3D DLO Shape Detection and Grasp Planning from Multiple 2D Views

3D DLO Shape Detection and Grasp Planning from Multiple 2D Views

Alessio Caporali, Kevin Galassi, Gianluca Palli

***Abstract*—In this paper, the estimation of the 3D shape of a deformable linear object such as an electric cable for the purpose of planning suitable grasping poses is proposed. These poses can be then exploited by a robot manipulator to grasp and manipulate the cable for assembling and manufacturing purposes. The proposed method is based on a previously developed algorithm called ARIADNE providing the segmentation and spline modelling on a single image of electric cables even in cluttered scenarios. By exploiting this result and by collecting measurements from different points of view by means of a 2D camera mounted on the robot end effector, a method to estimate the 3D shape of the cable is here presented. This method is experimentally evaluated in different scenarios showing the capability of providing reliable estimation of the cable shape. The result is then used to drive to robot to grasp the cable in controlled positions.**

***Index Terms*—Deformable Object Manipulation, Shape Detection, Robotic Vision, Robotic Manipulation**

I. Introduction

In many real scenarios, manipulating successfully Deformable Linear Objects (DLOs) requires an adequate characterization of their shape. This knowledge is usually achieved via vision systems based on sensors having 2D or 3D capabilities. Concerning the 3D devices, in recent years there have been advancements in sensor technology for both high-end industrial depth sensors and consumer-grade ones. Selecting the correct 3D vision device for a specific application is commonly a complex job, since their performances heavily depend on many aspects as the target environment, lightning conditions and surface materials. Additionally, it is difficult to benchmark one camera with respect to the others, since 3D capabilities are achieved with different technologies and comparing the data-sheets does not provide a complete picture. When dealing with thin objects such as wires and cables, such cameras are inadequate in characterizing the shape of the DLOs. In such applications, high-end industrial depth sensors are used (e.g. Zivid One+ S and Photoneo MotionCam3D) since they can achieve sub-millimeter depth accuracy. The issue here is that these sensors are usually expensive and not really compact, moreover they have specific working distance to be ensured from the object that may not fit with other task constrains, in particular in case of flexible robotic cells and mobile manipulators. On the other hand, 3D cameras and sensors, such as RGB-D cameras, stereo camera, laser scanners are becoming more popular and affordable day by day and they are actually employed in several applications. Recently, cheap 3D cameras such as the Intel RealSense D435 appeared in the market enabling broad evaluation and implementation of tasks like collision avoidance, pick and place of normal size objects, and many more. However, these cameras typically have limited capabilities in terms of 3D estimation, and thin objects like cable are often pretty invisible to them. For the reasons mentioned above, 2D cameras could be more effective since they are really cheap, very small and reliable in providing useful data due to the potential very high resolution. Moreover, the field of view can be easily adapted and the working optimal distance to the object can be very small.

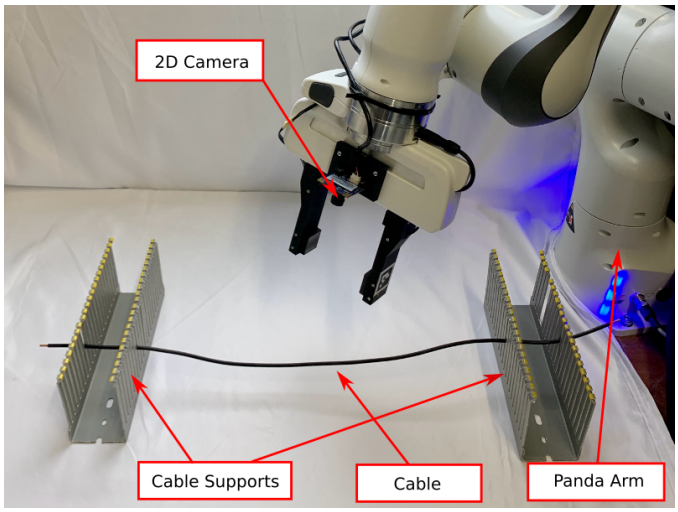


Fig. 1: Experimental Setup. A 2D camera is mounted on the robot end effector, the camera is positioned to capture objects between the fingers of the robot. The robot used in the experiments is a 7DOF Panda from Franka Emika.

In recent works, several approaches have been introduced related to the detection of DLOs both in 2D and 3D scenarios. Looking at the 2D ones, in [1] the authors introduced an approach for the segmentation of DLOs in complex cluttered backgrounds. Simpler approaches as [2], [3] and [4] can also be applied since DLOs appear in images as tubular structures. The 3D detection is instead mostly documented for ropes manipulation when dealing with problems as knots untangling [5]. Here the authors segment the background via color and plane fitting techniques and then built a graph representing

Alessio Caporali, Kevin Galassi and Gianluca Palli are with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI) of the University of Bologna, Italy.

This research received funding from the EC H2020 research and innovation program under grant agreement n. 870133 - REMODEL (Robotic Technologies for the Manipulation of Complex Deformable Linear objects).

Corresponding author: *alessio.caporali2@unibo.it*

(a) (b) (c)

Fig. 2: The input image (2a) is processed obtaining a binary mask (2b) and a 2D B-Spline model of the DLO (2c). In the remainder of the paper a white background is used for clarity in the figures.

the rope configuration. Concerning the manipulation of cables for industrial application, several approaches are discussed in literature as routing [6], planning and assembly in [7], [8] and [9].

In this paper, some preliminary results concerning the estimation of the 3D shape of DLOs via a common 2D eye-in-hand calibrated camera are presented. The main idea consists in exploiting multiple views of the same static scene. An algorithm called ARIADNE [1] is used to detect the DLOs from each input camera image through deep-learning based segmentation and providing a B-Spline representation of the DLO in pixel coordinates. Then, by collecting the estimated 2D splines provided by different points of views, the method proposed in this paper estimates the 3D shape of the DLO described by means of a spline in the 3D Cartesian space. The deformability of the wire has not been considered in the process of shape reconstruction, since the method proposed is vision-based only. However, in future works, a model of the wire deformation can be adopted to enhance the reconstruction capabilities in case of occluded portions of the cable. The error between the computed 3D model of the cable projected in each input image and the 2D input points in the original image is computed and utilized for the evaluation of the approach.

In Fig. 1 the setup used for the experimental evaluation of the approach presented in this paper is shown. The idea is to simulate the problem of cable grasping from racks. On the robot end-effector, is possible to see the used 2D camera.

The remainder of the paper is organized as follow: the cable segmentation and modelling, is described in Sec. II, while in Sec. III is presented the 3D DLO shape estimation. Finally, in Sec. IV, the method is experimentally evaluated.

II. Cable Segmentation and Modelling

For the detection and modelling of cables in images, we built on top of the approach first presented in [1]. There, the authors describe a framework for the segmentation and modelling of cables featuring complex backgrounds. The main drawbacks of the approach are three: 1) the need of the bounding boxes of the cables' terminals at the algorithm initialization, 2) the computational efficiency quite low, 3) the small robustness against difficult combinations of cables and backgrounds. The need of a separate tool for the detection of the cable terminals is a strong limiting factor. The computational efficiency of the approach is another relevant issue since a very long processing time, although not critical in our scenario, makes the approach less valuable. Lastly, we are planning to deploy a system in an industrial scenario where the backgrounds, lighting conditions and cables textures can not be controlled, thus a strong robustness against all these factors is mandatory.

Considering the issues just described, we decide to modify slightly the approach aiming at improving in all the aspects. We decide to adopt a deep convolution neural network for the background segmentation. The application of such approach allow us to solve, at the same time, both the terminal detection and robustness drawbacks. In fact, once the background has been removed, the terminal can be easily characterized (additional details at the end of this section). As network architecture, DeepLabV3+ [10] is chosen since it represents the state-of-the-art in semantic segmentation tasks. The network is trained on a synthetic dataset [11] built utilizing a chroma-key based approach which allows a generation of a huge dataset automatically. The dataset features both complex cluttered backgrounds and more simple ones. The lack of features in cables and DLOs in general, makes the creation of a good-quality dataset crucial. In Fig. 2b is shown an example of generated mask for the input image of Fig. 2a.

Given the binary mask of the image, in which background pixels (black) are separated from the foreground ones (white), a superpixel segmentation is applied on the latter. The idea of superpixels is to partition the image into local meaningful areas making the further processing easier and faster. We deploy the Simple Linear Iterative Clustering (SLIC) [12] algorithm in this regard which uses color and proximity pixel information in a 5D space for the segmentation. In particular, a modified version, called MaskSlic [13], that is capable of applying the superpixel segmentation only on a region of interest (i.e. the foreground pixels) is used, due to the availability of the binary mask previously computed. The

application of MaskSlic allows us to exploit completely the result of the initial semantic segmentation.

Based on the superpixel label map, a region adjacency graph (RAG) is built. It is an undirected and not weighted graph where each superpixel is represented as a graph node. Instead, the edges of the graph are used to model the neighbouring relationships between superpixels. Each graph node is augmented with the centroid information of the associated superpixel. Concerning the cables' terminals detection, they can be extracted simply from the graph as the nodes characterized by having only one neighbour. The make the entire 2D detection efficient at run-time, fast *walks* are performed on the graph aiming at organizing the set of graph nodes as an ordered sequence from one terminal to another of the DLO considered. Finally, the obtained *walk* is used to estimate a 2D spline representing the DLO in the current image plane.

A generic DLO shape can be represented in the Cartesian space by a 3-rd order spline basis as a function of a free coordinate u representing the position along the cable starting from an end point, where $u = 0$, to the opposite end where $u = L$, being L the length of the DLO

$$q(u) = \sum_{i=1}^{n_u} b_i(u) q_i \tag{1}$$

where $q(u) = [x(u)\ y(u)\ z(u)]^T$ is the vector of Cartesian coordinates of each point along the DLO, $b_i(u)$ is the i-th elements of the spline polynomial basis used to represent the DLO shape and q_i are n_u properly selected coefficients, usually called *control points*, used to interpolate the DLO shape through the $b_i(u)$ function basis. It is worth noticing that the same spline model can be used to represent the DLO in the image plane, i.e. the output of the ARIADNE algorithm, we will refer as $p(u) = [p_x(u)\ p_y(u)]^T$ as the 2D vector of pixel coordinates representing the estimated spline in the input image.

III. 3D DLO Shape Estimation

A robot equipped with an eye-in-hand 2D camera is used to perform pre-determined trajectories in a circular fashion around the object of interest. The trajectory is parametrized with a user-defined number of poses for the acquisition of sample images between two starting and ending reference poses. In details, the two reference poses are interpolated in order to generate a number n of final poses. The orientation between the two reference poses is interpolated utilizing the spherical linear interpolation method. Instead, the position is interpolated linearly.

For each acquired image, the detection pipeline described in Sec. II is applied and the computed spline of the frame under exam is stored together with the relative pose of the camera. The method to estimate the 3D coordinates of a single point observed from multiple points of view is illustrated first, then the method is extended to the estimation on the whole DLO shape.

A. Ray Tracing and Nearest Point Search

Let us consider the case in which a single unknown point x in the Cartesian space and expressed with respect to the world reference frame is observed by the camera mounted on the robot from multiple points of view. Provided that the camera frame with respect to world frame at the i-th points of view is

$${}^{w}T_{c_i} = \begin{bmatrix} {}^{w}R_{c_i} & {}^{w}t_{c_i} \\ 0\ 0\ 0 & 1 \end{bmatrix}$$

where ${}^{w}R_{c_i}$ is the rotation matrix and ${}^{w}t_{c_i}$ is the position of the camera frame origin in world coordinates, we assume the point x is seen in the image related to the i-th points of view at $p_i = [p_{x_i}\ p_{y_i}]^T$, being p_{x_i} and p_{y_i} the point pixel coordinates in the image.

A so-called unit ray v_i passing through the image reference frame origin and x can be expressed in the image frame considering the pixel coordinates p_i and the camera focal distance f

$$v_i' = \begin{bmatrix} p_{x_i} - c_x \\ p_{y_i} - c_y \\ f \end{bmatrix}, \quad v_i = \frac{v_i'}{\|v_i'\|} \tag{2}$$

where c_x and c_y are the pixel coordinates of the image center (assuming the camera frame is centered with respect to the image). Then, v_i can be expressed in the world frame by

$${}^{w}v_i = {}^{w}R_{c_i} v_i \tag{3}$$

Provided that n_p distinguished points of view are available, the estimation $\tilde{x}$ of the unknown point x can be obtained by looking for the point having the minimum distance from all the rays. By defining the symmetric V_i matrix

$$V_i = I - {}^{w}v_i {}^{w}v_i^T \tag{4}$$

providing the seminorm on the ray distance, the point location estimate $\tilde{x}$ is provided by nearest point search algorithm, i.e.

$$\tilde{x} = \left(\sum_{i=1}^{n_p} V_i \right)^{-1} \left(\sum_{i=1}^{n_p} V_i\, {}^{w}t_{c_i} \right) \tag{5}$$

B. Estimation of the 3D DLO Spline

Here the objective is to estimate the 3D spline providing the best fitting of the cable shape observed from multiple points of view by the 2D camera. We assume here that the ARIADNE algorithm provide us as input to this procedure a 2D spline on the image plane representing the observed DLO.

By assuming that the same portion of the DLO is seen in all the images (this can be achieved by moving the camera along the normal direction with respect to the main spline axis detected by each image and adjusting the mean distance from the DLO to a constant value as estimated by the progressive inclusion of successive images, this will be subject of future research), it is possible to apply the aforementioned method for the estimation of single point via ray tracing and nearest point search to a suitable set of sample point along the DLO spline. Let us call the set of spline samples $p_{ij} = p_i(u_j)$,

$j = 1, \ldots, n_s$, $i = 1, \ldots, n_p$, where n_s is the number of spline samples, n_p is the number of points of view, $p_i(\cdot)$ is the spline provided for by the i-th image and u_j are the spline sample points.

The vector of control points $q_v = [q_1 \ \cdots \ q_{n_u}]^T$ of the 3D spline $q(u)$ that optimally approximated the set of point estimates p_{ij} can be defined as

$$q_v = B^{\#} \tilde{x}_v \tag{6}$$

where $^{\#}$ represents the matrix pseudoinverse and

$$B = \begin{bmatrix} b_1(u_1) & \cdots & b_{n_u}(u_1) \\ b_1(u_2) & \cdots & b_{n_u}(u_2) \\ \vdots & \vdots & \vdots \\ b_1(u_{n_s}) & \cdots & b_{n_u}(u_{n_s}) \end{bmatrix}$$

$$\tilde{x}_v = \begin{bmatrix} \left(\sum_{i=1}^{n_p} V_{i1}\right)^{-1} \left(\sum_{i=1}^{n_p} V_{i1} \, {}^w t_{c_i}\right) \\ \left(\sum_{i=1}^{n_p} V_{i2}\right)^{-1} \left(\sum_{i=1}^{n_p} V_{i2} \, {}^w t_{c_i}\right) \\ \vdots \\ \left(\sum_{i=1}^{n_p} V_{in_s}\right)^{-1} \left(\sum_{i=1}^{n_p} V_{in_s} \, {}^w t_{c_i}\right) \end{bmatrix}$$

being V_{ij} the matrix computed according to eq. (4) for the j-th sample provided by the i-th image.

C. Evaluation of Estimation Error by Reprojection

To evaluate the estimation error, the estimated 3D DLO spline is reprojected on each image and the difference with respect to the initial estimation provided on the image plane by the ARIADNE algorithm is computed. Considering a generic 3D spline sample $q(u_j) = B\, q_v$, its homogeneous representation is represented by $\bar{q}(u_j) = [q(u_j)^T \ 1]^T$. The projected coordinates $\tilde{p}_{ij} = [\tilde{p}_{x_{ij}} \ \tilde{p}_{y_{ij}}]^T$ of the j-th spline sample on the i-th image plane can be written as

$$\tilde{p}'_{ij} = \begin{bmatrix} \tilde{p}'_{x_{ij}} \\ \tilde{p}'_{y_{ij}} \\ \tilde{p}'_{z_{ij}} \end{bmatrix} = A \left[{}^w R_{c_i}^T \mid -{}^w R_{c_i}^T \, {}^w t_{c_i} \right] \bar{q}(u_j) \tag{7}$$

$$\tilde{p}_{ij} = \begin{bmatrix} \tilde{p}_{x_{ij}} \\ \tilde{p}_{y_{ij}} \end{bmatrix} = \begin{bmatrix} \tilde{p}'_{x_{ij}} / \tilde{p}'_{z_{ij}} \\ \tilde{p}'_{y_{ij}} / \tilde{p}'_{z_{ij}} \end{bmatrix} \tag{8}$$

where

$$A = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

is the camera matrix containing the camera intrinsic parameters, such as the focal length and center point coordinates. Then, the overall error is provided by collecting all together in a single vector the error related to each single image, i.e. $e = [\cdots \ e_{ij} \ \cdots]^T$, $j = 1, \ldots, n_s$, $i = 1, \ldots, n_p$, where $e_{ij} = \|p_{ij} - \tilde{p}_{ij}\|$ is the distance between the corresponding initial spline sample provided by ARIANDE and the projection on the image plane of the estimated 3D spline sample. Finally, the mean error norm $\|e\|_{n_s n_p} = \sqrt{e^T e}/(n_s \, n_p)$ can be used to evaluate the quality of the estimation result.

IV. Results

To validate the proposed approach, several experiments are performed aiming at evaluating the performances and the capabilities of the system in correctly capturing the shape of the cable in 3D (Sec. IV-A) and hence in providing a reliable 3D model of the DLO for grasp tasks (Sec. IV-B). In addition, the quality of the estimation result is analyzed from two different point of views, the type of trajectory (Sec. IV-C) and the number of samples n_s employed (Sec. IV-D).

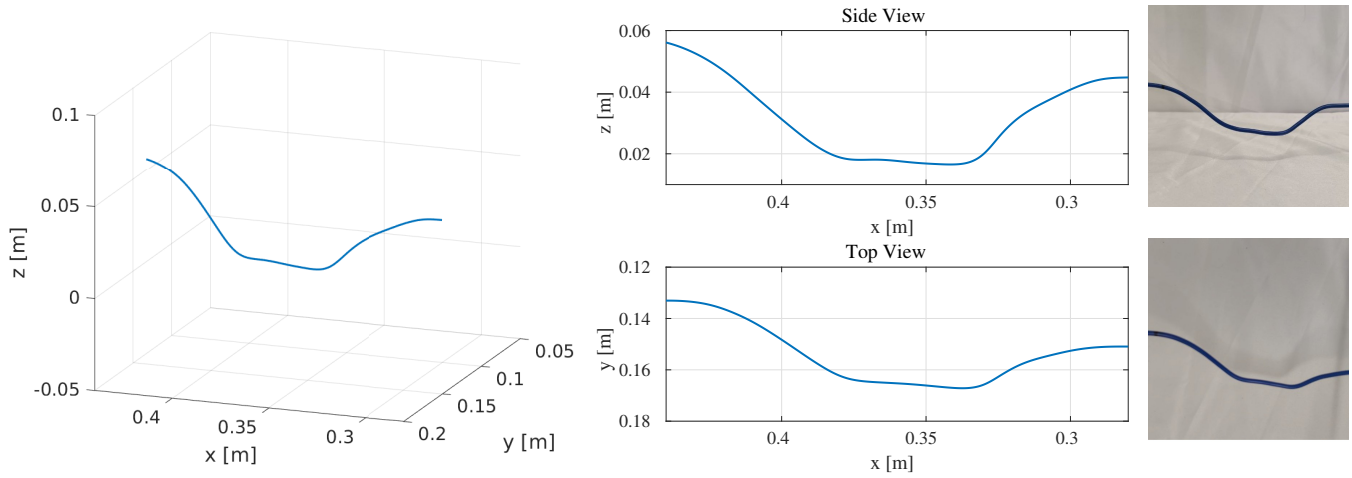
The Panda 7-DoFs robotic arm from Franka Emika equipped with his panda-hand parallel fingers gripper is adopted in the experiments reported in this paper, and the robot is controlled via ROS interface. A low-cost 2D camera with USB interface is mounted on the robot end-effector. In the experiments, the camera resolution is constrained to 640x480 pixels. The experimental scenario is composed by the robot placed near two racks, commonly used in electrical cabinets, holding the cable at a suitable distance from the underneath table. The cable under test is roughly aligned with the robot x axis. This assumption is required following the discussion made in Sec. III in order to see almost the same cable section in the camera images. Anyway, it is fairly simple to develop a correcting procedure aiming at bringing the cable aligned using the first sample in this regard. The 2D vision algorithm for the cable detection is based on the ARIADNE algorithm, thus a machine equipped with a Nvidia GPU is required. In our case the vision node is deployed on a separate computer having an Nvidia GTX-2080Ti. The communication and deployment of the robotic system is achieved thanks to the ROS middle-ware framework. The robot trajectory is planned thought the MoveIt! package. A ROS service is used for the communication between the planning node and the vision one. The vision node takes approximately 0.2 seconds for the detection and spline modelling of the cable in each image.

A. Cable Reconstruction

First, the reconstruction capabilities of the approach presented are tested from the point of view of the fidelity with respect the real shape of the cable. In Fig. 3a the 3D spline obtained by detecting the shape of a test DLO is shown. Then, in Figs. 3b and 3c the projections of the plot in the XZ and XY planes are shown together with two associated images displaying the real cable. From these pictures it is possible to appreciate the accuracy in the shape behavior for both views, meaning that the approach was able to capture well the cable characteristics. The figure shown was obtained by utilizing a linear trajectory of 10 centimeters along the y axis, with n_s equal to 15 and n_p equal to 20.

B. Grasp Evaluation

The second part of the evaluation activities consists in the estimation of the cable position using the method discussed and by concluding the experiment by attempting a grasp of the cable. The point of grasp is computed using the obtained 3D shape by taking the middle point of the 3D spline. Thus, given the parameter $u \in [0, 1]$ of the spline the point is taken at

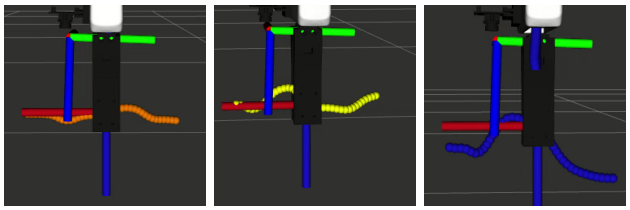


(a) 3D Cable Plot.

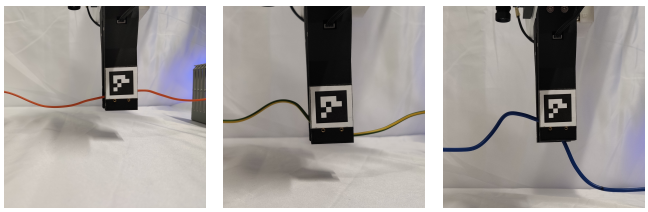
(b) 2D Cable Plots.

(c) Real Cable Shape.

Fig. 3: 3D shape reconstruction based on the 2D camera acquisition.



(a) Representation of the 3D cable shape estimation in RViz.



(b) $\phi = 2mm$ (c) $\phi = 3mm$ (d) $\phi = 4mm$

Fig. 4: The resulting method can be applied to obtain the grasping point in different condition in term of shape, dimension and colour of the desired cable.

$u = 0.5$. From the conducted test it results that the grasp can be achieved with cable of different plain color Fig. 4b, 4d or mixed colour such as a ground cable Fig. 4c. Another variation of the characteristic studied is differences in terms of cable dimension. Even in this case no issues have been observed and it is possible to conclude that the system is capable in handling cables of different sizes and colors.

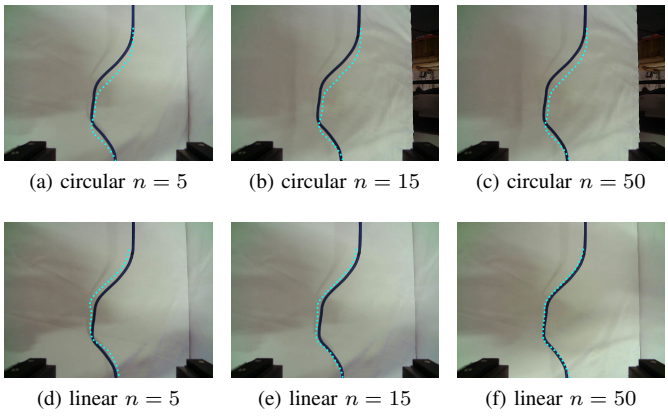
C. Comparison Between Linear and Circular Trajectories

How the robot trajectory during the acquisition of camera views affects the DLO shape estimation result is now analyzed. In this regard, a linear trajectory of 10 centimeters along the robot y axis in which the camera maintains a fixed orientation is compared to a circular one in which the camera rotates around an axis approximating the cable oriented as the robot x axis and characterized by a rotation of about 90 degrees. Fig. 5 displays the 3D points projected back on the plane of a test image for the sake of a visual comparison between real and estimated data. It is worth noticing that the pose corresponding to the test image utilized is not used during the 3D shape estimation procedure. Instead, it corresponds to a view from the left side of the cable and it is used to gain a better understanding of the cable's real shape. This figure shows how the linear approach seems to be able to produce a better reconstruction of the real cable. This phenomenon can be justified by considering that, in the linear case: 1) The camera orientation is kept fixed, hence the propagation of errors due to the extrinsic calibration in the algorithm is less significant; 2) It is easier to guarantee the correspondence of the points between the different samples, since in the circular trajectory the camera can only rotate around an estimation of the cable axis.

D. Effect of Samples Number

Another aspect investigated in this work is how the number of samples n_s can affect the estimation quality. The experiment is started by evaluating a number of samples n_s equal to 5, then progressively augmented to 15, 50 and 100 respectively. While the number of samples is varied, the same object to be reconstructed and the same robot trajectory defined by the same starting and ending pose, interpolated in a number of point n_p equal to the desired number of samples, are used. The test is performed for both the linear and circular trajectories.

As it can be seen in Fig. 6, it comes out that the number of samples affects the estimation quality, reducing the norm of the error from his highest value (in case of two samples) 4.8 pixel to 2.8 (with 100 samples) in the case of the linear trajectory. However, the larger error decrease is observed passing from 2 to 5 observation poses, and an acceptable error in the x axis can be obtained even with a number of samples equal to 5



(a) circular $n = 5$ (b) circular $n = 15$ (c) circular $n = 50$

(d) linear $n = 5$ (e) linear $n = 15$ (f) linear $n = 50$

Fig. 5: Comparison between linear and circular trajectories and relation with the number of poses used for the estimation.

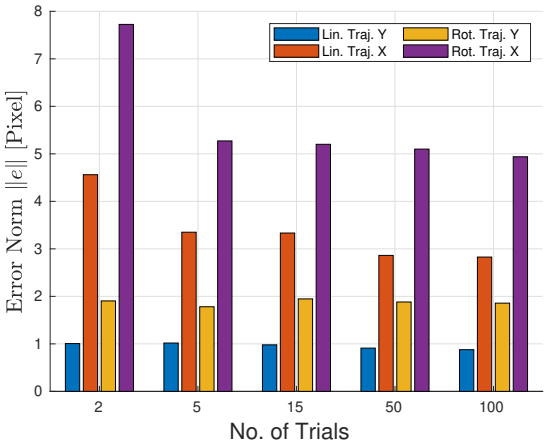


Fig. 6: The graph shows how the norm of the error over the x axis of the image frame decreases asymptotically with the increase of the number of pictures taken. The error in the y axis remains bounded.

or 15. This result is confirmed by the grasp test of Sec. IV-A where the grasp is successfully achieved with $n_s = 15$. For the circular trajectory, the error values are in general higher than the ones provided by the linear trajectory, in accordance with the discussion of Sec. IV-C. For this type of trajectory, the improvement with a bigger number of samples is not substantial.

V. Conclusion

In this paper, we evaluated the use of a 2D camera to reconstruct the position in the Cartesian space of a deformable linear object with a precision that allows the perform grasps with a robotic manipulator. An adequate knowledge of the extrinsic and intrinsic parameters of the camera is a requirement, however this parameters are also a prerequisite for other methods involving 3D camera as well. At contrary, the cable parameters are not a requisite for the 3D estimation. The results can be accomplished with a suitable estimation in the range of few image pixels even with a limited number of images samples, in the order of 5. This result makes the proposed method suitable for the implementation of a more realistic task, that will be evaluated in future research. The precision of the algorithm is limited by how much the camera is capable to takes the same portion of the cable in each sample image, in facts if different portions of the cable are observed during the acquisition, the shape estimation will be very likely diverted. To solve this issue, a new recursive version of the algorithm will be implemented in future research to obtain a more robust measure and to take into account the presence of outliers in the data acquisition process. Another step forward in future research will be to consider scenarios in which multiple cables are present, considering also the possibility of intersection between cables. In this case, the different cables must be identified separately and the corresponding spline estimations organized accordingly.

References

[1] D. De Gregorio, G. Palli, and L. Di Stefano, "Let's take a walk on superpixels graphs: Deformable linear objects segmentation and model estimation," in *Lecture Notes in Computer Science - Asian Conference on Computer Vision*. Springer, 2018, pp. 662–677.

[2] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI 1998*.

[3] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Tran. on Medical Imaging*, pp. 501–509, 2004.

[4] V. Pătrăucean, P. Gurdjos, and R. G. Von Gioi, "A parameterless line segment and elliptical arc detector with enhanced ellipse fitting," in *ECCV 2012*, pp. 572–585.

[5] W. H. Lui and A. Saxena, "Tangled: Learning to untangle ropes with rgb-d perception," in *Proc. of the IROS*, 2013, pp. 837–844.

[6] K. Galassi and G. Palli, "Robotic wires manipulation for switchgear cabling and wiring harness manufacturing," in *Proc. IEEE Int. Conf. on Industrial Cyber-Physical Systems*, 2021.

[7] D. De Gregorio, R. Zanella, G. Palli, S. Pirozzi, and C. Melchiorri, "Integration of robotic vision and tactile sensing for wire-terminal insertion tasks," *IEEE Tran. on Automation Science and Engineering*, vol. 16, no. 2, pp. 585–598, 2018.

[8] G. Palli, S. Pirozzi, M. Indovini, D. De Gregorio, R. Zanella, and C. Melchiorri, "Automatized switchgear wiring: An outline of the WIRES experiment results," in *Advances in Robotics Research: From Lab to Market*. Springer, 2020, pp. 107–123.

[9] A. Shah, L. Blumberg, and J. Shah, "Planning for manipulation of interlinked deformable linear objects with applications to aircraft assembly," *IEEE Tran. on Automation Science and Engineering*, vol. 15, no. 4, pp. 1823–1838, 2018.

[10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[11] R. Zanella, A. Caporali, K. Tadaka, D. De Gregorio, and G. Palli, "Auto-generated wires dataset for semantic segmentation with domain-independence," in *Proc. IEEE Int. Conf. on Computer, Control and Robotics*, 2021, pp. 292–298.

[12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[13] B. Irving, "maskslic: regional superpixel generation with application to local pathology characterisation in medical images," *arXiv preprint arXiv:1606.09518*, 2016.