State of the Art

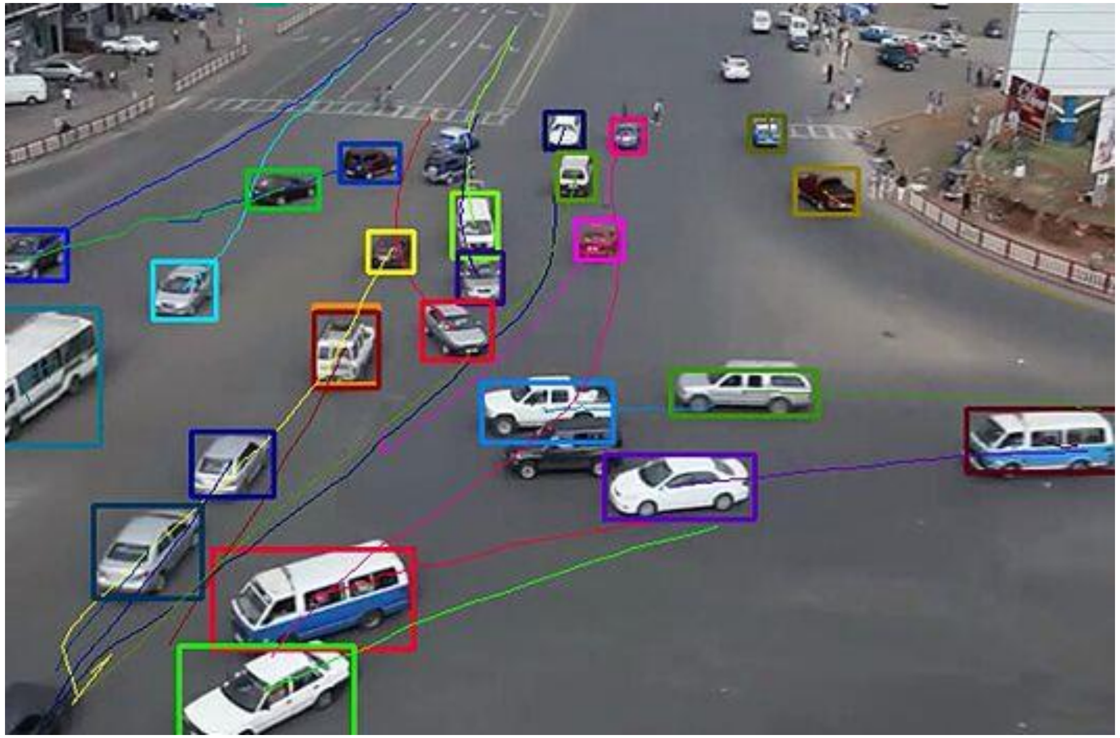
Ben Bartlett

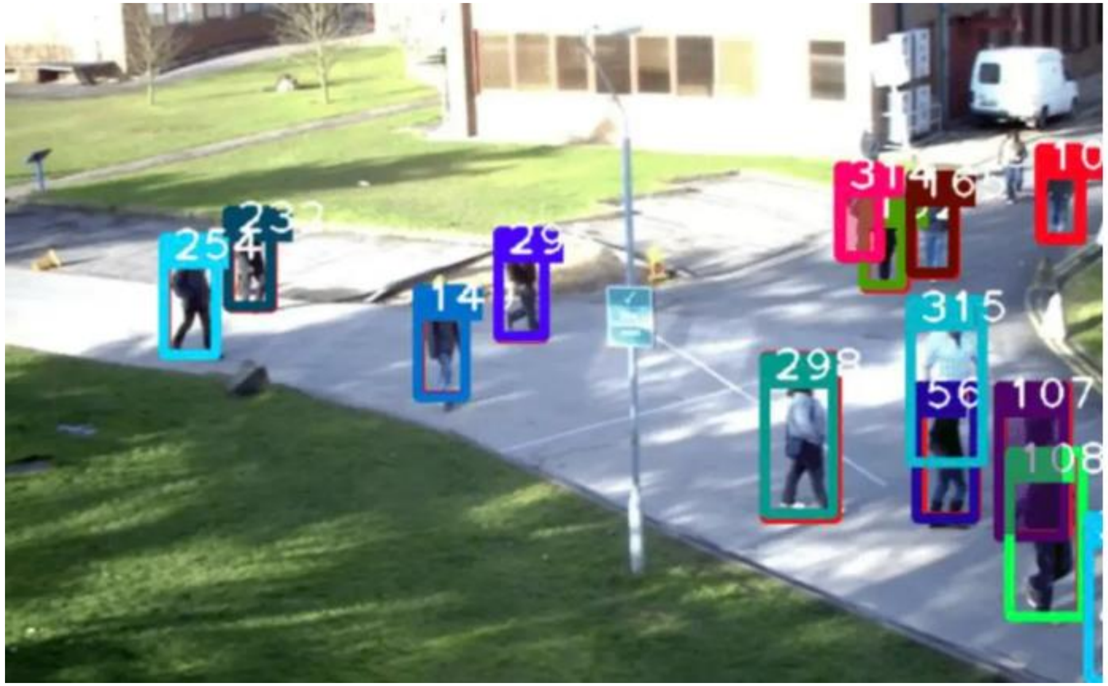
Visual Tracking in Computer Vision

Aimed at estimating the trajectory of an object in a video sequence. "Detections over time and their relationship"

Key Challenges:

- Changes in illumination
- Occlusion
- Motion Blur
- Presence of similar objects in the background




254
14
29
298
315
56
107
108

Visual Tracking

Typical algorithms:

- Correlation Filter-based Trackers (MOSSE, KCF)
- Deep Learning Based Trackers (SiamFC, SiamRPN, FairMOT)
- Optical Flow Methods (Farneback, Lucas-Kanade)
- Probabilistic methods (Kalman Filter, Particle Filter, DeepSORT)

Deformable Object Tracking

Examples:

Cables, Cloth, Soft tissues

Methods:

Model Based / Model Free:

- Model Based
- Template Matching

Feature Based

Deep learning

Tracking a Deformable Cable with RGB-D Cameras

Possible Methods

- Optical Flow with depth information
- Pose Estimation
- 3D reconstruction
- Physics-Based Simulation
- Graph based model
- Neural Network

Literature review from PhD thesis

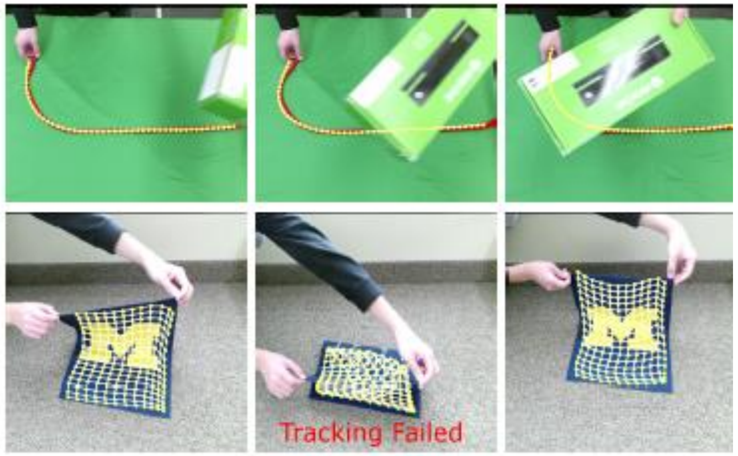
- ICP for matching point clouds, assumes the object didn’t move, typical in LiDAR/Photogrammetry model matching/joining
- SIFT for matching images, frame to frame, assumes the object is purely related by scale, rotation and illumination transformation
- Harris Corner Detector similar but only performs well on rotation differences not so much, scale
- PWP3D known 3D model detection, model projected into scene and rotated/translated until a match occurs.

Literature review from PhD thesis

- CPD – similar idea to ICP but represent point set while maintaining some structural coherence. Completed using Gaussian Mixture Models. While not physics based, an initial connectivity model is needed as input.

Paper:

- Occlusion-robust Deformable Object Tracking without Physics Simulation



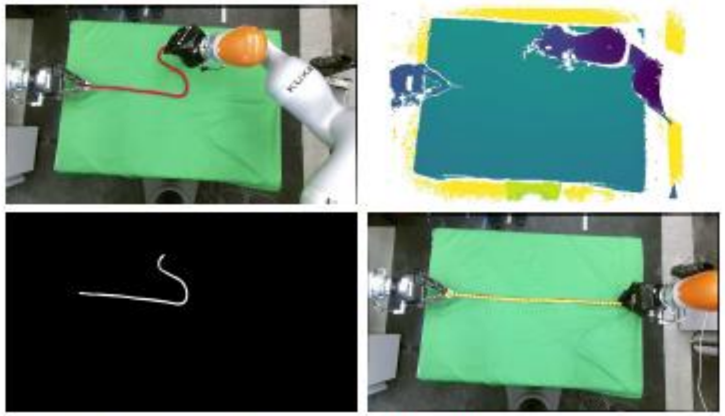


Fig. 2. The input for each time step. TL: RGB image. TR: depth image. BL: Mask image. BR: The input model for first time step, where yellow dots are vertices and white lines are edges.

Literature review from PhD thesis

- Growing Neural Gas network for tracking by an element of prediction trained on frame-to-frame data. Removes need for Kalman filter/nearest neighbour

To note: **shape deformation monitoring** better search term than deformable object tracking is suggested where this involves pedestrians, faces etc

Paper:

- Soft Object Deformation Monitoring and Learning for Model-Based Robotic Hand Manipulation

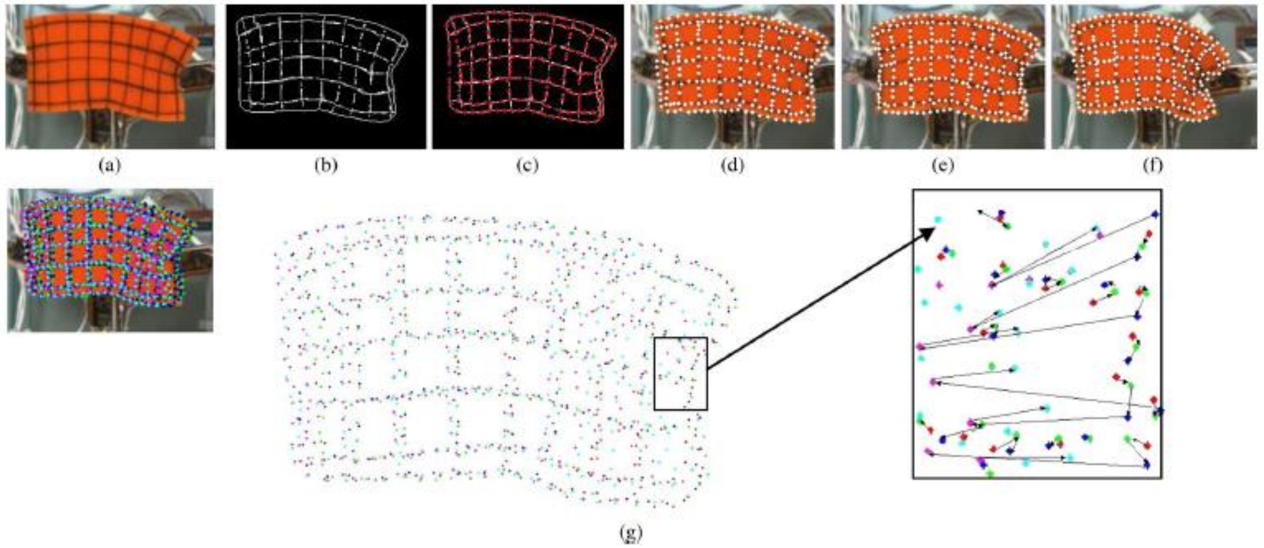


Fig. 9. Contour and grid point tracking for the orange sponge: (a) Initial frame, (b) grid and contour points, (c) initial growing neural gas, (d)–(f) tracking using a sequence of neural gas networks, and (g) trajectory of tracked points and detail of the one-to-one correspondence of tracked points.

Literature review from PhD thesis

- Dynamic model input robust tracking with occlusion, use of kinect

<u>Paper:</u>

- Real-time State Estimation of Deformable Objects with Dynamical Simulation

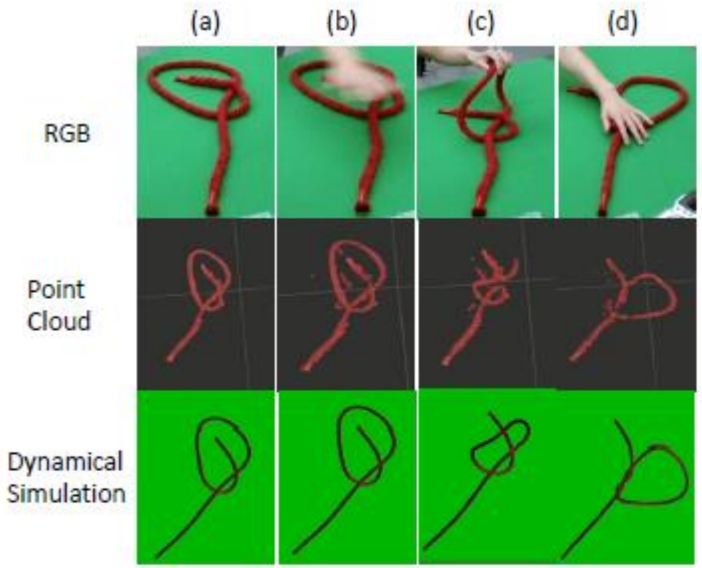


Fig. 6. With point cloud recovery. Tracking is robust to large occlusion.

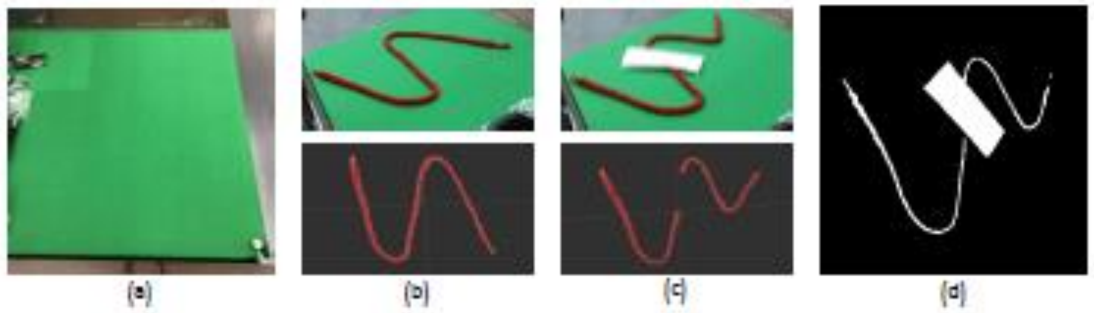


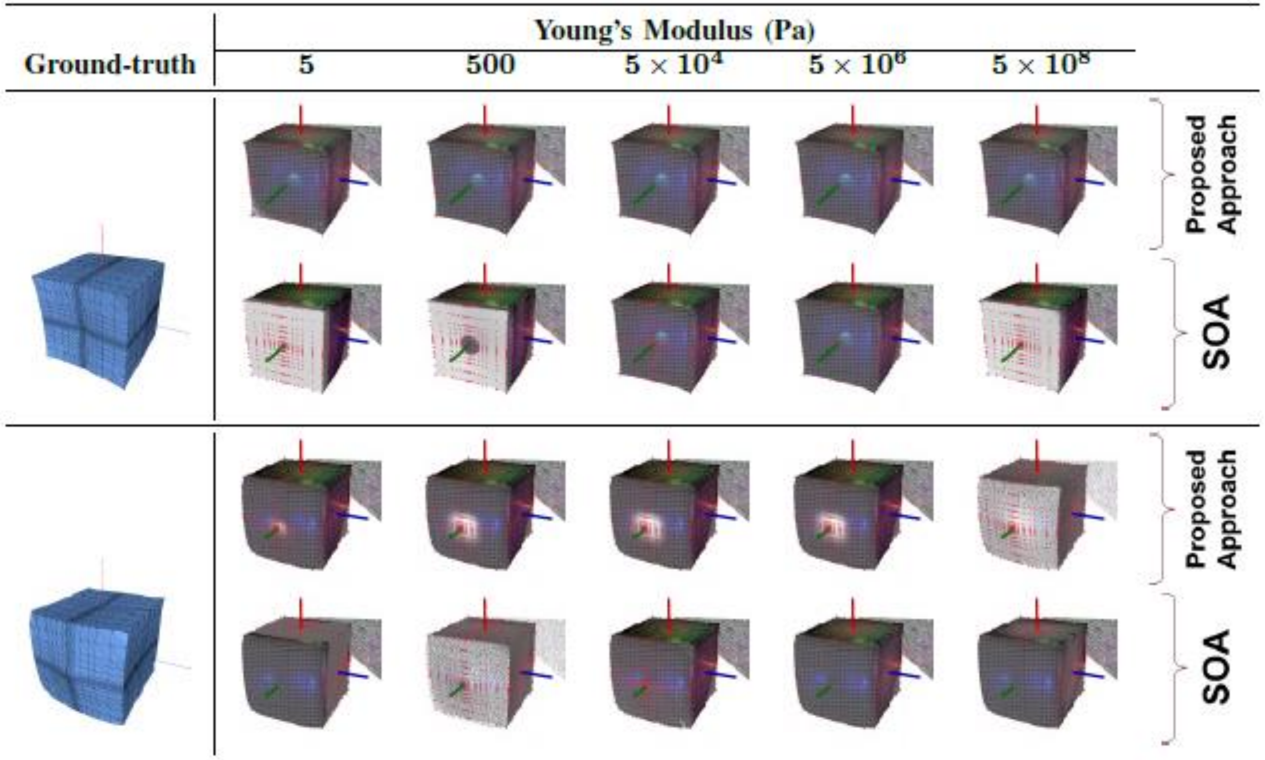
Fig. 3. (a) Background. (b) $t-1$ frame. (c) t frame. (d) Foreground mask.

Literature review from PhD thesis

- Onto Model based, coarse model, FEM simulated with closed loop optimization with real object feedback

<u>Paper:</u>

- Tracking of Non-Rigid Objects using RGB-D Camera



(a) Tracking deformation of the *cube* across different values for Young's modulus

Literature review from PhD thesis

- RGB_D Mass-Spring Model used in conjunction with real time mesh to estimate the deformation without the need for well textured objects

Paper:

- Real-time Deformation, Registration and Tracking of Solids Based on Physical Simu

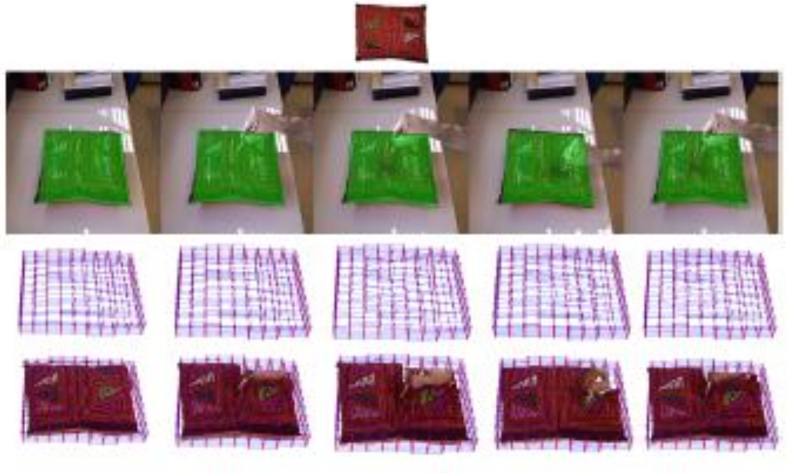


Figure 6: Visual results of 3D recovery shapes: *Teddy Bear* (a), *Sponge Bob* (b), *Ball* (c) and *Cushion* (d). For each of the four examples: the model, the original image with the projection of the recovered 3D mesh (first row), the recovered 3D mesh in terms of the physics model (second row) and the same recovered mesh with the raw data acquired from the depth camera (third row).

General Literature Study

- Visual Tracking of Deformable Objects
- Visual surface deformation monitoring
- Template Matching
- RGBD
- Physics based deformable object tracking
- Non rigid shape tracking
- Non physics based deformable object tracking
- FEM method tracking
- Deformable Object in video
- Deep Learning
- Cable Specific

Trends in literature

Use cases:

- Manipulation tasks
  - Pizza making robots
  - Biomedical stitching
  - Knot tieing
  - Panel Box cable detection
- Objects utilized/Look at
  - Humans
  - Dough
  - Sheets (plastic, cloth)
  - Cables/connectors

Visual Tracking of Deformable Objects

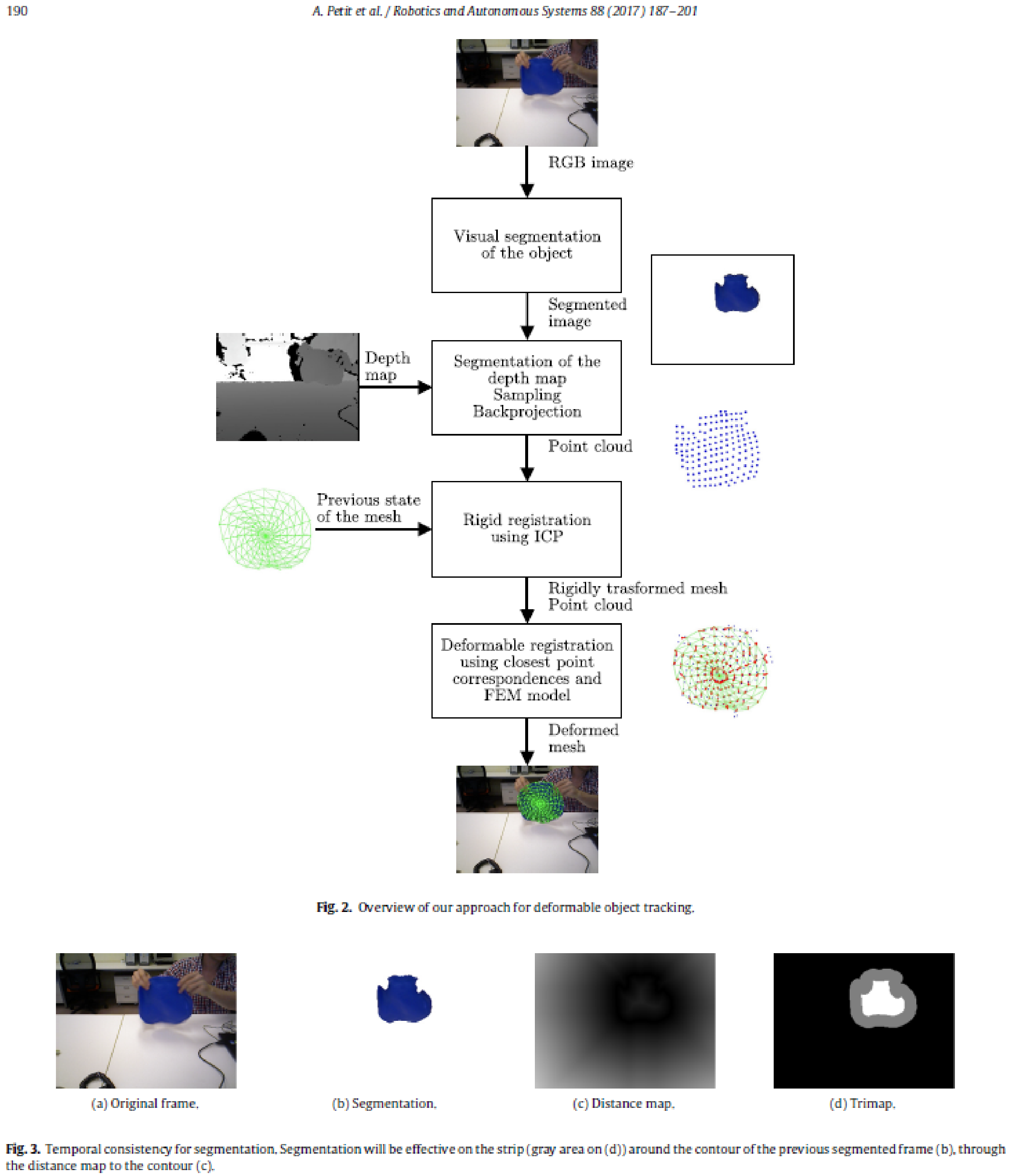
- Tracking Deformable Objects with Point Clouds
    - Application of physics models to track across a sequence of point clouds

Visual Tracking of Deformable Objects

- Object Tracking Using Deformable Templates
  - Prototype based deformable template for tracking. Manual initialization of template near object, matching process completed until convergence
  - Discussion of this method, vs snake model

RGBD

- Tracking elastic deformable objects with an RGB-D sensor for a pizza chef robot
    - Strong workflow with visual segmentation to segment the point cloud. Utilization of the last mesh for rigid ICP matching followed by deformable registration with FEM model.



Non-Rigid Shape/Object Tracking

- Robust Non-rigid Motion Tracking and Surface Reconstruction Using L0 Regularization
  - Not the best search term mostly human tracking

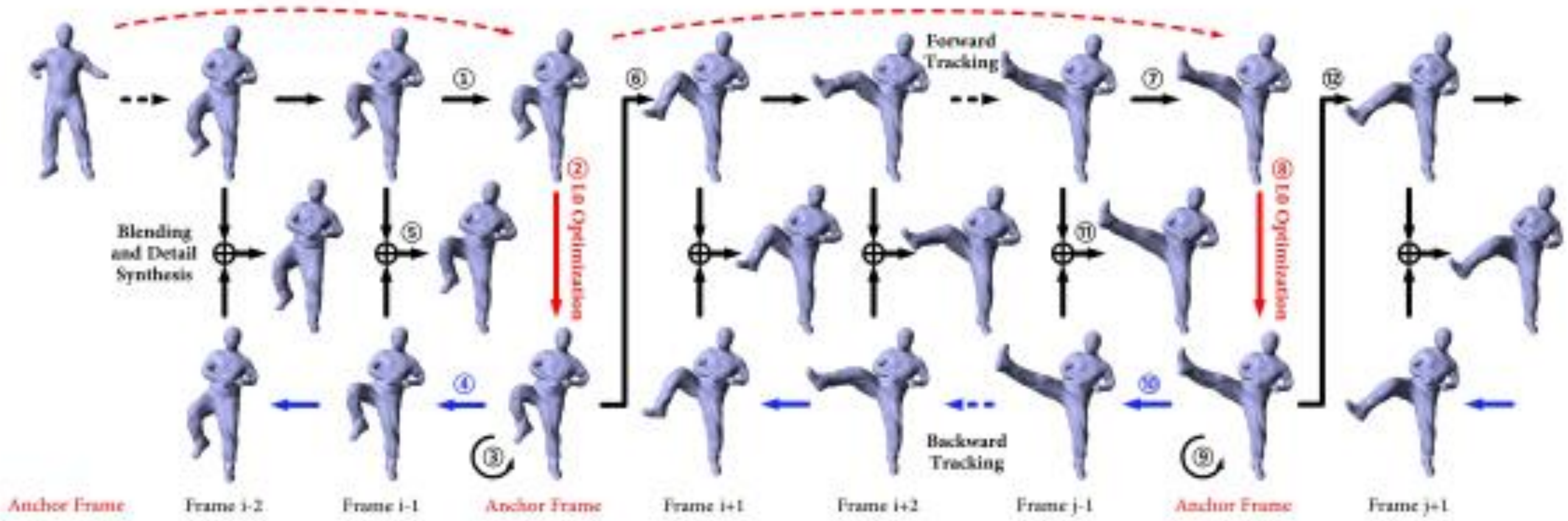


Figure 2. The pipeline of the proposed method. Basically, it is a forward-backward tracking scheme using combined L_2 and L_0 regularization for non-rigid deformation. The first row shows the forward tracking while the last row shows the backward tracking. The middle row is the blending between the two results for temporal smoothness. Please refer to Sect.3 for detailed description.

FEM and Video Tracking

- Poor Search Terms with respect to the problem

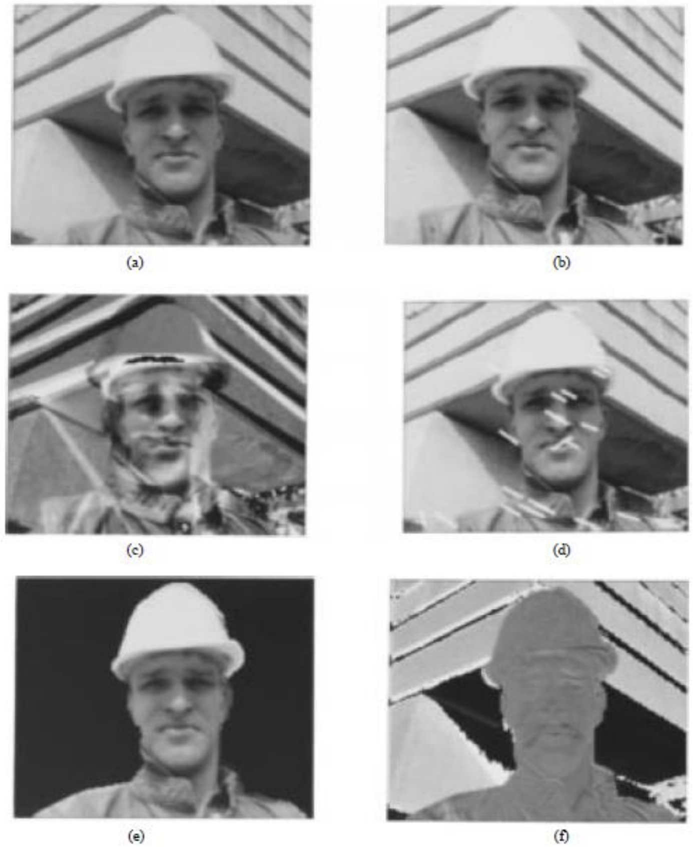


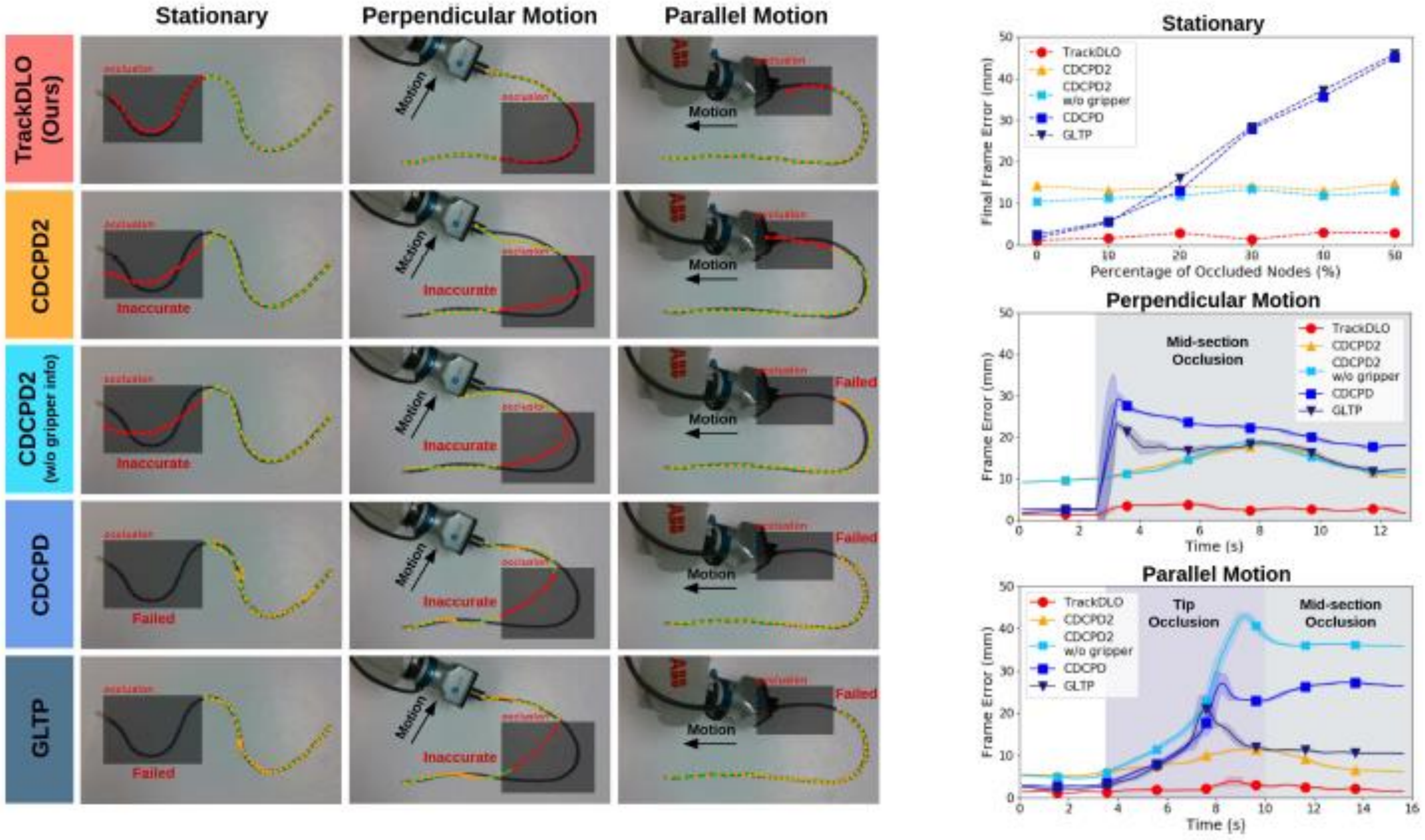
Fig. 7. (a) Frame 1 of "foreman." (b) Frame 10 of "foreman." (c) Frame difference. (d) Initial feature vectors. (e) Prediction of frame 1 inside mask. (f) Difference between original and predicted images.

Cable Specific

Realtime Robust Shape Estimation of Deformable Linear Object

- Good comparison of algorithms for tracking cable
- https://github.com/UM-ARM-Lab/cdcpd
- https://github.com/songdevelop/gltp
- https://github.com/RMDLO/trackdlo

- https://deformable-workshop.github.io/icra2024/
  - Workshop at ICRA on deformable object manipulation, next ICRA is Atlanta, 19th to 23rd July 2025



ig. 7. Compared to CDCPD2 with and without gripper information, CDCPD, and GLTP, TrackDLO accurately estimates the state of the DLO under scaled, tip, nd mid-section occlusion. Among these algorithms, TrackDLO had the lowest frame error.

Cable Specific

- Realtime Robust Shape Estimation of Deformable Linear Object
  - Key point labelling, interpolation and optimization.
  - More of a digital twin solution but could be a good place to start

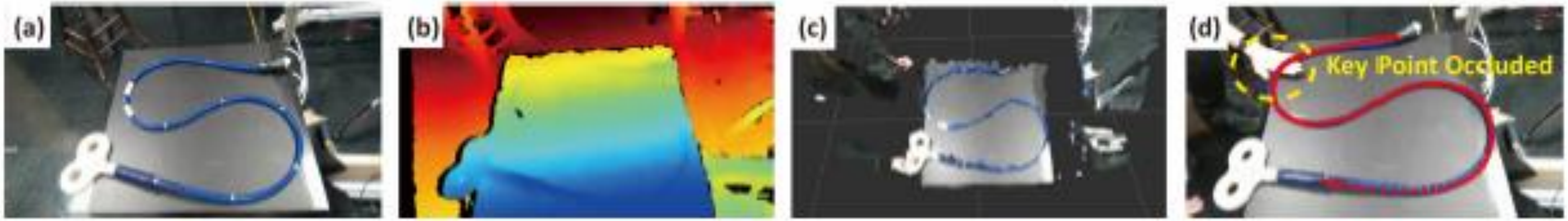


Fig. 7. (a)-(c) are the RGB image, depth image, and point cloud of the DLO, respectively. (d) is the estimation results overlaid on the RGB image. The red dots in (d) are the interpolation results. One key point is blocked but the estimation result still matches the actual shape.

Cable Specific

Robust Deformation Model Approximation for Robotic Cable Manipulation

- https://changhaowang.github.io/IROS2019/SPRRWLS.html
- Real time able tracking with gaussian mixture model based no rigid registration to track with noise, outliers and occlusions.
- Tracking, point cloud t , t+1 SPR registration used to registered to new point cloud with new estimation of tracking point positions
- https://github.com/thomastangucb/SPR
- https://thomastangucb.github.io/IJRR2018/

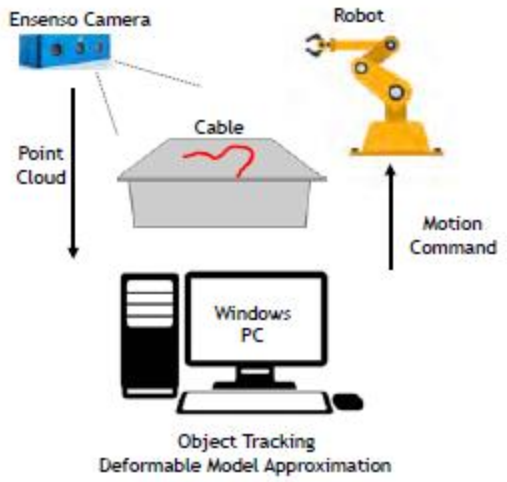


Fig. 5. The testbed setup

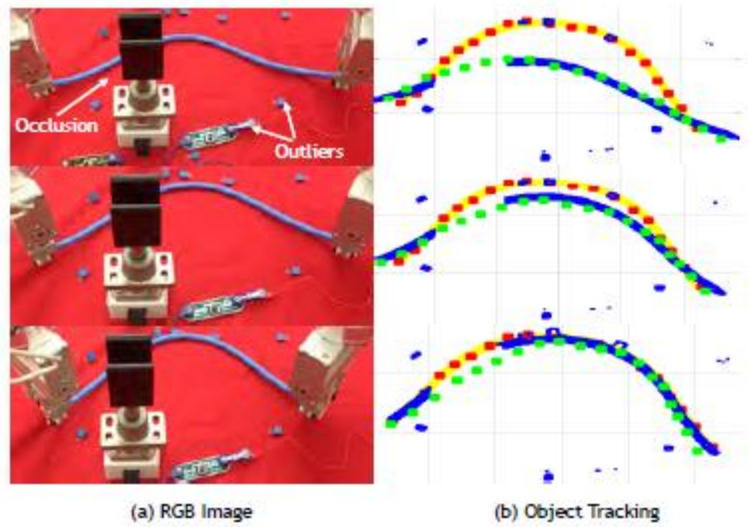


Fig. 6. SPR cable tracking in the presence outliers and occlusions. In (b), blue dots represent the perceived point cloud; yellow dots represent the target shape; red squares represent the desired feature points; and green squares represent current feature points.

Cable Specific

- Three-dimensional pose estimation of deformable linear object tips based on a low-cost, two-dimensional sensor setup and AI-based evaluation.
  - 2 Real sense cameras but don't use the depth in this paper…
  - Use of synthetic data with 3D model in Blender and ray tracing for cameras (could lead to the discrepancies)

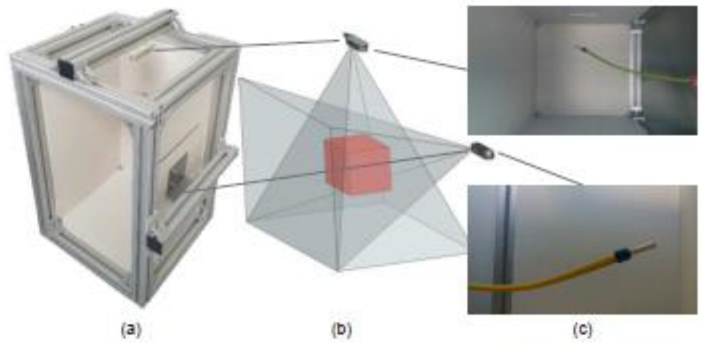


Fig. 2. (a) physical setup of the photobox; (b) working space of the setup; (c) resulting RGB-images from two views

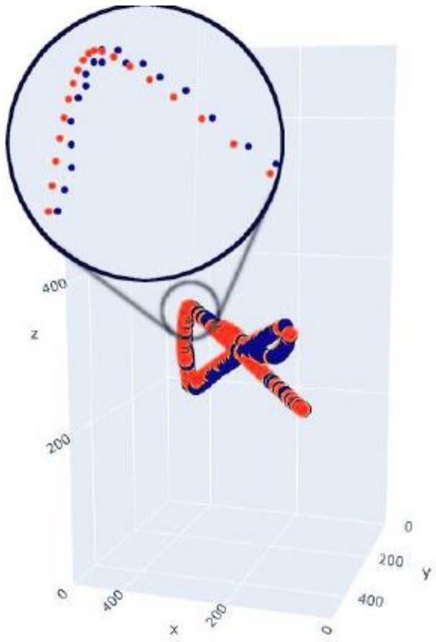
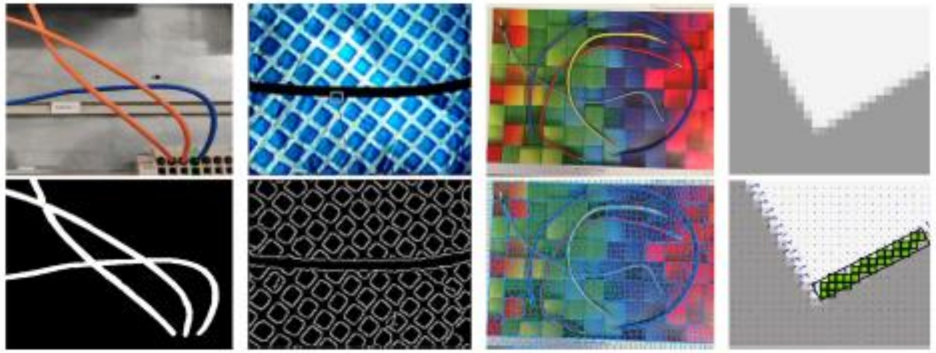


Fig. 6. Visualization of the position of a real (red) and predicted (blue) keypoint from a synthetic test dataset. The axes are given in millimeters.

Cable Specific

- An approach based on machine vision for the identification and shape estimation of deformable linear objects
  - Review on a lot of DLO detection and shape identification to look at in more detail
  - Focus on this paper, while not about manipulation or tracking explicilty etc, it has a strong element of detection cables which is required for tracking.
  - [An approach based on machine vision for the identification and shape estimation of deformable linear objects - ScienceDirect]()



(a) Background subtraction [13] (b) Edges detection [14] (c) Superpixel segmentation [9] (d) Other techniques [15]

Fig. 1. Machine vision techniques used as initial step for the shape estimation of DLOs.

Table 2
Characteristics and capabilities of published vision-based DLO detection approaches (2/2).

Ref	Approach	Prev info	Outcome	Camera	Speed
	Background subtraction				
[13]	Image differencing + contour following algorithm	None	DLO base points	2D grayscale	< 10 ms
[8]	Color filtering + pointcloud discretization + CPD	Prev state	Skeleton points	3D stereo	< 100 ms
[12]	Color filtering + Enhanced GMM-EM regularization + enhanced CDCPD	Prev state	Skeleton points	3D RGBD	< 100 ms
[14]	Color filtering + FCN + geometric finetuning	Training	Skeleton points	RGB + lidar	–
[15]	Background subtraction + geometric optimal control	Endpoints	3D skeleton	2D RGB	< 1 s
[17]	CNN objects detection and segmentation	Training	Points of interest	2D RGB	< 100 ms
[18]	DLO segmentation (CNN) + cross points determination (CNN)	None	B-Splines	2D RGB	–
[19]	CNN pixel classification + DBSCAN clustering	Training	Linear segments	2D RGB	< 100 ms
[20]	DLO segmentation (CNN) + vertices sampling + edges sampling + graph analysis	Training	DLO points + overlapping info	2D RGB	< 100 ms
[22]	Object classification and segmentation (SOM) + sequential thinning skeleton function [40]	Training	Skeleton points	2D RGB	–
[23]	Background subtraction + thinning skeletonization + graph structure analysis + intersections analysis	DLO diameter	Topological model	3D stereo	–
[24]	Background subtraction + skeletonization + contour extraction + DLO fitting and pruning + merging	None	DLO points	2D RGB	< 1 s
[25]	3D background subtraction (ROI intensity filter) + enhanced SPR method	WH info	Linear segments	3D stereo	< 10 s
[26]	3D background subtraction (plane fitting) + DLO discretization by region-growing + cost function minimization to find the correct segments order	Training cost function weights	Linear segments	3D RGBD	–
[27]	3D background subtraction + Pointcloud sorting + skeletonization + Joints estimation	None	Multibody model	3D stereo	–
	Edges detection				
[28]	Edge detection (Canny) + classification (MLP and SVM) + lines detection (Hough transform)	None	Border lines	2D RGB	< 100 ms
[29]	Edge detection (Canny) + lines detection (Hough transform)	None	Border lines	2D RGB	< 100 ms
[10]	Edge detection (Canny) + local minimization of an energy function of the DLO	Training + init config	Vertices + edges + material frames	3D stereo	–
[30]	Edges detection + skeleton line calculation + stereo merging	None	3D skeleton	3D stereo	–
[31]	Edge detection (Canny + NN) + Template-based tracking using the BFGS method to minimize the error	BEM model	Object template (BEM)	2D grayscale	< 1 s
[33]	Deformable template matching + transformations energy function minimization	Template	Triangulated polygons	2D grayscale	> 1 min
[34]	Edge detection (Sobel) + DLO borders intersection with adaptative tracking windows	Thickness + prev state	Borders + linear regression	2D grayscale	–
[35]	Edges detection (clustering) + parallel contours evaluation	None	Axis line	2D grayscale	< 100 ms
[36]	Multi-scale filtering + Hessian eigenvalues analysis	None	Filter (visual)	DSA, MSA[a]	–
[37]	Ridges detection and grouping by Hessian eigenvalues analysis	None	Segmentation	2D RGB	–
	Superpixel segmentation				
[9]	Superpixel segmentation + biased random walks	Endpoints	B-Splines	2D RGB	< 1 s
[11]	Multiscale superpixel segm + cost function to group them	None	Linear segments	2D RGB	–
	Other techniques				
[38]	Candidates selection (region growing), validation and modeling	None	Lines and arcs	2D grayscale	–
[39]	Cost function minimization by dynamic programming in successive evaluation windows to determine the contour	Reference points	Contour segmentation	DSA, MRI[a]	–

[a] *The images taken by these cameras have been analyzed as 2D grayscale images.*

Cable Specific

- Deformable Linear Objects 3D Shape Estimation and Tracking From Multiple 2D
  - Currently just deals with static scenes but a trend for 2 cameras, 2D only above and to the side of the scene is seen.

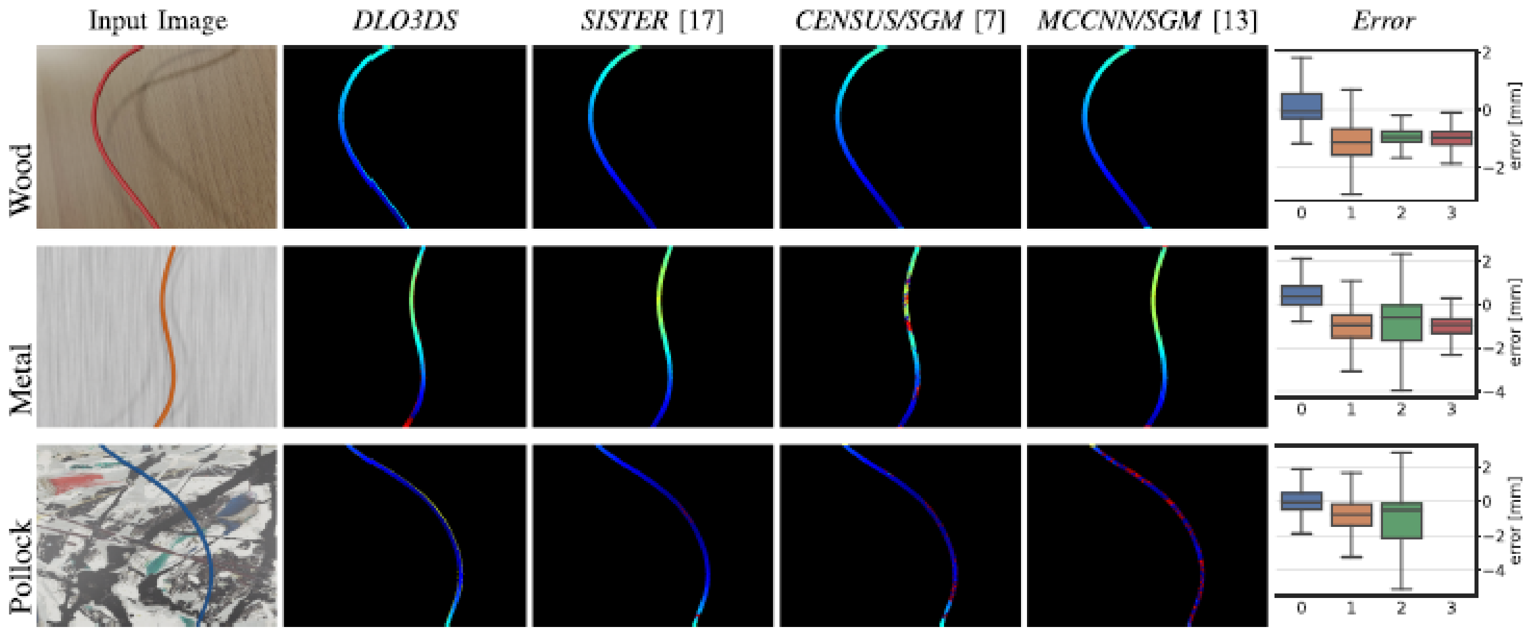


Fig. 8. Comparison with baseline methods in the form of depth images and boxplot. Plot legend: 0) *DLO3DS*, 1) *SISTER*, 2) *CENSUS/SGM* and 3) *MCCNN/SGM*. The display of the *MCCNN/SGM* boxplot in the third row is avoided due to large errors.