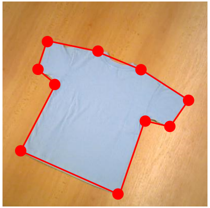
Garment perception and its folding — overview

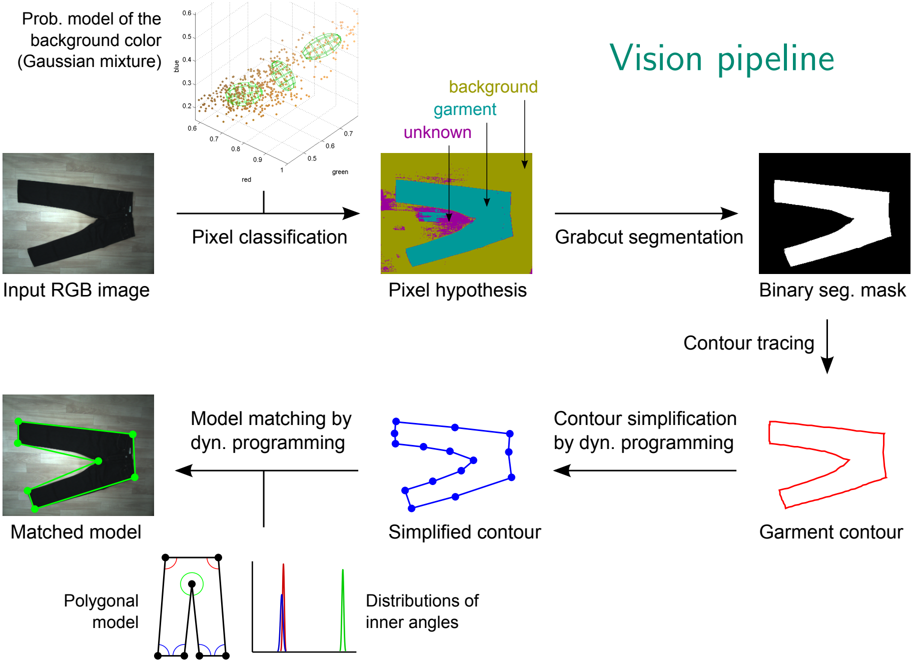
- Problem: Automated pose estimation of the garment spread on the table from a single RGB image followed by its folding using a dual-arm robot

- Solution:
  - Computer vision pipeline for the garment landmarks estimation. The pipeline includes the garment segmentation, its contour extraction, the contour simplification and matching of a polygonal model to the contour.
  - Positions of the landmarks are used to plan a folding move which is performed by our dual-arm robot.
  - Pose of the garment is checked after each fold.
  - State of the art results in terms of accuracy and speed.

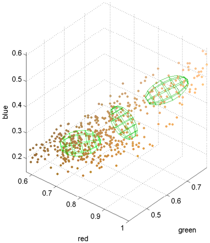
J. Stria, D. Průša, V. Hlaváč et al.

Vision pipeline



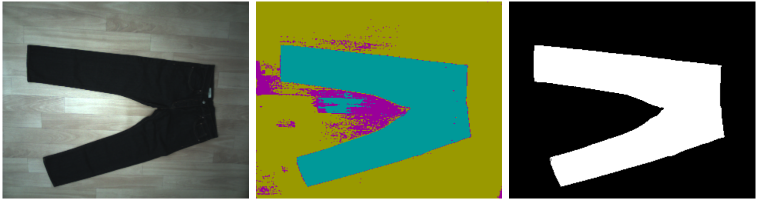
Learning the table color model

- Assumption: Color of the table is known, invariant and dissimilar from the color of the garment.
- We learn a probabilistic model of RGB triples from training images of the table. The distribution is modeled by the mixture of 3D Gaussians (GMM).
  - RGB triples are split into components using the palette design algorithm [Orchard and Bouman, 1991].
  - Prior probabilities, mean vectors and covariance matrices for individual components are estimated according to the maximum likelihood principle.



Segmenting an unknown image

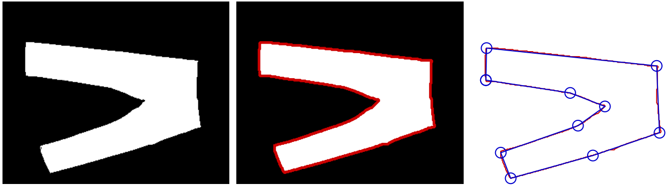
- While dealing with previously unseen image, we split its pixels into three disjoint sets based on their probability in the learned GMM of table color:
  - Garment (foreground) pixels have low probability in GMM.
  - Table (background) have high probability in GMM.
  - Unknown pixels have medium probability in GMM.



- The created trimap forms the input of the grabcut segmentation algorithm [Rother et al., 2004] which gives the final binary segmentation mask.

Contour processing

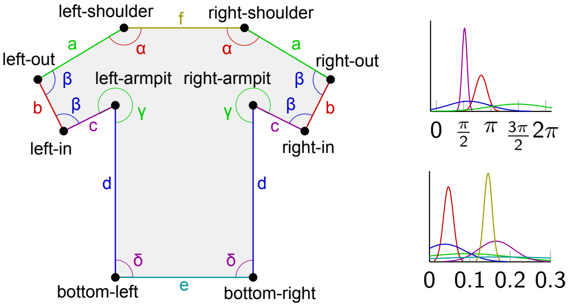
- Boundary of the segmentation mask is traced by Moore's algorithm [Gonzalez et al., 2009]. The resulting contour consists of several hundreds or thousands points corresponding to pixels of the segmentation mask.



- The contour is approximated by a polygon having at most tens of vertices.
  - Suboptimal algorithm based on dynamic programming approach for the optimal approximation of open curves [Perez and Vidal, 1994]
  - Minimization of the total sum of distances between the original contour points and edges of the approximating polygon

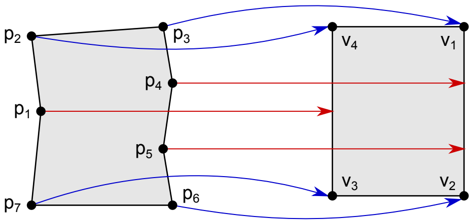
Polygonal models

- Various categories of clothes are represented by polygonal models. The model vertices correspond to the desired landmark points.
- Each model is appended distributions of inner angles adjacent to its vertices and distributions of relative lengths of its edges.
  - Modeled by independent normal distributions
  - Learned from training data by the maximum likelihood principle



Mapping contours to models

- The observed simplified contour is mapped to the polygonal model.
  - Mapping preserves clockwise ordering of contour points and model vertices.
  - Not all contour points are mapped to vertices. Some of them are mapped to the edges of the polygonal model.



- Each valid mapping is assigned certain cost.
  - Based on the similarity of the inner angles and the relative edge lengths between the observed contour and the polygonal model
  - The similarity is evaluated using the learned angles and edges distributions

Matching costs

- Vertex matching cost $V^m_{i,j,k}$ is defined for each triple of contour points p_i, p_j, p_k and each model vertex v_m:

$$V^m_{i,j,k} = -\lambda_V \log \mathcal{N}(|\angle p_i p_j p_k|; \mu_m, \sigma^2_m) \tag{1}$$

- Edge matching cost $E^m_{j,k}$ is defined for each pair of points p_j, p_k and each polygonal model vertex v_m:

$$E^m_{j,k} = -\lambda_E \log \mathcal{N}\left(\frac{\|\overline{p_j p_k}\|}{\sum_{i=1}^{n} \|\overline{p_i p_{i\oplus 1}}\|}; \nu_m, \tau^2_m\right) \tag{2}$$

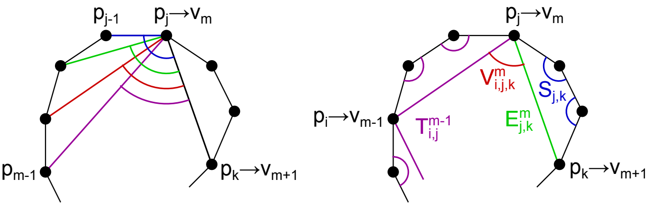
- Segment matching cost $S_{j,k}$ is defined for each pair of simplified contour points p_j, p_k in the following way:

$$S_{j,k} = -\lambda_S \sum_{i \in I_{j,k}} \log \mathcal{N}(|\angle p_{i\ominus 1} p_i p_{i\oplus 1}|; \xi, \phi^2) \tag{3}$$

Finding the optimal mapping

- The optimal mapping is found by dynamic programming.
- Iterative evaluation of the total cost $T^m_{j,k}$ for matching the contour points $p_1 \ldots p_k$ to the model vertices $v_1 \ldots v_{m+1}$ so that it holds $f(p_j) = v_m$ and $f(p_k) = v_{m+1}$ and points mapped to the first and last vertex are fixed.

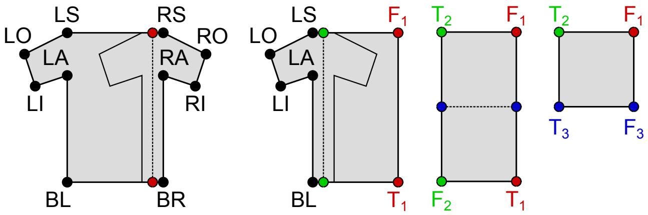
$$T^m_{j,k} \leftarrow \min_{i \in \{m-1 \ldots j-1\}} \left(T^{m-1}_{i,j} + V^m_{i,j,k} \right) + E^m_{j,k} + S_{j,k} \tag{4}$$



- Polynomial time complexity vs. exponential number of possible mappings

Folded models

- Pose of the garment is checked after each fold by matching its contour to the folded polygonal model. The folded models are derived automatically from the original model based on the planned fold.
- The original model and the derived models after the first ($F_1 \rightarrow T_1$), second ($F_2 \rightarrow T_2$) and third ($F_3 \rightarrow T_3$) fold:



- Distributions of the inner angles and relative lengths for the folded polygonal models are adjusted to the actual garment.

Vision experiments

- The polygonal models were trained on 340 images. The proposed method was tested on 170 images (41 towels, 45 pants, 45 short-sleeved shirts, 39 long-sleeved shirts).
  - Dataset acquired by ASUS Xtion camera
  - Resolution 1280×1024
  - Manually annotated ground truth positions of the landmarks

- The whole pipeline takes 1.5–4.5 seconds for a single image.
  - On average 0.83 second for segmentation
  - 0.5–3.5 seconds for contour processing
  - On average 0.14 second for model mapping
  - Significant improvement compared to 30–150 seconds needed by the past state of the art method [Miller et al., 2011]
- 165 out of 170 images were correctly segmented which is 97%.

Precision of the matching procedure

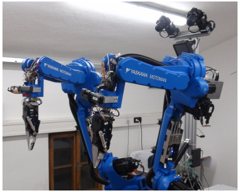
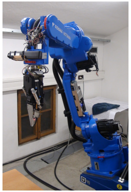
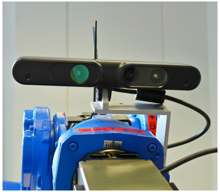
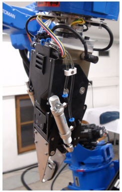
- Displacements of the vertices found by model matching and the manually annotated landmarks projected to the canonical images:

- Displacements for individual categories of clothes (in centimetres):

Error	Towel	Pants	Short-sleeved	Long-sleeved
Median	0.31	0.46	0.49	0.50
Mean	0.35	0.52	0.88	1.03
Std. dev.	0.03	0.11	1.17	2.14

- Improved by 20% compared to the original method [Stria et al., 2014]

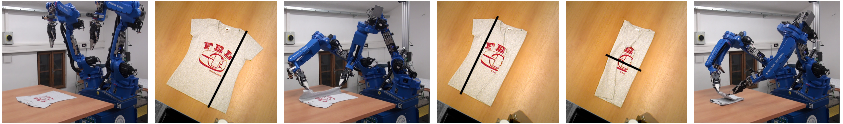
CloPeMa robot

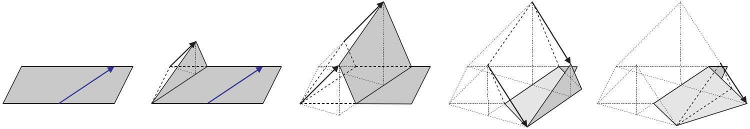
- Stereo vision head (two Nikon DSLR cameras D5100)
- The robot body utilizes components used in welding industry supplied by YASKAWA Motoman: two arms MA1400 are attached to the turntable R750 and powered by two control units DX100.
- Both arms equipped with combined color and depth sensors (ASUS Xtion)
- Self-made grippers with wrist force/torque sensors, finger tactile sensors and macro scale photometric stereo
- The robot is operated by the Robotic Operating System (ROS).

Robotic folding

- The identified landmark points are used to plan a folding move.



- The planning utilizes RRT-Connect algorithm [Kuffner and LaValle, 2000] to find collision free trajectories for both arms.
- The inverse kinematics is computed by Open Motion Planning Library.
- Gravity based folding proposed by [van den Berg et al., 2011]
  - In each moment, part of the garment is hanging vertically, part of it is laying horizontally on the table. Both parts are immobilized.
  - The grippers follow triangular paths (upwards and then downwards).



Folding videos (speeded up 6x)

CloPeMa project

- CloPeMa = Clothes Perception and Manipulation
- Funded by the European Comission in FP7.
- 3 years (February 2012 – January 2015)
- Web: `www.clopema.eu`
- Consortium members:
  - Czech Technical University in Prague
  - Centre for Research and Technology Hellas
  - Universita Degli Studi di Genova
  - University of Glasgow
  - Neovision Ltd.
- The main objective of the project is to advance the state of the art in the autonomous perception and manipulation of fabrics, textiles and garments.

References I

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2009).
Digital Image Processing Using MATLAB, 2nd ed.
Gatesmark.

Kuffner, J. J. and LaValle, S. M. (2000).
RRT-connect: An efficient approach to single-query path planning.
In *Proc. Int. Conf. on Robotics and Automation (ICRA)*.

Miller, S., Fritz, M., Darrell, T., and Abbeel, P. (2011).
Parametrized shape models for clothing.
In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*.

Orchard, M. and Bouman, C. (1991).
Color quantization of images.
IEEE Trans. on Signal Processing, 39(12).

References II

Perez, J.-C. and Vidal, E. (1994).
Optimum polygonal approximation of digitized curves.
Pattern Recognition Letters, 15(8).

Rother, C., Kolmogorov, V., and Blake, A. (2004).
Grabcut – interactive foreground extraction using iterated graph cuts.
ACM Trans. on Graphics, 23(3).

Stria, J., Průša, D., and Hlaváč, V. (2014).
Polygonal models for clothing.
In *Proc. Towards Autonomous Robotic Systems (TAROS)*.

van den Berg, J., Miller, S., Goldberg, K., and Abbeel, P. (2011).
Gravity-based robotic cloth folding.
In *Proc. Int. Workshop on the Algorithmic Foundations of Robotics (WAFR)*.

Contacts

- Garment Perception and its Folding Using a Dual-arm Robot
- Jan Stria, Daniel Průša, Václav Hlaváč, Libor Wagner, Vladimír Petrík, Pavel Krsek, Vladimír Smutný
- Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
- striajan@fel.cvut.cz
- http://cmp.felk.cvut.cz/∼striajan/iros2014