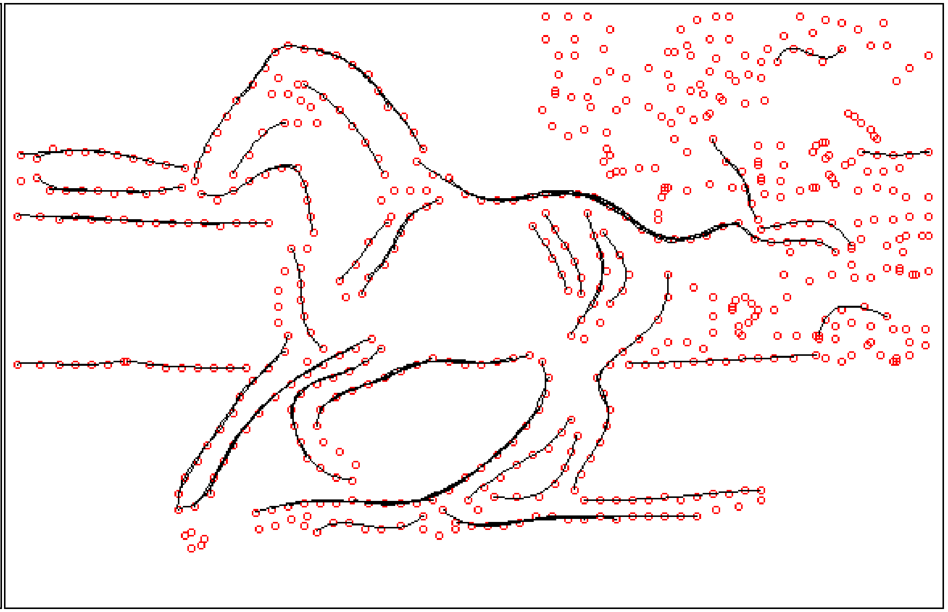
Preprocessing: Smooth Curves

- Parts of segment chains are approximated with polynomial curves





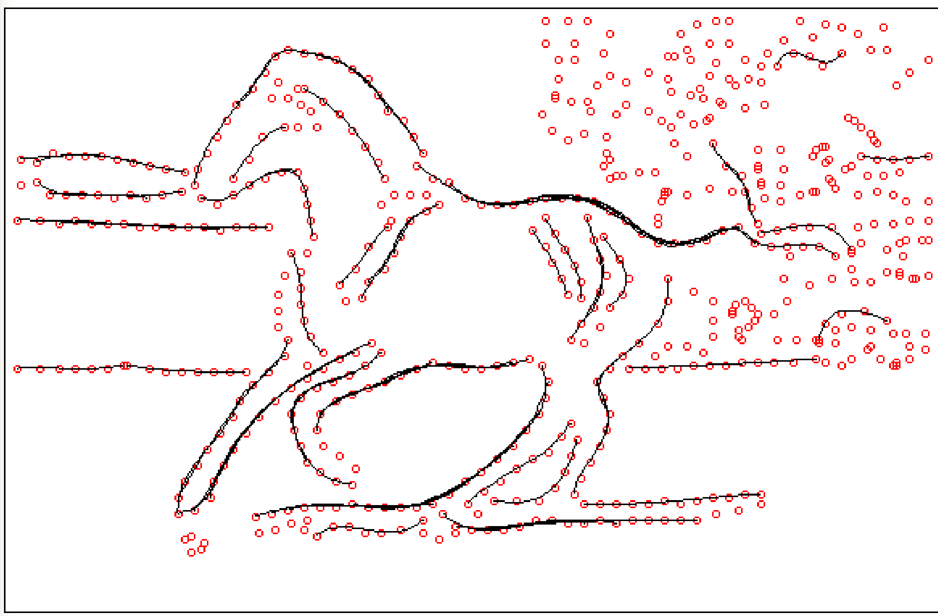
Segment chains



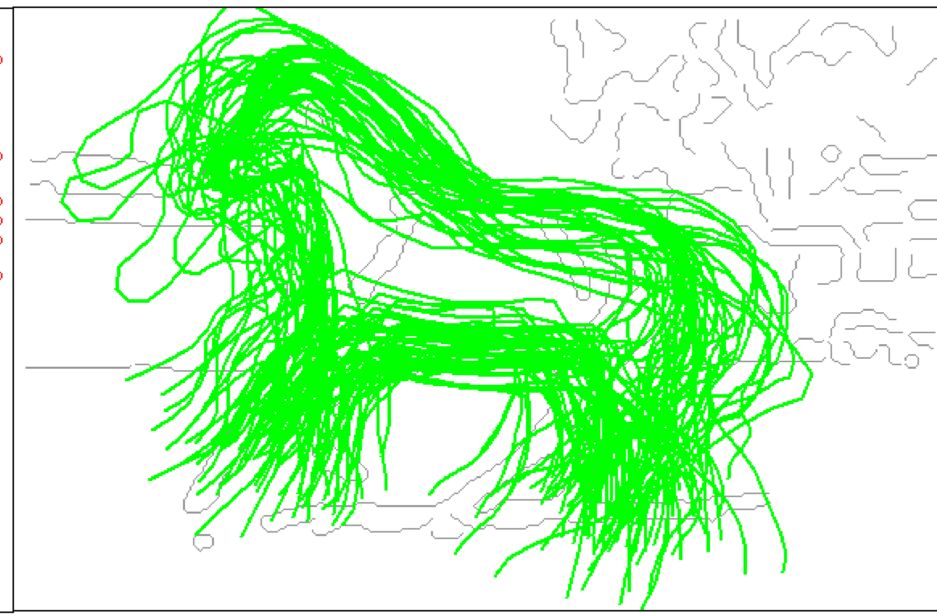
Smooth Polynomial Curves

Shape Candidates

- Match smooth curves from edge detection to parts of the PCA model
  - Find transformation and PCA coeffs in a least square sense
  - Uses weighted least square fitting from Rogers & Graham, ECCV'02
- From one or more smooth curves
- Best fit N_{cand} candidates are kept after non-max suppression



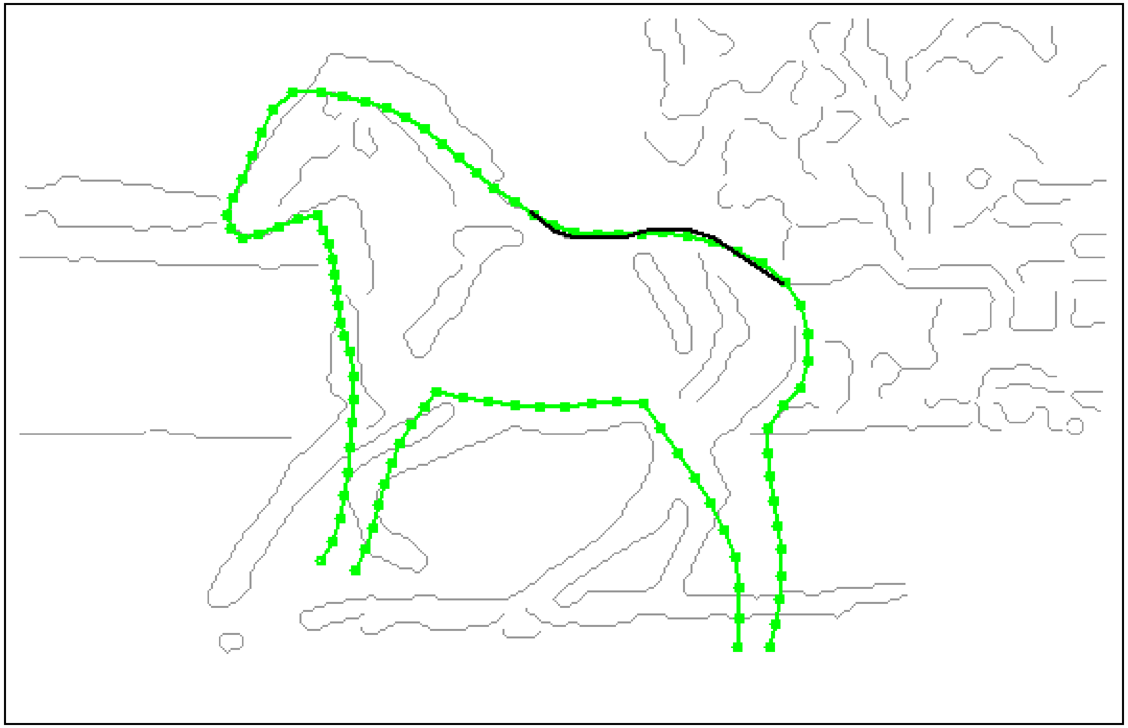
Smooth Polynomial Curves



Best 50 PCA Shape Candidates

One Curve CG

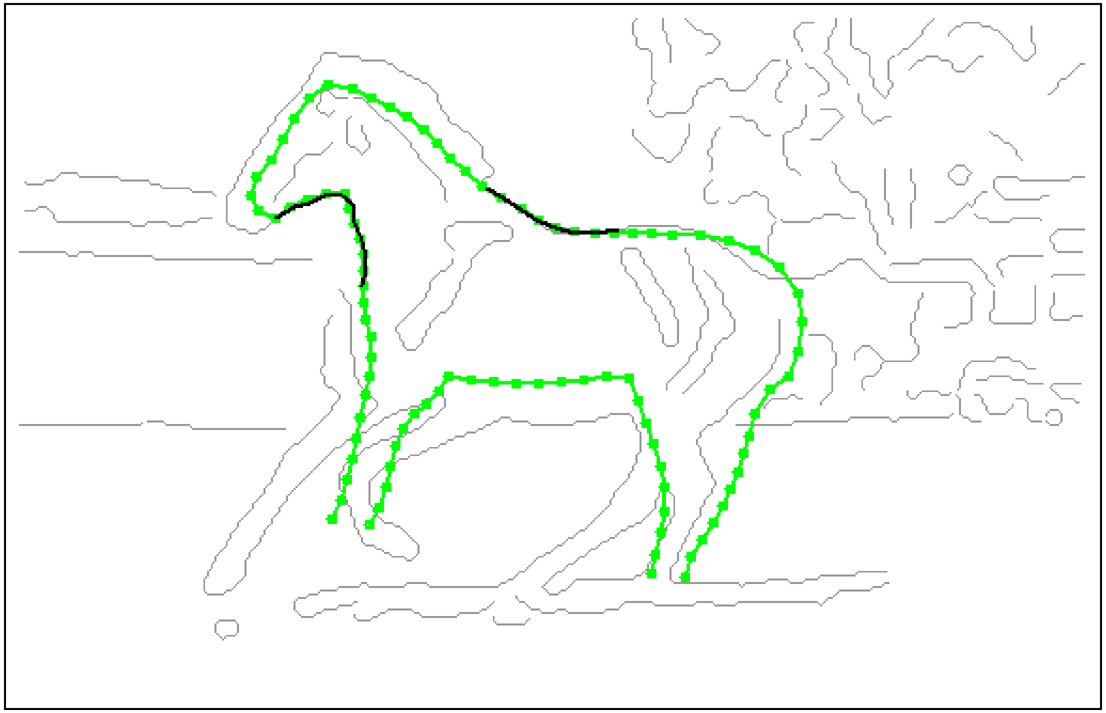
- For each long smooth curve
  - Match it to different parts of the PCA
  - Keep only matches that fit well
- NMS over all obtained candidates to keep best N_{cand}



Best candidate of CG1

Two or More Curve CG

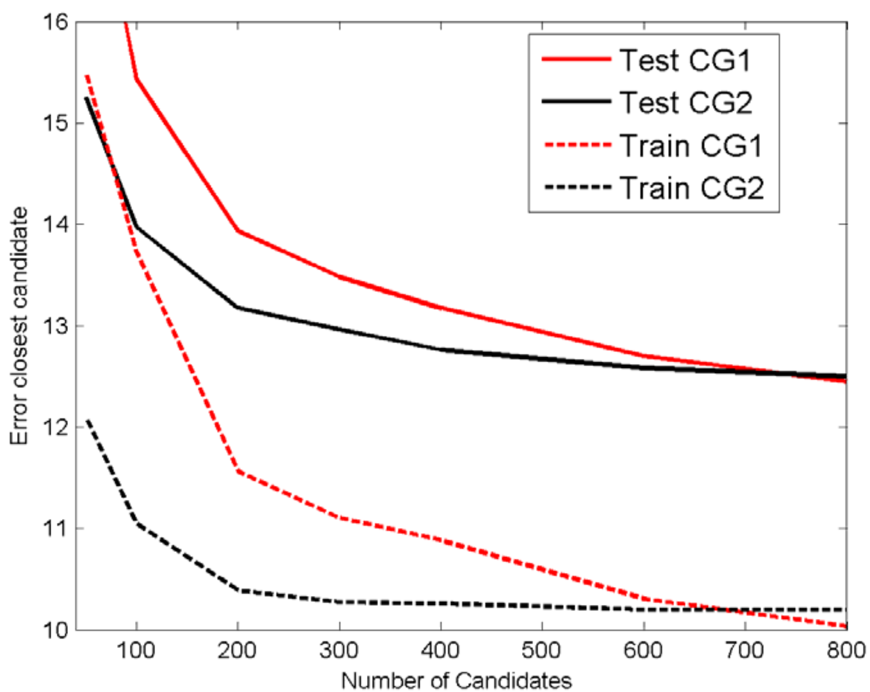
- For each curve candidate from previous CG
  - For each smooth curve close enough
  - Match it to closest points on the candidate
  - Refit PCA
  - Keep only matches that fit well
- NMS over all obtained candidates to keep best N_{cand}



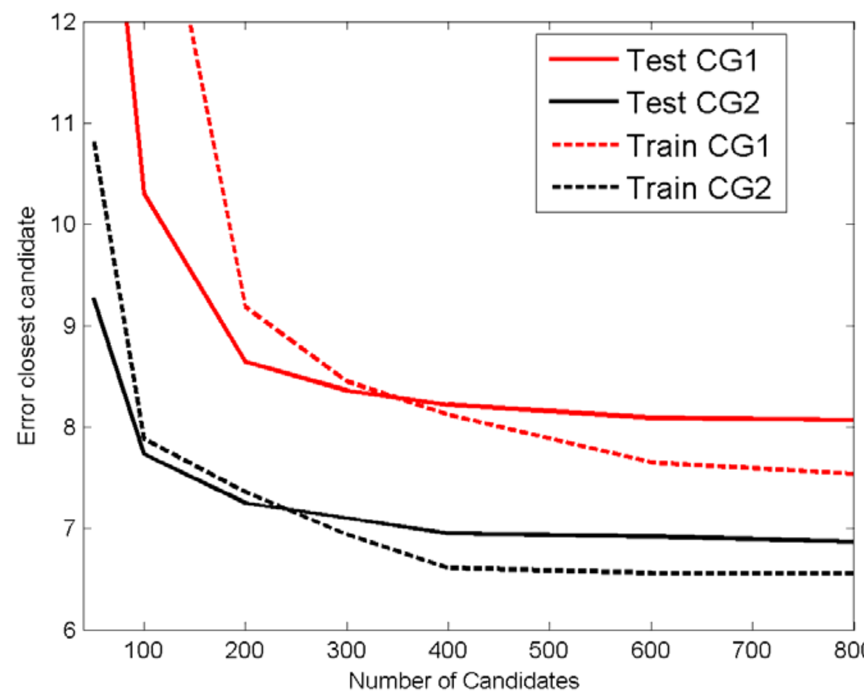
Best candidate of CG2

Learning-based Optimization

- Model and algorithm parameters are tuned on training set for best results
  - Less than 20 parameters totally
- CG parameters are tuned with a different measure
  - Smallest distance of a candidate to GT
- Error on test set follows same trend



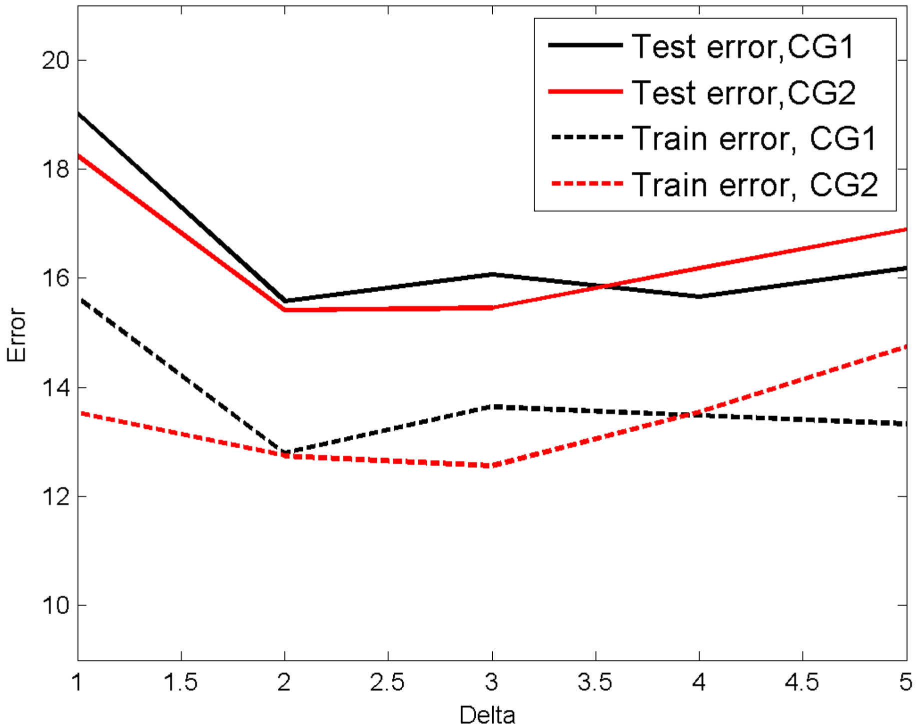
Error vs N_{cand} for horses



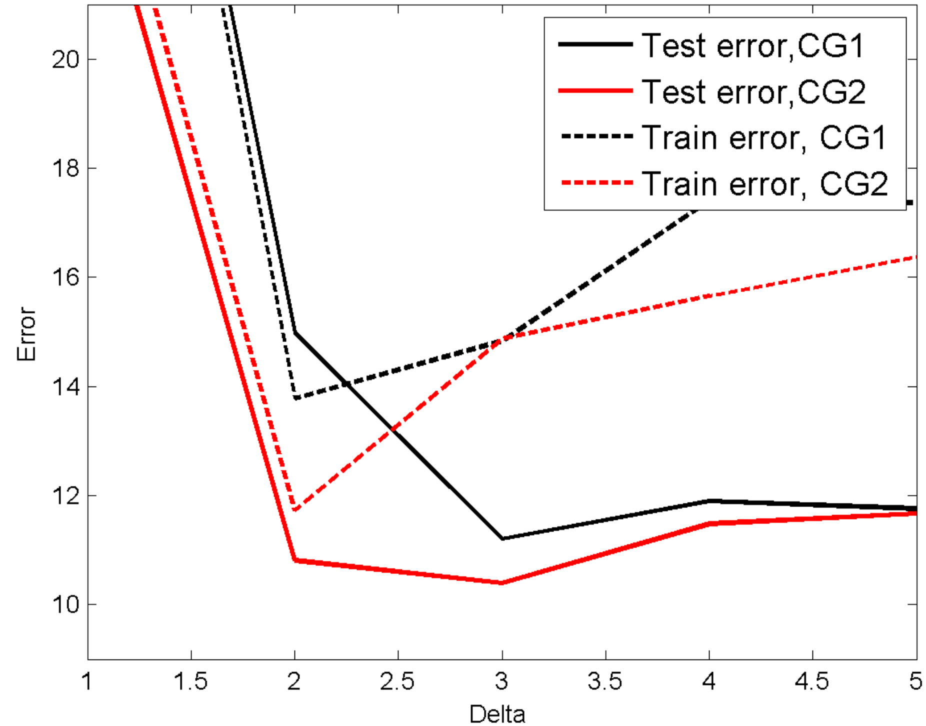
Error vs N_{cand} for cows

Learning-based Optimization

- Segmentation parameters are tuned on training set for best results
  - 5 parameters
  - Error measure is average pt-pt error of the result on the training set
  - Coordinate descent optimization
- Error on test set follows same trend



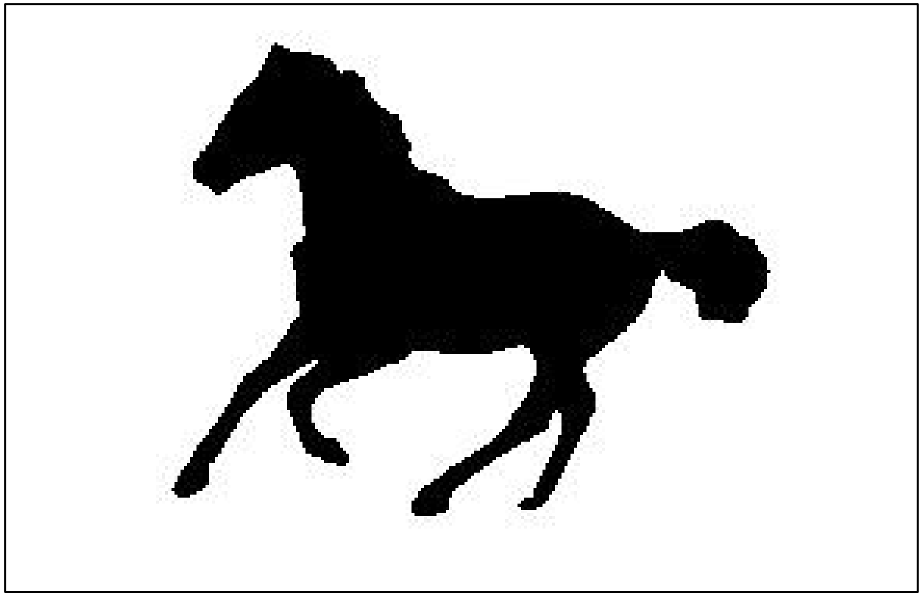
Error vs δ for horses

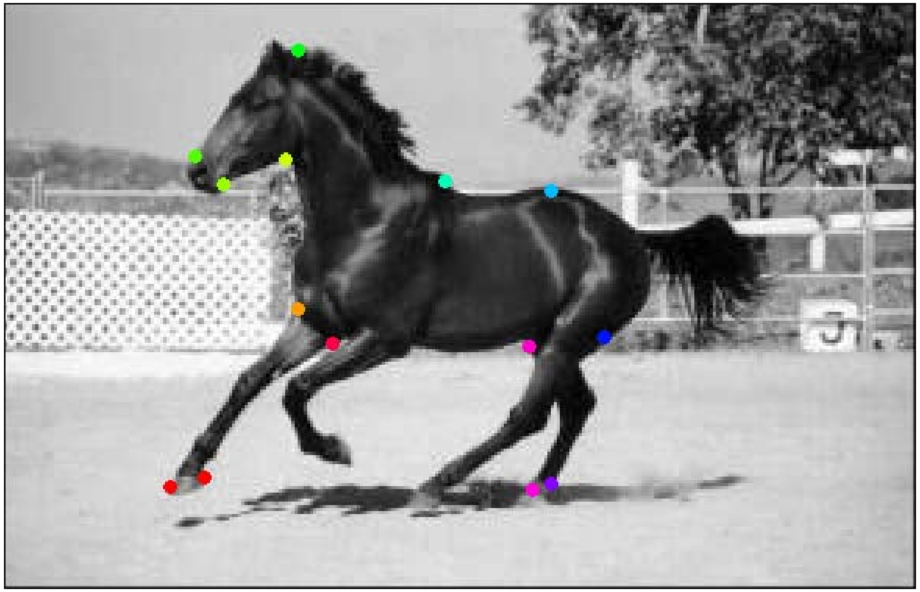


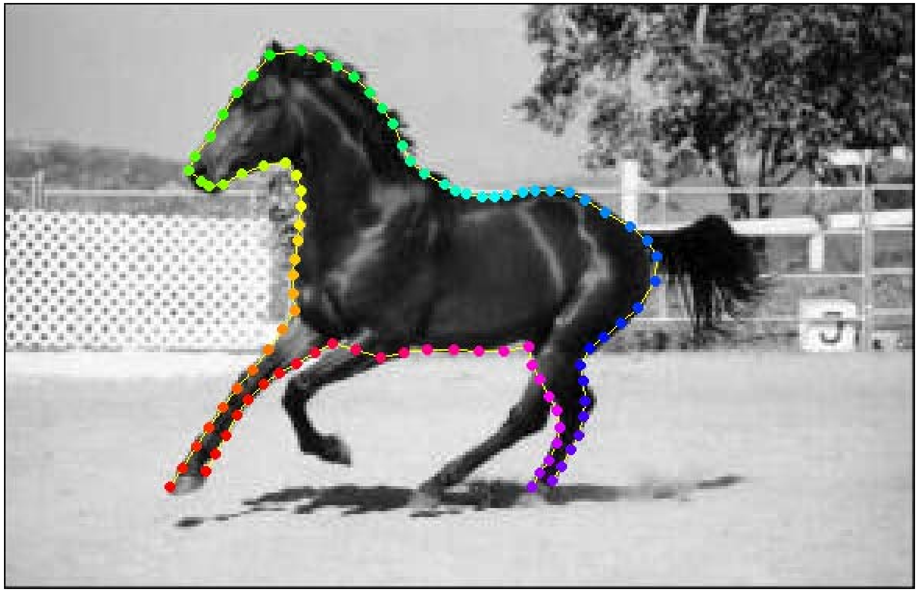
Error vs δ for cows

Weizmann Dataset

- 327 horse images
  - Similar size and orientation
  - Boundary manually delineated
  - 50 train, 50 validation, 227 test
- Manual Annotation
  - 14 control points on each horse
  - Smooth curves btw control pts
  - 96 interpolated boundary points
  - Same legs annotated as Zhu et al, 2009

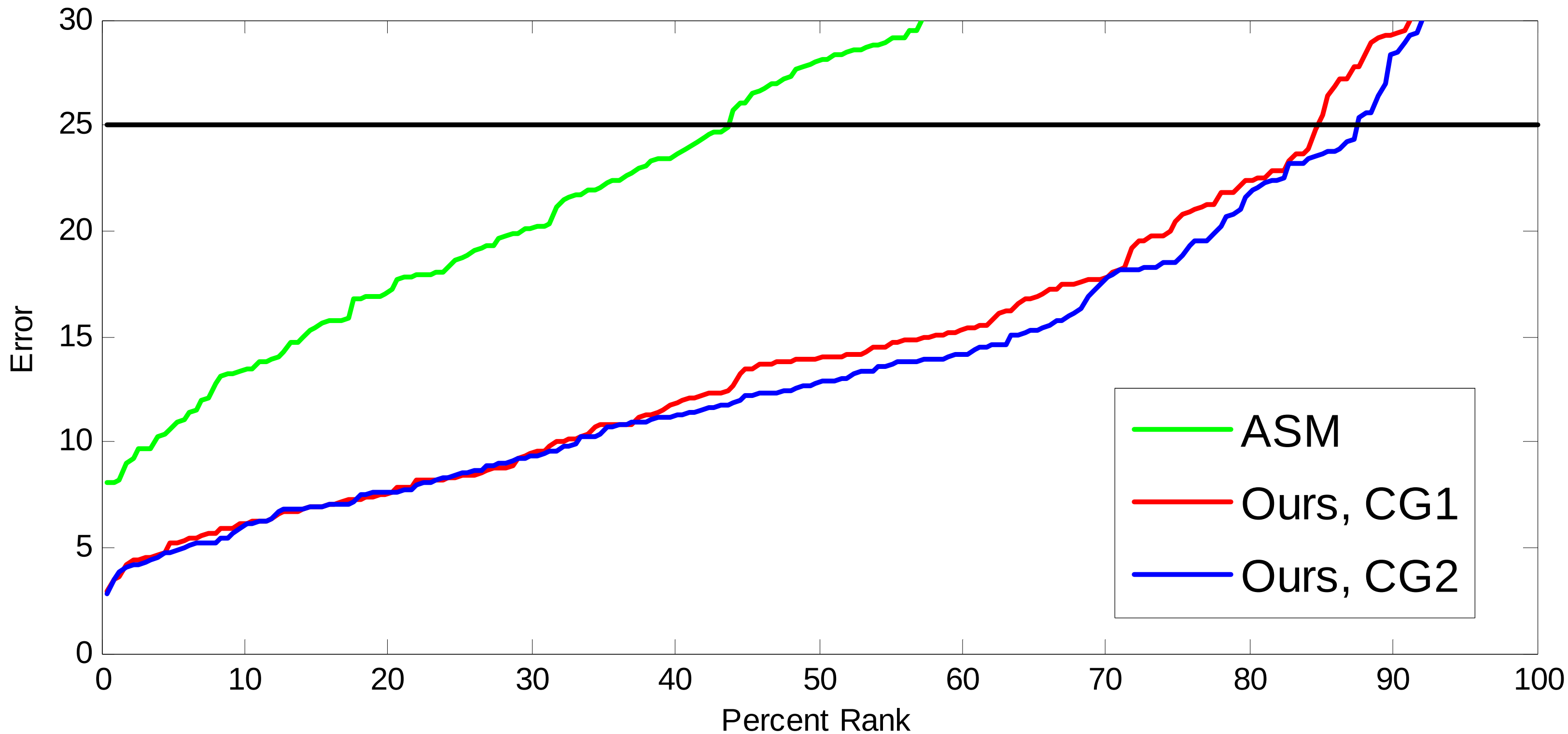




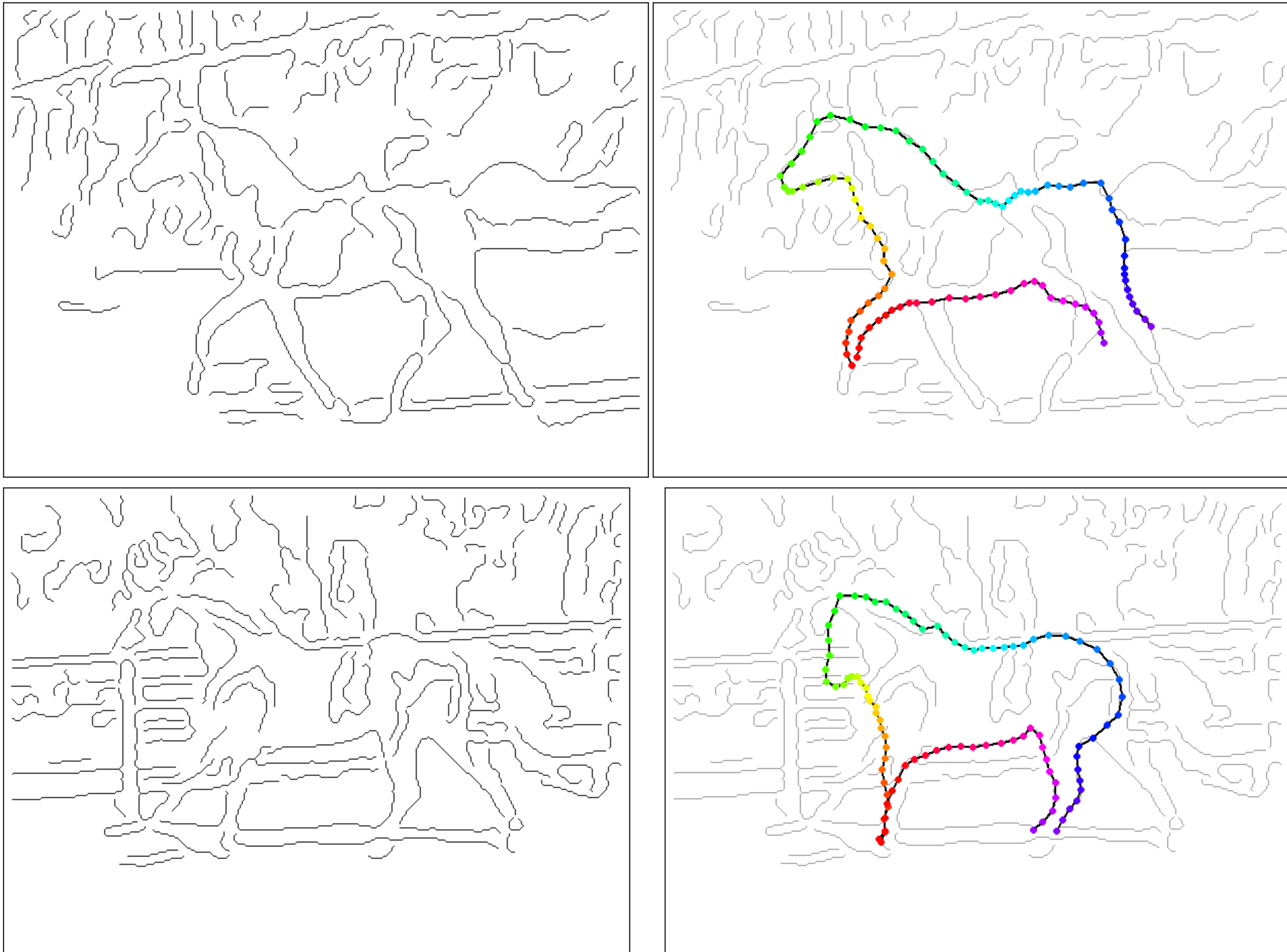


Quantitative Evaluation Horses

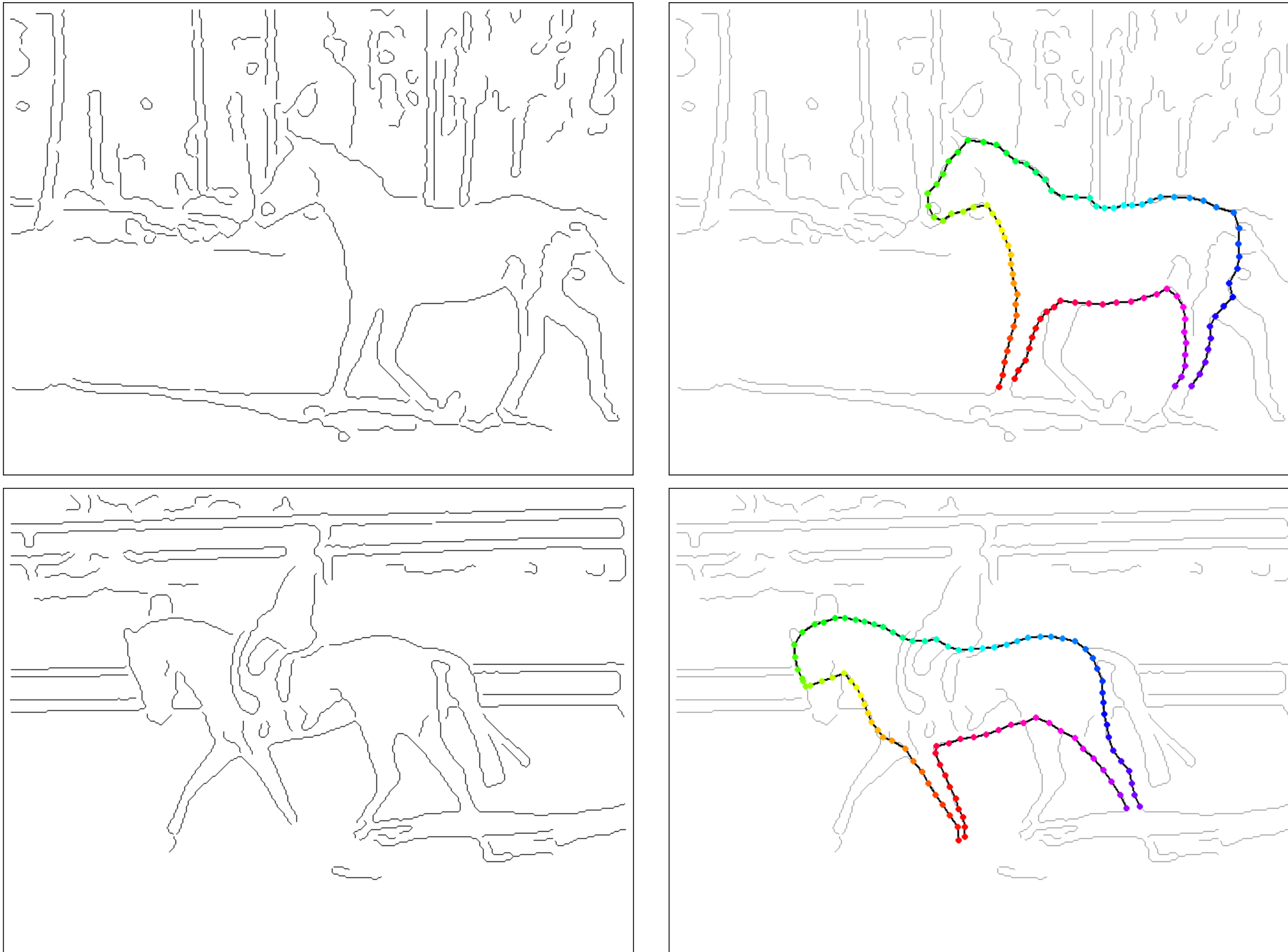
Method	Train images	Test images	Contour points	Train error	Test error	Time/img (sec)
ASM	50	227	96	25.35	29.05	<1
RCM	1	227	27	-	18.7	3
RCM	50	227	27	-	16.04	23
Ours, with CG1	50	227	96	12.79	15.58	44
Ours, with CG2	50	227	96	**12.74**	**15.36**	69
CG2 no head/legs	50	227	60	8.21	11.42	20



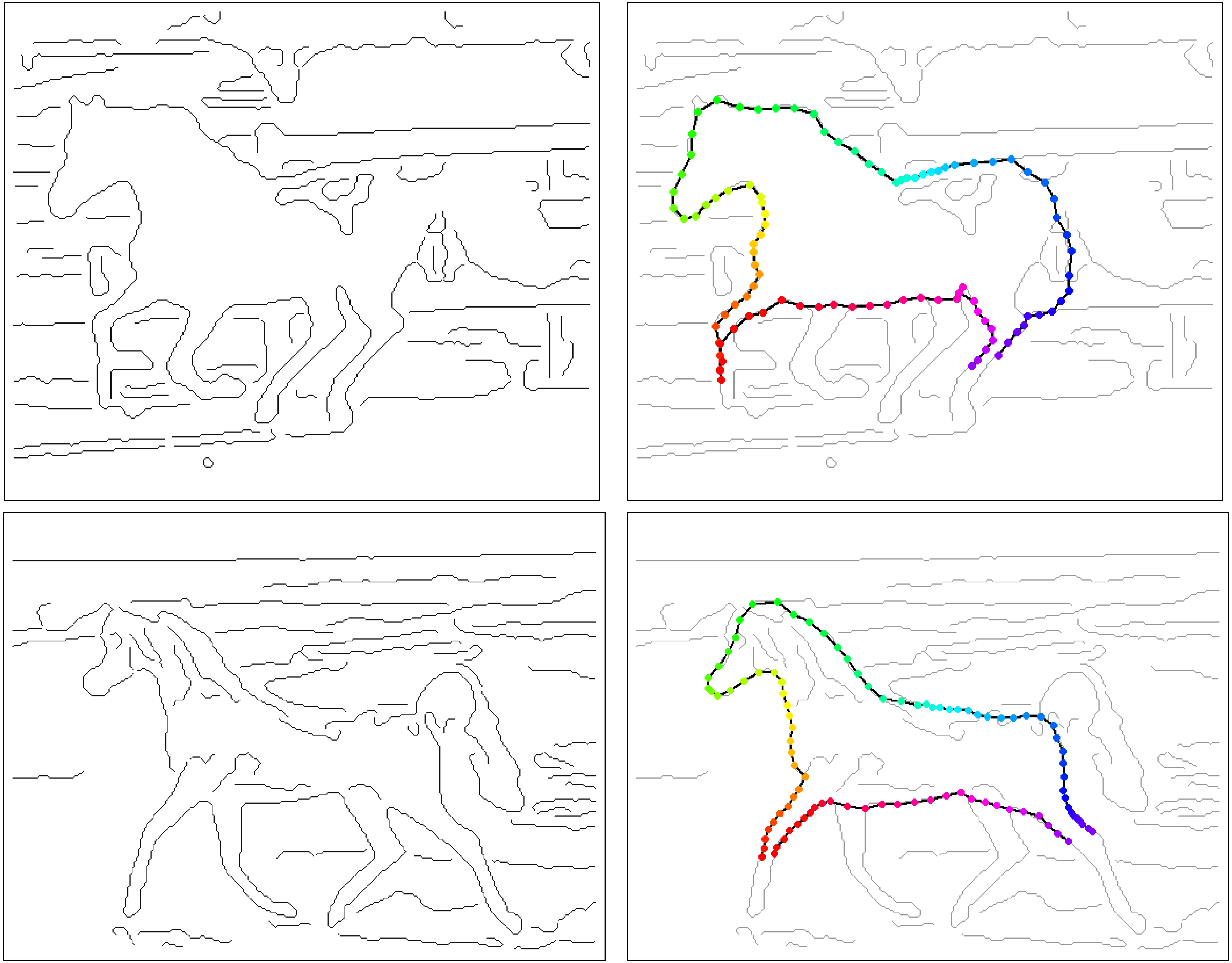
Example Results



Example Results

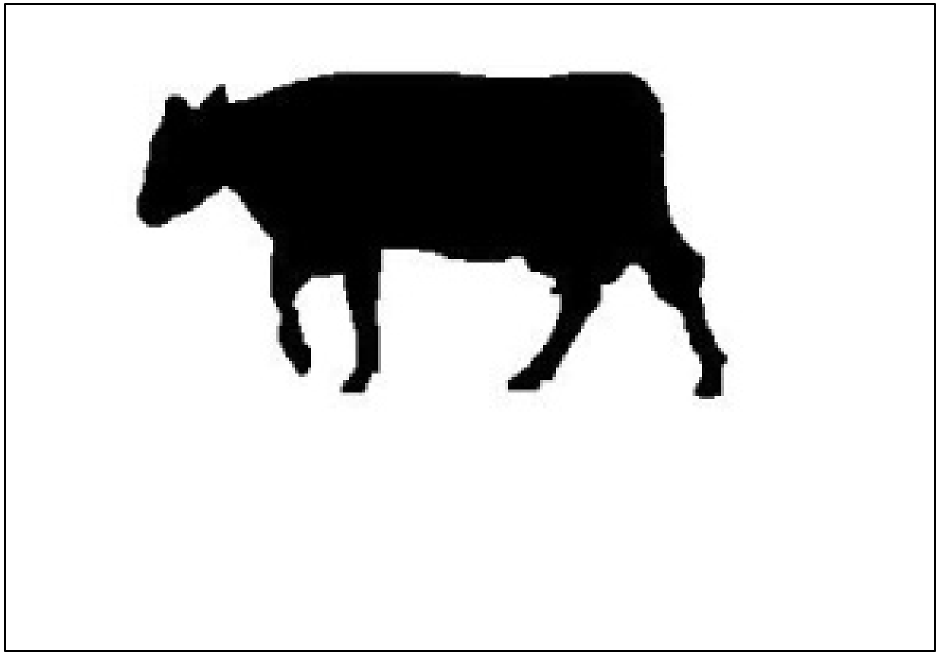


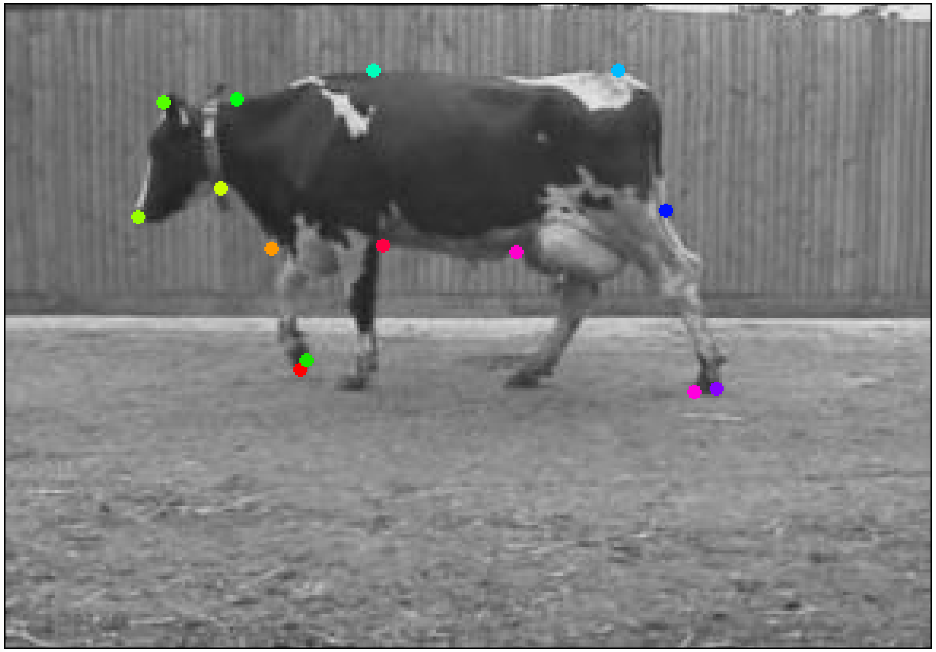
Example Results

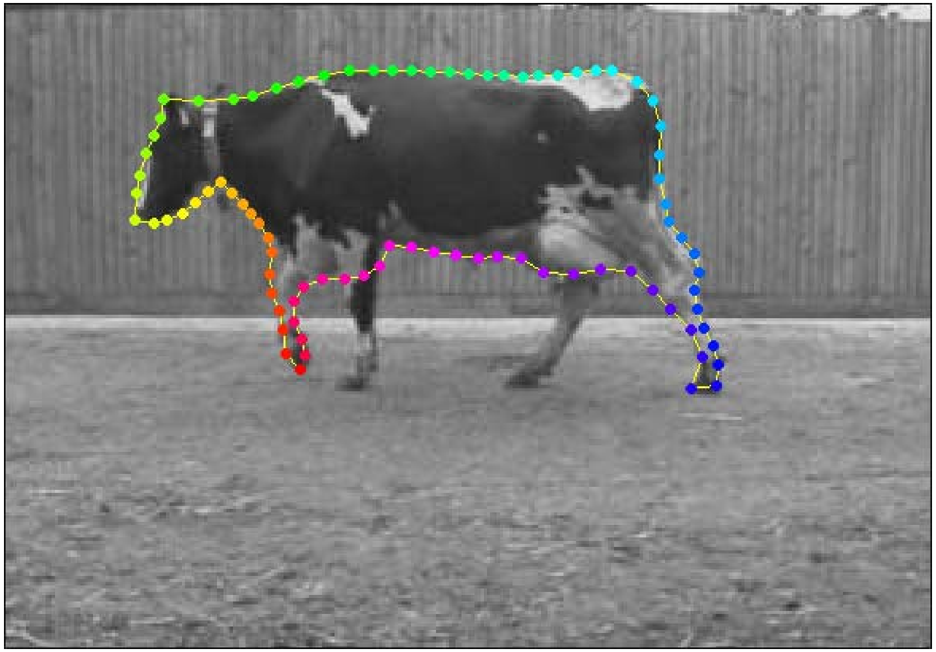


Cow Dataset

- 111 Cow images
  - Binary manual segmentation
- First 25 images for training
- 87 Point annotation

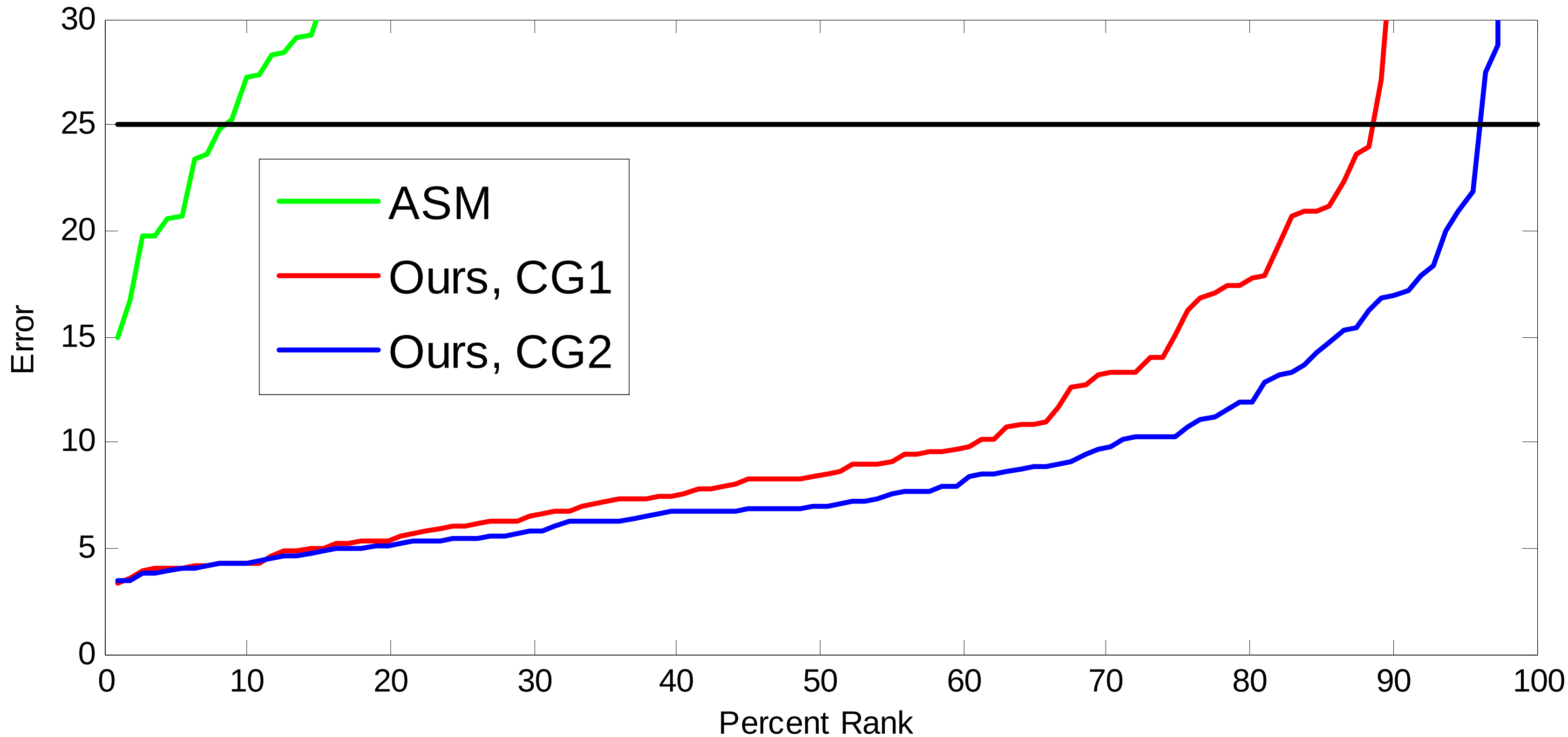




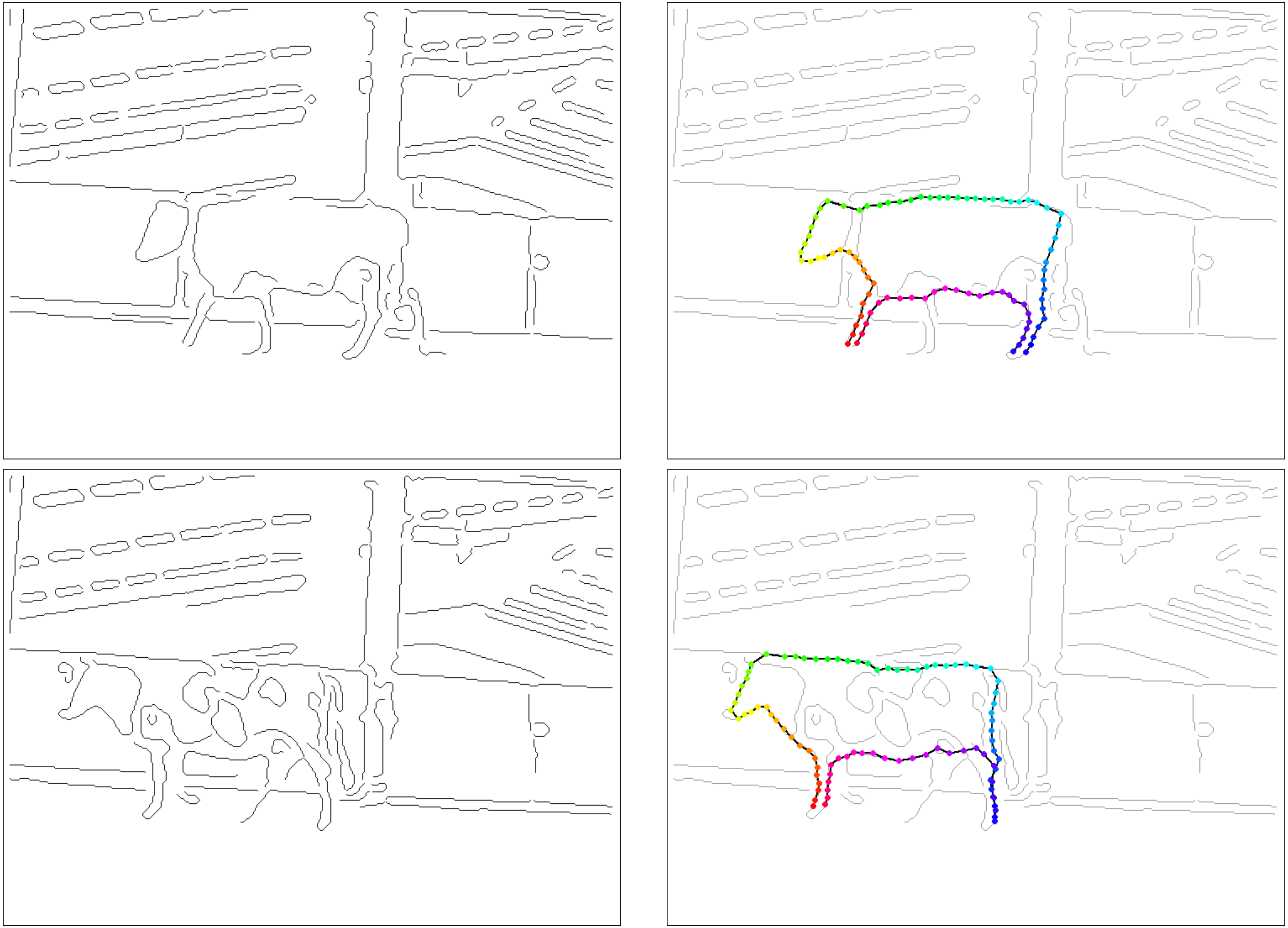


Quantitative Evaluation Cows

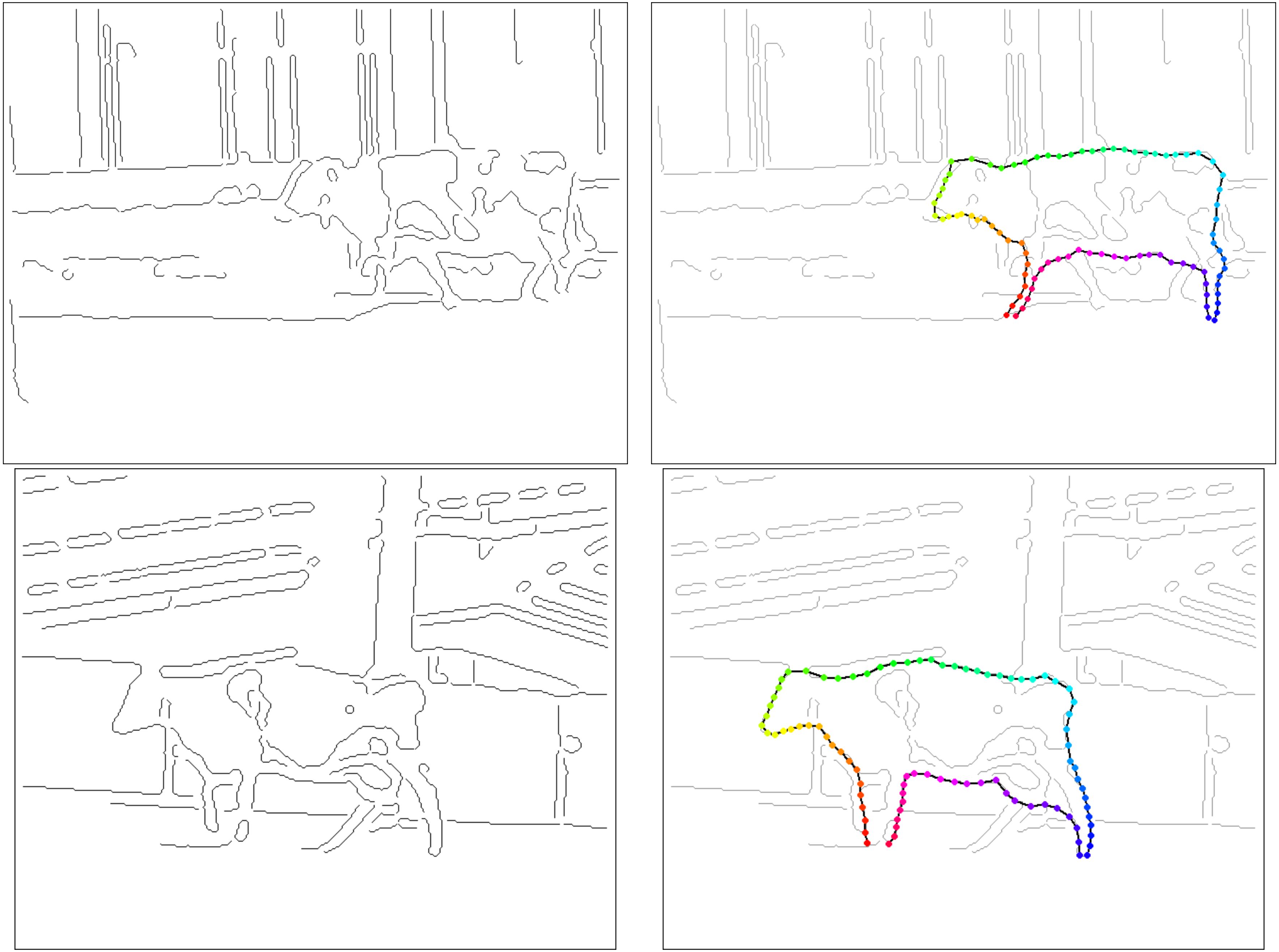
Method	Train images	Test images	Contour points	Train error	Test error	Time/img (sec)
ASM	25	111	87	48.81	49.23	<1
RCM	1	111	27	-	15.8	3.5
Ours, with CG1	25	111	87	13.78	14.98	14
Ours, with CG2	25	111	87	**11.73**	**10.81**	28



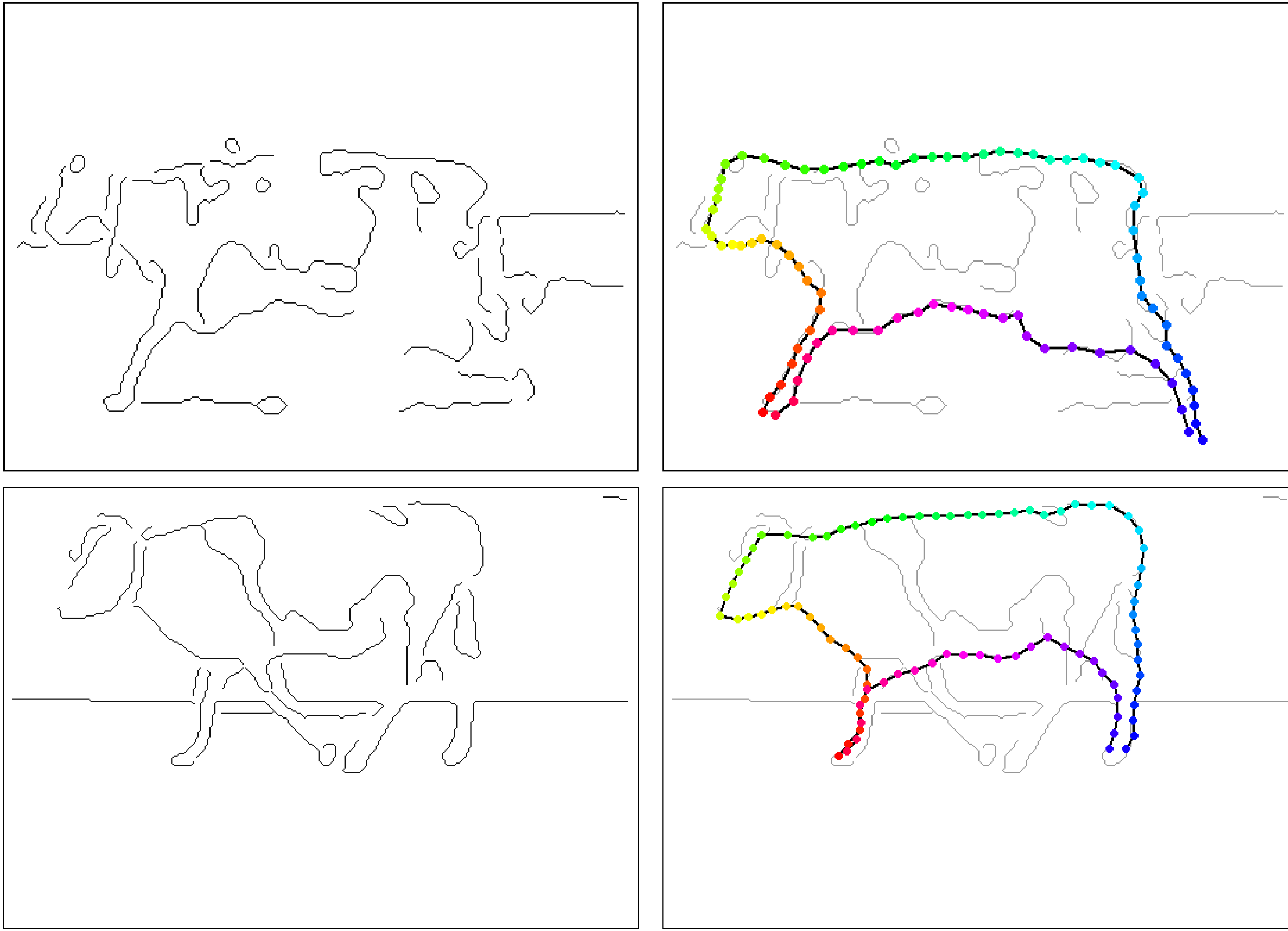
Example Results



More Results

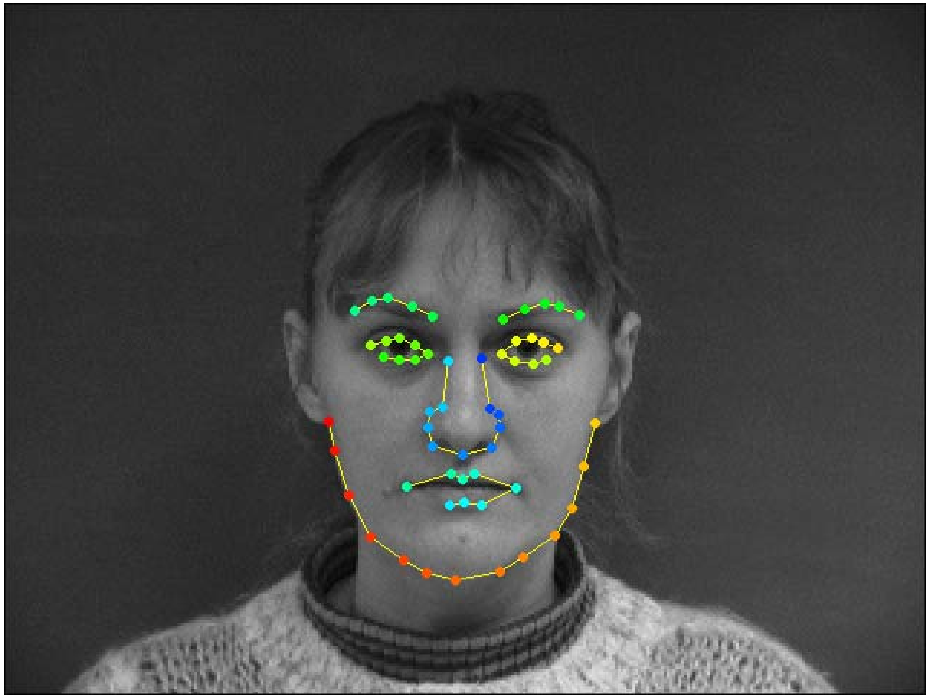


More Results



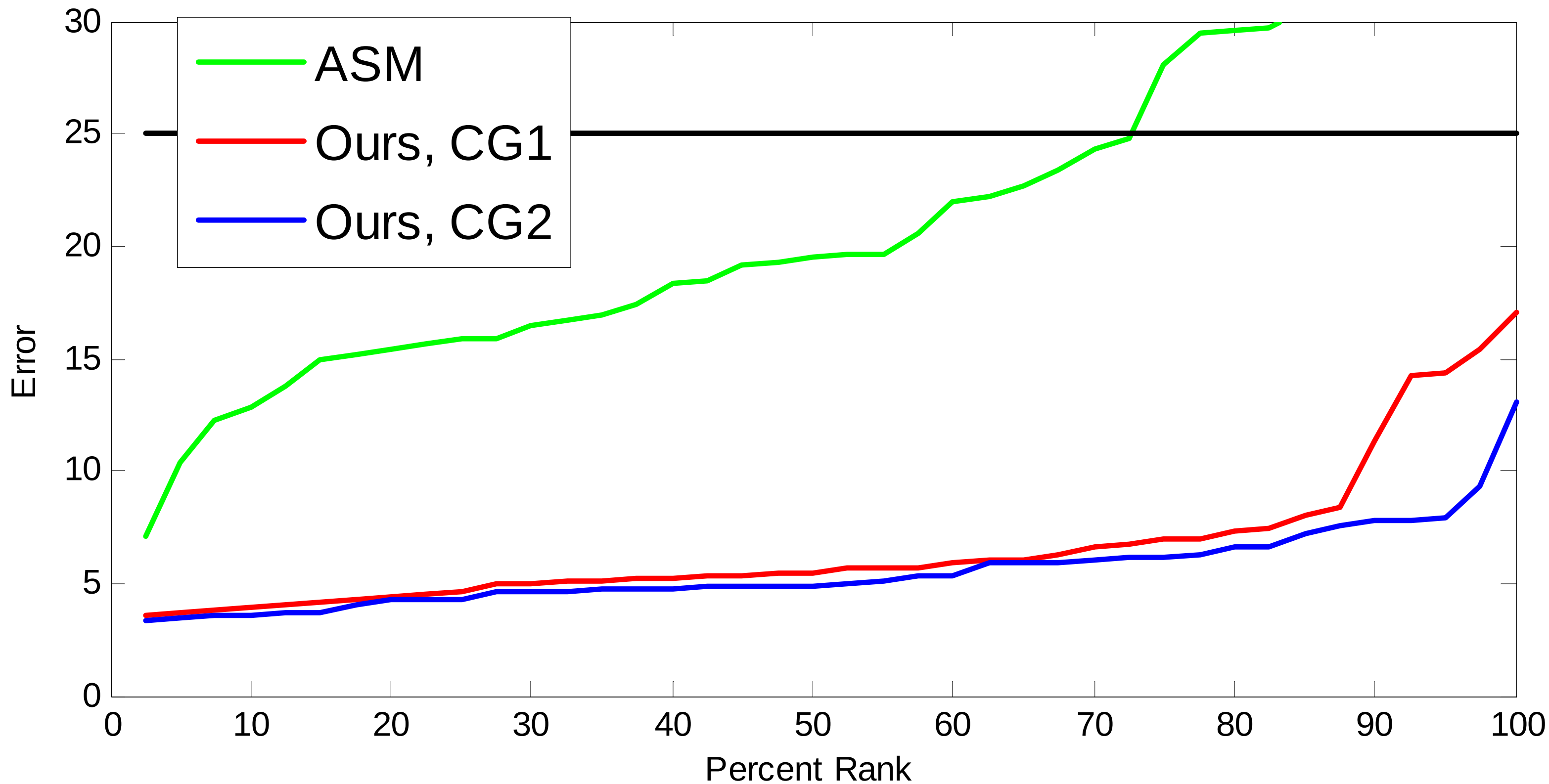
IMM Face Dataset

- Stegman et al, TMI 2003
- 40 frontal face images
  - 58 Landmarks
- Cross-validation



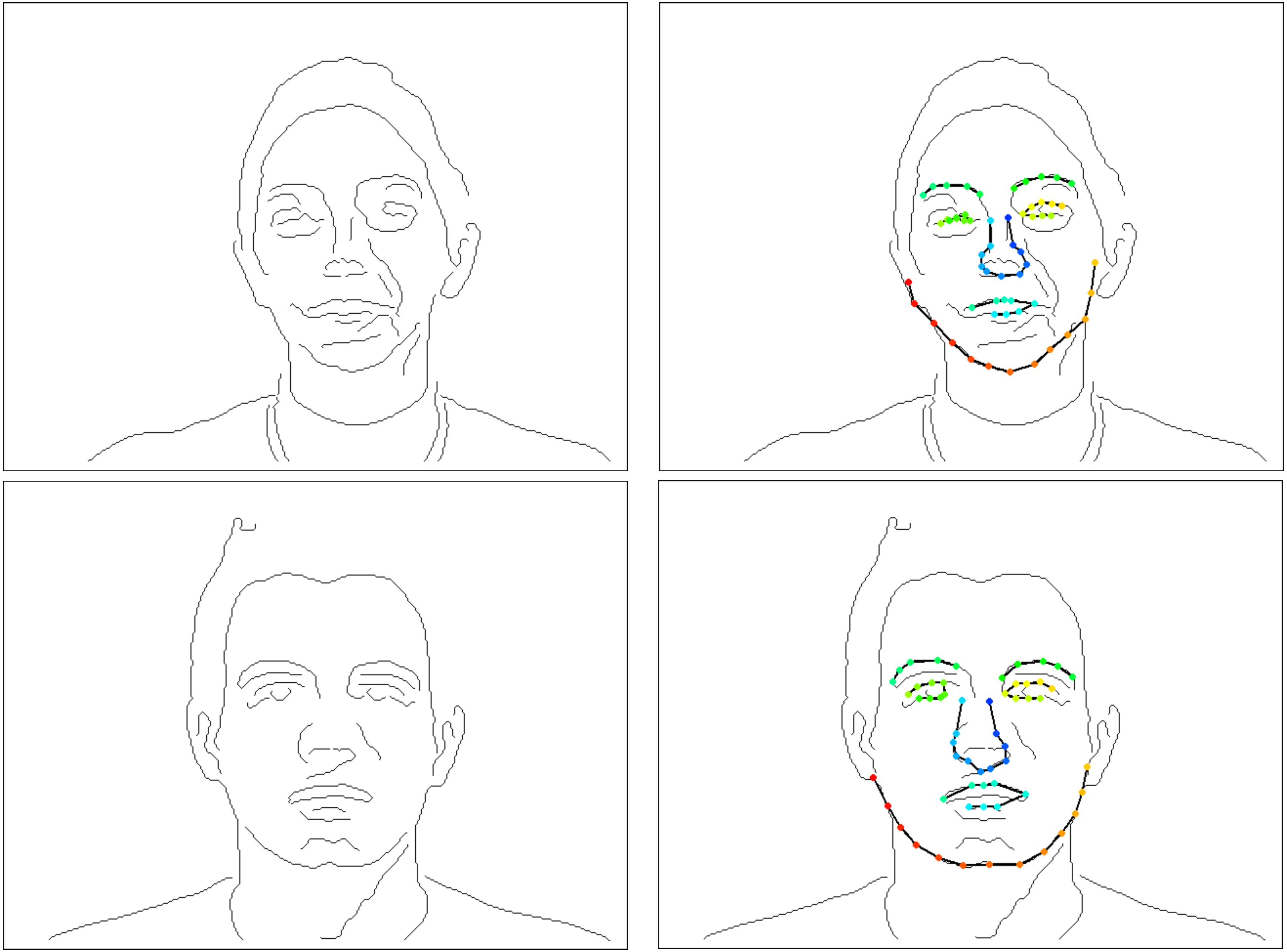
Quantitative Evaluation Faces

Method	Uses intensity	Automatic init.	X-val. folds	Train error	Test error	Time/img (sec)
ASM	No	Yes	4	21.47	21.56	0.08
Stegman	B/W	No	37	-	3.14	0.13
Stegman	Color	No	37	-	**3.08**	0.28
Ours, with CG1	No	Yes	4	6.54	6.64	0.33
Ours, with CG2	No	Yes	4	5.30	5.57	0.43



Example Results

Example Results



Conclusion

- A simple PCA model+ MRF deformation
  - Accurate
  - Can be used for object parsing from point clouds
- Local optimization initialized at good locations
  - Data-driven method for generating candidate locations
- Competitive with state of the art in object parsing
  - Not using any intensity information

Future Work

- Better shape model
  - Part based model plus deformation
- Shape deformation beyond normals.
  - Allow some control points to move in 2D (Kainmueller et al, MICCAI 2010)
- Use intensity information
  - Learning-based data term
- 3D Object Parsing
  - Parsing 3D faces from 2D Images
  - 3D Liver or spleen segmentation in CT/MRI

References

- A. Barbu. Hierarchical Object Parsing from Structured Noisy Point Clouds. To appear in *IEEE Trans. PAMI* 2013
- T. Cootes, C. Taylor, D. Cooper, J. Graham, et al. Active shape models-their training and application. *CVIU*, 61(1):38–59, 1995.
- Y. Li, L. Gu, and T. Kanade. A robust shape model for multiview car alignment. *CVPR*, 2009.
- M. B. Stegmann, B. K. Ersboll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. Medical Imaging*, 22(10):1319–1331, 2003.
- L. Zhu, Y. Chen, and A. Yuille. Learning a hierarchical deformable template for rapid deformable object parsing. *IEEE Trans. PAMI*, 2009.