Harnessing Vision and Touch for Compliant Robotic Interaction with Soft or Rigid Objects

Ana-Maria Cretu and Pierre Payeur

Abstract The chapter discusses recent research achievements related to sensing issues and interfacing techniques to enable safe interaction of commercial-grade robot manipulators with objects exhibiting rigid or soft surfaces. The main challenges are described, including the identification of proper combinations of vision and touch sensor technologies, and their placement and trajectory with respect to the objects of interest to enable safe navigation and close interaction. Various selective data acquisition procedures are also examined to ensure fast and sufficient monitoring of the interaction behaviour of the object under forces imposed by a robotic manipulator or a multi-finger gripper. Issues related to sensor calibration and data fusion are detailed. Potential solutions are presented in the context of various interaction tasks, including adaptive surface and contour following, object characteristics identification, and dexterous robot hand manipulation of soft objects using the Barrett hand. Numerous experiments demonstrate the validity of the proposed solutions.

1 Introduction

Modern robotic systems to be employed in industrial, security and space applications require the development of a new generation of autonomous robot manipulators able to intelligently perform sophisticated manipulation tasks [1] in environments that are often unknown, variable or unstructured. Over the past

A.-M. Cretu (✉)
Computer Science and Engineering Department,
Université du Québec en Outaouais, Gatineau, Canada
e-mail: ana-maria.cretu@uqo.ca

P. Payeur
School of Electrical Engineering and Computer Science,
University of Ottawa, Ottawa, Canada

B. George et al. (eds.), *Advanced Interfacing Techniques for Sensors*,
Smart Sensors, Measurement and Instrumentation 25,
DOI 10.1007/978-3-319-55369-6_9

decades, a huge research effort was invested in the design and development of robot systems able to sense and react intelligently to their environment and safely handle rigid and deformable objects. Building such systems that can interact autonomously with unknown objects is a complex task, requiring a combination of sensing technologies, control systems, knowledge of computer and mechanical engineering, as well as an understanding of human abilities that could be mimicked in order to produce more flexible, general and intelligent solutions.

Human vision-touch experience shows the ability of vision in assisting grasping, handling and manipulation tasks. In a similar manner, autonomous robot manipulators can count on the coordination of these two sensory capabilities to adapt to unpredictable situations and work efficiently in unknown environments. Vision sensing, as provided by stereo cameras, RGB-D sensors or range scanners provides rich information on the geometry and topology of the objects to be manipulated. Along with advanced image processing techniques, it can also enable the monitoring or tracking of soft object deformations under forces exerted by the manipulator. Visual feedback can improve the grasping and manipulation process by guiding the robot manipulator and assist in the estimation of the relationship between the object and the end-effector. Integrating visual feedback with touch (contact, force) sensing also compensates for the inaccuracy of vision systems alone due to occlusions and the inability of vision sensors to provide force measurements. Moreover, the use of vision sensing in the system can guide the touch probing towards areas of relevant features in order to shorten the exploration time which can be long, as the manipulator must execute multiple complex motions to collect tactile data.

Most of current research effort in the robotics literature is focused on manipulation and grasping of rigid objects. Relatively few researchers yet dedicated their interest to the interaction with deformable objects, while in fact numerous real-world objects are mostly unsymmetrical, compliant, and exhibit alterable shapes. The robotic manipulation of deformable objects still offers an important challenge to the robotics community and makes it a subject of significance for the development of future generations of autonomous robots. This chapter discusses challenges in rigid and deformable object grasping and manipulation based on a combination of vision and touch sensor technologies. The authors' research groups investigated their placement and trajectory with respect to the objects of interest to enable safe navigation and close interaction, the various selective data acquisition procedures to ensure fast and relatively complete monitoring of the interaction behaviour of the object under forces imposed by a robotic manipulator, as well as object modelling techniques. These issues and some potential solutions are exemplified in the context of various interaction and manipulation tasks, including adaptive surface and contour following, surface characteristics identification, and dexterous robotic hand manipulation of soft objects using the Barrett hand.

2 Challenges of Robotic Interaction with Soft or Rigid Objects

While the interaction with objects exhibiting soft or rigid surfaces is one of the fundamental capabilities of autonomous robot systems, the design and development of comprehensive autonomous robotic systems able to interact with surfaces and manipulate objects, in particular soft deformable objects, without human intervention remains a challenging task. As briefly highlighted in the introduction, such complex interaction can only take place with the assistance of multisensory data acquisition systems that combine vision and touch (tactile, force-torque) measurements. Such sensors allow for the recuperation of crucial information on the interaction, including the location (pose) of the object in the environment, the occurrence of a contact between the manipulator and the object, the size and shape of the object, its material properties, the magnitude and position of forces exerted by the manipulator or the detection of slippage of the object from the manipulator. A coordinated fusion of this information opens door to dexterous manipulation. However, there are several issues that complicate the automation of this information acquisition and fusion.

In the case of large rigid objects, if multiple vision sensors are involved, a calibration process is required prior to their use. An object model is indispensable to represent the geometry of the object and to enable its close inspection or the interaction with it. Ideally, this model needs to be compact to support the robot's operation in real time. Moreover, because local tactile probing is time consuming, intelligent selective algorithms should be employed to only select areas of interest, such as areas where the local geometry changes, for enhancing the sensing procedure. Path planning algorithms have to be employed accordingly to guide the interaction with the object.

In contrast to the manipulation of rigid objects which has been extensively studied in the literature and for which well-established procedures exist, the investigation of the manipulation of soft deformable objects represents a more recent undertaking. While several 1D and 2D solutions tackle the issue of grasping and manipulation of soft objects [2, 3], few researchers have addressed the manipulation and grasping of 3D objects [4–7]. This is due to its complexity and to the fact that a majority of researchers hope to tackle simpler 1D and 2D modelling problems before generalizing to a 3D solution. One of the most critical issues is the difficulty to estimate or predict in real-time the deformation properties of the object [8]. These properties tend to vary greatly among various objects. Their understanding and their prediction, ideally without making assumption on the material (such as linearity, homogeneity and isotropy), is necessary to coordinate the motion of the manipulator and its interaction with the object. If touch sensors are involved as well (i.e. force-torque sensors), a synchronization of visual and force (or tactile) data is required as different sensing technologies work at different sampling rates. As well, in this case, as for rigid objects, the probing should be restricted only over areas of interest. Furthermore, monitoring the coupling between the contact forces

and the object deformation is also necessary to study the impact of the position, the magnitude and the angle of forces applied to the object over the various stages of the shape deformation. In order to implement and evaluate the interaction tasks of a soft object with a robot manipulator, a soft object model is required to represent the deformation characteristics during the physical interaction. In classical models, the deformation characterization generally implies the approximate identification of elastic parameters of the model, generally a mass-spring model [5, 9] or a finite element representation [10–12], by comparing the real and simulated object subject to interaction and aiming to minimize the differences. However, these approaches work only by making assumptions on the object material, such as linearity, homogeneity or isotropy, which do not transpose well to multiple materials such as foam or rubber. These justify our interest in the development of methods that do not make assumptions on the material of the object, but rather directly employ experimental data to make decisions on the properties of the object and capture implicitly the deformation behaviour.

Once the model is developed, a control scheme has to be proposed to ensure smooth interaction with the object. For rigid object exploration or contour following, a path planning algorithm is required to guide the motion of the manipulator to achieve the desired task. In case of dexterous grasping and manipulation of soft objects, this operation needs to be performed robustly in spite of possible uncertainties in the robot environment in which a deformable object is neither located at a precise position, nor modelled with high accuracy. For such dexterous manipulation, it is important to consider the difference between the ways of handling a rigid or a deformable object, in particular the major distinction between the definitions of grasping and manipulation respectively [13]. The manipulation of a rigid object requires only the control of its location and therefore grasping and manipulation can be performed independently. Grasping of a rigid object requires the control of grasping forces only, while manipulation of a freely moving rigid object results in the change of its position and orientation. On the other hand, grasping and manipulation interfere with each other in the manipulation of deformable objects. Handling of a deformable object requires controlling both the location of the object and its deformation. But grasping forces yield the deformation of a deformable object, which may change the shape and location of the object. Hence contact between fingers and the object may be lost and grasping may be compromised due to the deformation at the fingertips. Therefore, in the handling of deformable objects, grasping and manipulation must be performed in a collaborative way.

These issues will be exemplified in the context of practical interaction (manipulation) applications in Sect. 4, after the following section describes some of the relevant work on robotic interaction with rigid or soft objects, respectively.

3 Related Work on Robotic Interaction with Rigid or Soft Objects

The literature reports on relatively few research that has been performed in the context of 3D soft object modelling and on robotic interaction with 3D rigid and deformable objects. In [7], an approach is proposed for the in-hand modelling of 3D rigid objects using RGB-D data. An estimate of the position of a robot manipulator, the object and the Kinect sensor is produced at each frame by a Kalman filter based on depth and visual information. These estimates enable the segmentation of the object and its model is built using a series of surfels. Also using surfel models, the authors of [14] propose a registration method on multi-resolution surfel maps that provides a dense displacement field between deformable object shapes monitored in RGB-D images. Petit et al. [11] explore the issue of real-time tracking of 3D elastic objects in RGB-D data. Assuming that a prior segmentation of the object of interest is available, the object is tracked using a graph-cut approach. The iterative closest point (ICP) method is then applied on the resulting point-cloud to estimate a rigid transformation from the point-cloud to a linear tetrahedral finite element model (FEM) representing the object. Linear elastic forces exerted on vertices are computed from the point cloud to the mesh based on closest point correspondence and the mechanical equations are solved numerically to simulate the deformed mesh. A linear isotropic 3D deformable object in interaction with a three-fingered robot hand is modelled by Zaidi et al. [12] as a mass-spring system based on a tetrahedral mesh. The object deformations and the contact points estimation is based on tracking the node positions by solving the dynamic equations of Newton's second law. The authors of [15] measure the stiffness of a 3D planar elastic object by the curvature of surface points from the object geometry and describe the local deformation in terms of a level curve set. In the same line of research, the authors of [16] inscribe markers on the surface of a paper to track its folding in the visual data input. The paper in interaction with a robot hand is represented as a 2D grid of nodes connected by links that specify the bending constraints, namely a resting distance between two nodes and the stiffness coefficient that are tuned manually. Choi et al. [9] propose to tune elasticity parameters of moving deformable balls, painted red against a blue background, by tracking their global position in a video stream and optimizing the differences between the real object captured and its mass-spring representation. In [17], models are acquired and tracked via a webcam. While visual features alone work correctly for some objects, many objects lack sufficient texture for this type of tracking. Sparse sets of oriented 3D points along contours of objects manipulated by a robotic manipulator are monitored in Kraft et al. [18] using a stereo camera, and then predicted based on the motion induced by the robot. Schulman et al. [19] track deformable objects from a sequence of point-clouds by identifying the correspondence between the point-cloud and a model of the object composed of a collection of linked rigid particles, governed by dynamical equations. An expectation-minimization algorithm aims at finding the most probable node positions for the model given the measurements. Tests are

performed in a controlled environment, against a green background that limits its applicability to normal conditions. A solution for robot manipulation of elastic objects that allows to control simultaneously the object's final position (i.e. points of interest over the object and its centroid) and its deformations (i.e. compression distance between points of interest, folding angle and normalized curvature of the object, as estimated by the curve passing through 3 points of interest) is proposed in [8]. In Hur et al. [20], a 3D deformable spatial pyramid model is introduced to find the dense 3D motion flow of deformable objects in RGB-D data without assuming a prior model or template for the object. The point-cloud is corrected with a depth hole-filling algorithm and treated with a Gaussian filtering prior to the computation of a series of perspectively normalized descriptors. The 3D deformable spatial pyramid finds dense correspondences between instances of a deformed object by optimizing an objective function, in form of an energy corresponding to a Markov random field that takes into consideration the translation, the rotation, the warping costs and the descriptors matching costs.

4 Vision and Touch Sensing Systems for Soft Object Interaction

The main challenge in developing autonomous robotic systems able to handle deformable objects originates from the fact that a series of interconnected problems have to be solved, starting with data acquisition, data fusion, data modelling, simulation and validation of objects properties, and the definition and tuning of a control scheme to safely handle the manipulation or the interaction with an object. Figure 1 illustrates these issues in the context of a combined use of vision and

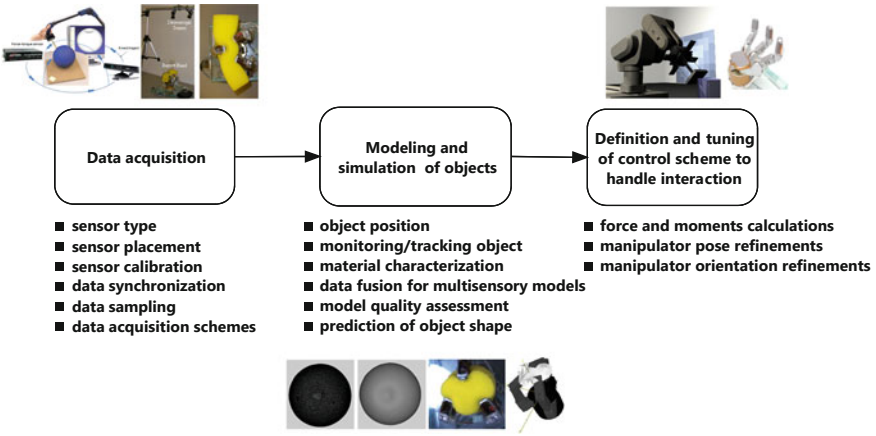

Fig. 1 The series of interconnected problems in rigid or soft object interaction with a manipulator

tactile sensing to enable an autonomous interaction and/or manipulation of rigid or soft objects.

These issues and some possible solutions are discussed in the following subsections.

4.1 Data Acquisition

In the data acquisition process, there are multiple aspects to be taken into account in practical applications, including: the sensor technologies to be used, the placement of the sensors in the environment, the calibration between multiple sensors and sensor technologies, data sampling strategies and selective acquisition schemes to allow for the collection of only relevant data and acceleration of sensing.

4.1.1 Sensor Type and Placement

Realistic, plausible models for objects require the acquisition of experimental measurements using physical interaction with the object in order to capture its complex behavior when subject to various forces. Tests can be carried out based on the results of instrumented indentation tests and usually involve the monitoring of the evolution of the force (e.g. its magnitude, direction, and location) using a force-feedback sensor (i.e. force-torque) (Fig. 2a) or measuring forces applied by the fingers of a robot hand (Fig. 2b, c) accompanied by a visual capture of the deformed object surface to collect geometry data.

In order to collect 3D geometry data, a classical solution offering high precision, are laser scanners. However, they are expensive and the acquisition is often lengthy. In this case, algorithms should be employed to only collect relevant data (Sect. 4.1.3). Stereo systems (Fig. 2b) provide good results, but at the price of a significant computational load and they are prone to important feature matching constraints which often lead to low density depth maps [21].

Moreover, most of the current sensors cannot capture color and depth simultaneously. To overcome these limitations, several attempts have been made to capitalize on the use of the RGB-D Kinect sensor (Fig. 2a, c). The sensor proves to be a simple, fast and cost-effective alternative to collect high density depth maps and the associated color information in a fraction of a second. In spite of the low resolution of the depth map, it generally offers enough precision for most robot manipulation tasks.

Visual data provided by Kinect, has been successfully used for the reconstruction of 3D point clouds of objects by merging data from multiple viewpoints (Fig. 2a), for rigid [7] and non-rigid objects [24, 25] as well. A few open-source [26] and commercial solutions [27] are also available. To collect a full 3D model the sensor is turned around the object of interest following a trajectory similar to the one marked by blue arrows in Fig. 2a and integrating the partial collected point

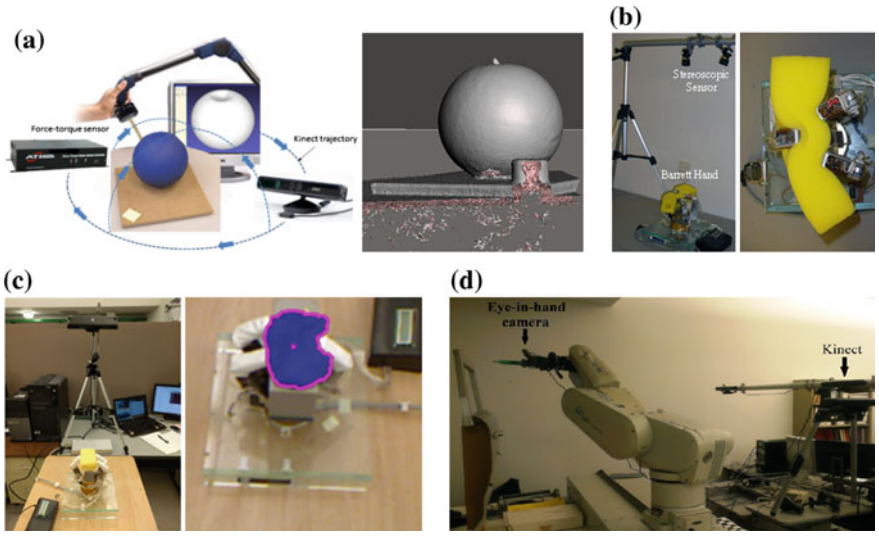

Fig. 2 Multi-sensory vision and tactile data acquisition platforms systems for: **a–c** soft object deformation—**a** Kinect sensor and force-torque sensor collecting 3D data on an object of interest, **b** object handled with a Barrett hand observed by a stereo-system [22], **c** 3D soft object deformation monitoring using a Kinect sensor—and **d** contour following operation [23]

clouds. Alternatively, a partial 2.5D point cloud obtained from a single sensor is sufficient to track contours and detect the object material characteristics (Fig. 2c).

Vision data obtained by Kinect can be also used to locate the object in an unknown environment and to guide the robot arm in proximity to the object. Located behind the robot at a given distance, it can provide the global shape and depth information in complex contour following tasks (Fig. 2d) [23]. This information can, in this case, complement the higher accuracy measurements on the contour location recuperated from an eye-in-hand camera.

4.1.2 Calibration

In the case of large objects, a single Kinect cannot be used to capture the entire surface. When multiple sensors are grouped and operated as a collaborative network of imagers in order to enlarge the overall field of view and allow for modelling large objects, such as automotive vehicles (Fig. 3), a precise mapping between the color and depth components of all the Kinect sensors must be achieved. The internal and external calibration processes proposed in [28] can be used in such situations. The internal calibration corresponds to estimating intrinsic parameters for the color and IR cameras inside a given Kinect, while the extrinsic process provides accurate estimates of the extrinsic parameters in between any respective pair of Kinect devices.

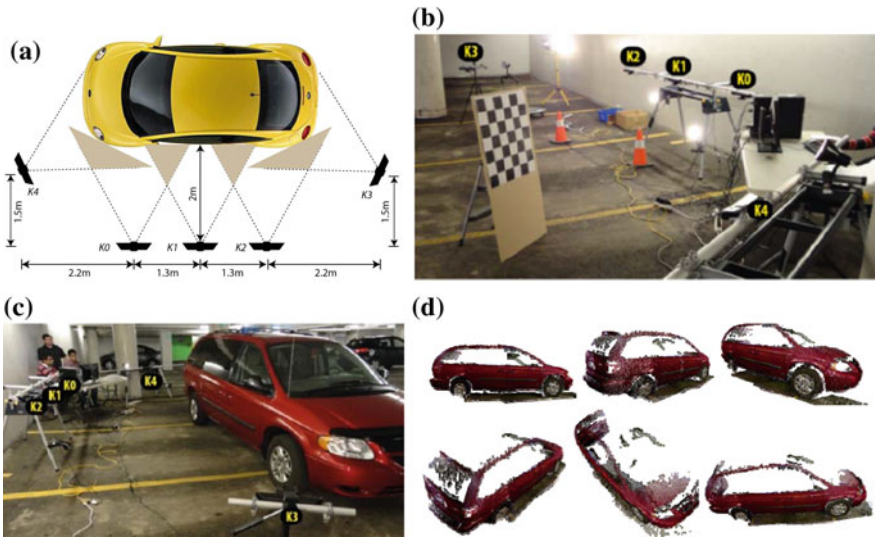

Fig. 3 **a** Sensor system using 5 Kinect sensors K0-K4 for vehicle inspection [28], **b** calibration of sensor using checkerboard, **c** data acquisition over vehicle, and **d** views of reconstructed vehicle

4.1.3 Data Acquisition Schemes

In terms of visual data acquisition schemes, several options are possible. Uniform sampling offers a straightforward solution to ensure complete coverage of a surface. However, in order to achieve adequate sampling density over regions where the local geometry is most likely to vary, the sampling density must be uniformly high over the entire surface and this may lead to inefficiency in certain applications. Each point of the object has an equal chance of being measured in random sampling, but only a lower number of points are actually collected. With an increase in the percentage of sampled points, the cost gets higher to eventually become equal to that of uniform sampling. As well, sampling points randomly might lead to missing important features. In stratified sampling, spaced samples are generated by subdividing the sampling domain into non-overlapping partitions and then by sampling independently from each partition. Such a technique ensures that an adequate sampling is applied to all partitions. It can also be employed in the context of post-processing of large point clouds or meshes [29, 30], where a subdivision of models into grid cells occurs and sample points falling into the same cell are replaced by a common representative. However, all these methods are not meant to be incorporated in the actual sampling procedure, but they rather post-process collected data.

Meant to be incorporated directly in the sampling procedure, a framework to achieve automated selective scanning over large workspaces [31] is illustrated in Fig. 4. A self-organizing neural network architecture, namely a growing neural gas network, adaptively selects regions of interest for further refinement from a cloud of

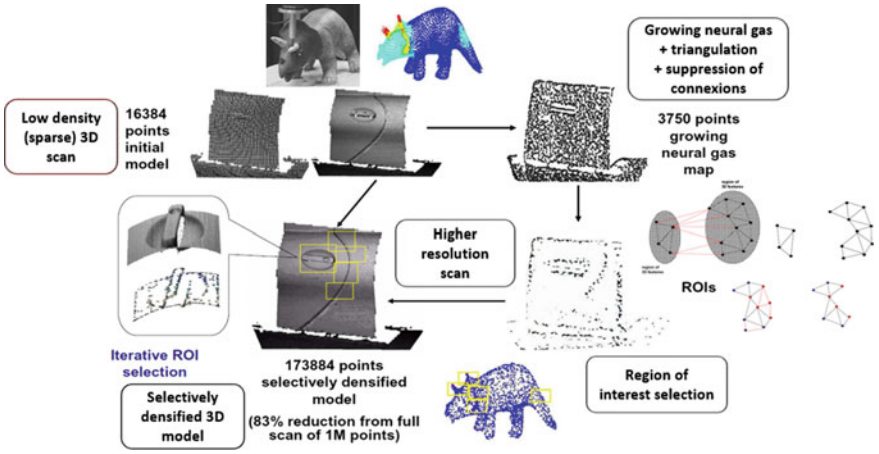

Fig. 4 Selective vision and tactile scanning scheme (adapted from [31])

3D sparsely collected measurements. Starting from an initial low resolution scan of an object, the network is employed to model the resulting point cloud. Those regions that are worth further sampling in order to ensure an accurate model are detected by finding higher density areas in the resulting map. This is achieved by applying a Delaunay tessellation to the resulting growing neural gas output map and by subsequently removing from the tessellation all the triangles that are larger than a set threshold. The latter is automatically computed based on the length of vertices for every triangle in the tessellation.

Rescanning at higher resolution is performed for each identified region (shown in yellow over the car door model in Fig. 4) and a multi-resolution model is then built by augmenting the initial sparse model with the higher resolution data from regions of interest. In this way, a much more compact model can be achieved (i.e. for the car door model in Fig. 4 only 17% of the total number of points that would have resulted from a full-scan) and that contains accurate details only in the regions of interest.

In terms of data acquisition schemes for tactile measurements, current research concerns itself with computer generated objects and their simulation. Conducting strain-stress relationship measurements for objects made of materials that exhibit nonlinear behavior is extremely challenging. Therefore, many applications leave the choice for the selection of elastic parameters to the user, or values are chosen according to some a priori knowledge regarding the deformable object model. This is a subjective process that cannot be applied where accuracy is expected. When measurements of elastic behavior are performed, often a single probing of the object is collected. While this procedure gives satisfactory results for objects made of homogeneous materials, it is unsuitable for objects that are non-homogeneous and have varying elastic properties in different parts of their bodies. Furthermore, the procedure for the acquisition of tactile measurements from each point of an object is

extremely time-consuming. These two aspects explain the considerable interest in finding fast sampling procedures for the measurement of the tactile properties of 3D object surfaces. Appropriate sampling control algorithms should be able to minimize the number of the sampling points by selecting only those points that are relevant to the elastic characteristics.

Due to the human vision-touch experience showing the ability of vision in assisting grasping, handling and manipulation tasks, visual information and particularly the regions of interest into visual information can be used for the collection of tactile measurements. This approach is also justified by the fact that changes in the geometry are very often associated with changes in the elastic behaviour of objects. Using the same framework shown in Fig. 4, if the growing neural gas network is applied not only over the geometry data, but is supplemented with compliance information (an approximate measure of elasticity), during the learning procedure, the model contracts asymptotically towards the points in the input space, respecting their density and thus taking the shape of the object encoded in the point cloud. The regions of interest are identified in a similar manner to the one followed for visual data, but removing from the tessellation not only all the triangles that are larger than a set threshold, but also those which have the same compliance. Due to these properties, if tactile measurements are collected over these identified regions of interest (marked with yellow boxes over the triceratops model in the bottom of Fig. 4), the density of the tactile probing points is higher in the regions with more pronounced variations in the geometric shape. The advantage of such a model is not only to identify relevant sampling points, but also to allow for the determination of clusters of sampling points with similar geometric properties, due to its ability to find an optimal finite set that quantizes the given input space. This provides a robust mechanism that can be extended to model non-homogeneous objects.

It is expected that the collection over points of interest inspired from a visual attention mechanism in vision data [32] could also improve the tactile data acquisition process. The consideration of various aspects derived from psychological studies could also be included in advanced intelligent sensing systems to enable the next generation of intelligent autonomous robotic manipulators. For example, the bias of visuo-haptic estimates towards vision, that is the fact that stimuli are judged to be slightly softer under vision-only condition than under touch-only condition and that the haptic softness perception is more reliable with deformable as compared to rigid surfaces [33] can efficiently guide sensing strategies. Additional testing is required before confirming the effectiveness of these procedures.

4.1.4 Data Cleaning and Synchronization

Data collected using the vision sensors, such as Kinect, often contains undesired elements, such as a background or a surface over which an object is placed, some fixed landmarks required by the software to merge 3D data from multiple viewpoints, or the probing tip when a force-torque sensor is used, as it can be noticed in

Fig. 2a. These can be eliminated in part automatically (e.g. supporting surface and landmarks). However, when a tactile sensor probing tip touches the surface and gets acquired as part of the object, a manual intervention might be required to remove the tip and fill the resulting holes. A mesh processing software (e.g. Meshmixer [34]), can be employed for this purpose.

Due to the different sampling rates found in vision (3D data collection on one side and image analysis to recuperate the angle of the probing tip with respect to the surface on the other side) and force-torque sensors (force magnitude measurements), a synchronization process is also required in order to associate the correct surface deformation with the corresponding force magnitude and angle measurements. This can be achieved by calculating a mean of all the recorded force magnitude and also of the angle of measurement over the time it takes for the 3D object model to be collected. The deformed object model can be considered as a result of the application of a force with a magnitude equal to the mean magnitude and applied at an angle equal to the mean angle value.

4.2 Object Modelling and Simulation

4.2.1 Object Position Recuperation and Segmentation

The object of interest is normally selected in the visual environment using user guidance. A user-selected point can guide the segmentation algorithm towards the location of the object of interest. Such user guidance is common in current tracking literature, going from more extreme approaches in which a prior segmentation of the object of interest is assumed to be available in [11], to cases when the user is asked to crop the object in the initial frame [9]. Other solutions capitalize on the automation of the process, by exploiting the fact that the manipulation of objects takes generally place in relatively controlled environments. Therefore the solutions need to be insensitive to smooth changes in lighting, contrast and background, but do not have to deal with multiple moving objects, or with severe changes in the environment. One such solution based on growing neural gas [6] is illustrated in Fig. 5, where the segmentation is treated as a clustering problem based on color information (HSV color components) and spatial features (X and Y coordinates of each pixel in a color image extracted from a video stream). The HSV color space is chosen because it represents better the color similarities and is able to more accurately identify pixels on the same surface in spite of some differences between their colors due to non-uniform illumination or shading effects. A growing neural gas is adapted over the color and spatial information and the resulting map is then classified as one of two categories: object of interest or background based on the mean HSV value computed for the two clusters and making the assumption that generally the background is darker in color than the object of interest. The latter assumption is generally satisfied due to the controlled environment in which the experiments are performed. To identify the color of the object of interest, the mean

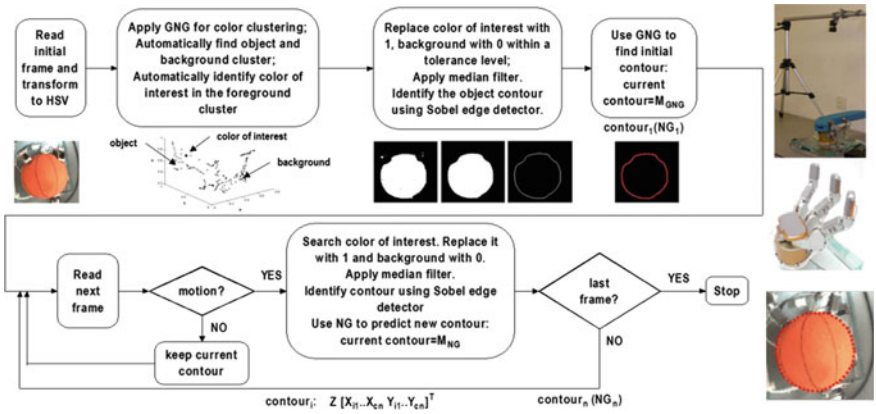

Fig. 5 Object segmentation and tracking (adapted from [6])

is computed for all HSV values in the corresponding cluster. The identified color is then searched in the initial image and over all images in the sequence where movement occurs and all pixels with this color code or a very similar code (within a tolerance level required due to different lighting conditions and due to the fact that the object edges are perceived darker in the image because of shadow effects) are replaced with 1 and the rest with 0 in order to segment the object of interest in subsequent frames. A median filter is finally applied on the result to reduce isolated patches of color and the contour of the object is identified based on the filtered image with the aid of the Sobel edge detector.

4.2.2 Monitoring/Tracking the Object

Once the soft object contour is extracted, it can be tracked over the video sequence as it progressively deforms. To achieve this, a second growing neural gas is initially used to detect the optimum number of points on the contour that accurately represent its geometry. This compact description is employed as an initial configuration for a sequence of neural gas networks that track the contour over each frame in the image sequence in which motion occurs. In each case, a new neural gas network is applied, initialized with the contour of the object in the previous frame, to predict and adjust the position of its neurons to fit the new contour. As illustrated in the flowchart of Fig. 5, this process is repeated until the end of the sequence (i.e. last frame). Due to the choice of a fixed number of nodes used in the neural gas network and to the proposed learning mechanism, the nodes in the contour retain their correspondence with specific points throughout the deformation. This one-to-one correspondence of the points during tracking helps to avoid their mismatch during deformation and ensures a unified description of the contour throughout the frames. Methods such as fast level sets [35] are also an interesting alternative to the

proposed neural gas solution for tracking. However, in this case, the one-to-one correspondence of the contour representation cannot be guaranteed.

The resulting contours (in a number equal to the number of frames with motion) representing each neural gas network can be analyzed in order to detect the object material properties (Sect. 4.2.3) or further associated to the measured interaction parameters (e.g. position of the fingers of a robotic hand and applied force magnitude at each finger) for a comprehensive description and prediction of the object's deformation under manipulation (Sect. 4.2.4).

4.2.3 Object Material Characterization

The contour of the object recuperated from video data or the profile of an object as recuperated from a laser scanner can be used to characterize the object elastic properties based on the following observations: elastic objects return to their initial shape or profile once the interaction with them stops (Fig. 6). Therefore, in order to detect if the object is elastic, the final deformation profile, after the interaction stops is compared to the initial deformation profile, collected in the beginning of the measurement procedure, before any force is applied. If the two contours are almost identical, within a certain tolerated noise margin, the object is elastic. The comparison between the initial profile, the profile under force and final profile after force removal can also be exploited to detect plastic and elasto-plastic deformations. If these are different (more than a threshold to cover the noise in the recuperated profiles), it means that either a plastic or an elasto-plastic behavior occurred. The distinction between the plastic and elasto-plastic behaviors can be made by comparing the final deformation contour with the contour while force is applied. If they are identical, it means that a plastic deformation occurred.

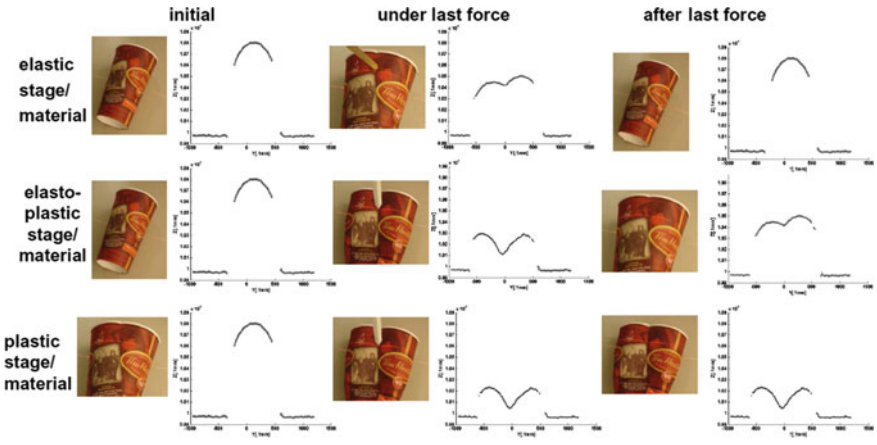

Fig. 6 Object material characterization [31]

If they are different, the material exhibits elasto-plastic properties or the object is within its elasto-plastic deformation stage. If the three profiles are identical, the object is rigid. If a one-to-one correspondence is maintained in the tracked contours, as it is the case of the neural gas solution in Sect. 4.2.2, this comparison is trivial. If one has to deal with profiles or contours of different lengths, an efficient solution to automatically compare them is dynamic time warping [36].

4.2.4 Data Fusion and Deformation Prediction for Multisensory Vision and Tactile Models

Capitalizing on the automated selective scanning framework in Sect. 4.1.3, data selectively collected over regions of interest in terms of vision and tactile measurements can be fused in a representation based on tactile patches (Fig. 7). Such a representation is coherent with psychological studies that have shown that the synthesis of a complex shape is based on the geometric properties of simpler primitives and that this phenomenon occurs in human vision and tactile sensing as well [37]. In such a model, regions of interest from vision are probed at higher resolution and the geometric component of the multi-resolution object is based on the sparse collected data enhanced with these regions. The identification of regions where changes occur in the elastic behavior can lead into the separation of the object in "tactile patches" each exhibiting different elastic properties (i.e. the bottle cap and the bottle body in Fig. 7). A feedforward neural architecture is then used for each patch to capture the relationship between the measured parameters (force magnitude, angle of application, point of interaction of the probe with the object, and object pose with respect to measurement equipment) and the object surface

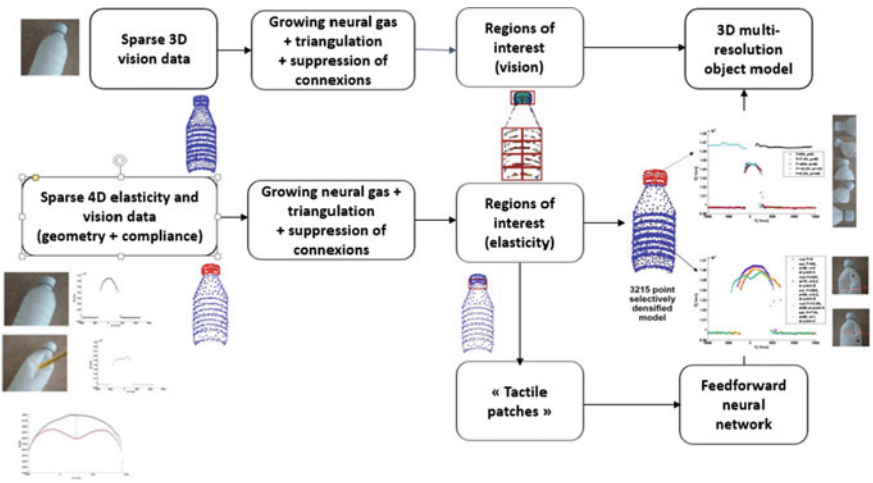

Fig. 7 Modelling of 3D objects as tactile patches (adapted from [31])

deformation. Due to its properties, any of the neural networks is able to provide real-time estimates of the elastic behavior (providing the deformation profile) for those points where the behavior was not probed, therefore eliminating the need for any interpolation of values that normally occurs in any classical model for deformable objects.

The use of neural networks also avoids the problem of recuperating explicitly elastic parameters, which is almost impossible to solve for highly nonlinear elastic materials. The proposed scheme deals easily with piecewise homogeneous materials, due to the existence of tactile patches.

In a similar manner, feedforward architectures can capture and predict the local deformations when the deformed contour is recuperated from a sequence of images of an object under interaction with a robot hand (Fig. 8). If instead of tracking the contour or the profile of an object, the deformation of a 3D object is monitored using a Kinect sensor turning around an object of interest under the interaction of forces exerted with an ATI force-torque sensor (Fig. 2a), a solution to capture implicitly the object behavior capitalizing on a stratified sampling procedure based on the deformation depth, followed by a neural gas-tuned simplification [39] is illustrated in Fig. 9.

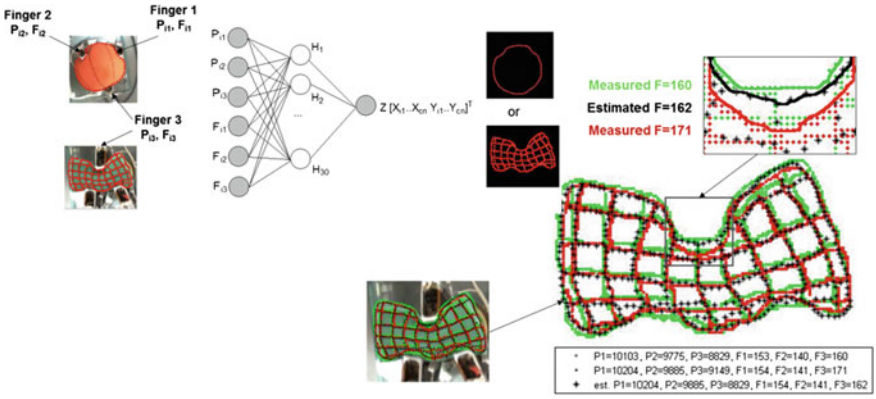

Fig. 8 Prediction of object shape under manipulation with a robot hand (adapted from [38])

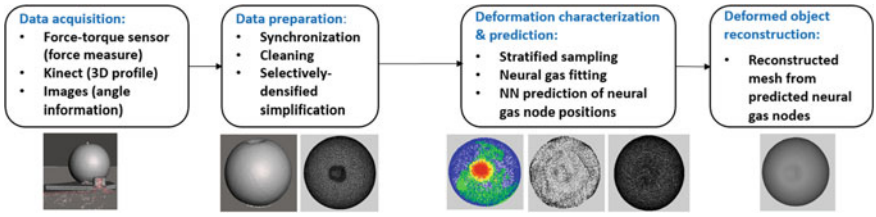

Fig. 9 Data-driven representation of deformable objects under interaction with force-torque sensor

After the cleaning and synchronization of collected data, in order to better characterize the deformation around the probing tip, instead of using the entire collected point cloud, a selectively-densified mesh is first constructed, in which the area around the point of interaction between the probing tip and the object surface is preserved at higher resolution, while the other areas are simplified. This ensures the small deformed area around the probing tip has a higher density of points with respect to the rest of the object's surface. Each deformed mesh is then clustered according to the distance to the initial non-deformed mesh (i.e. blue in Fig. 9 is the closest, and progressing to green, yellow, orange and red as the distance gets higher). Points are sampled randomly but in various proportions from each cluster to identify the adequate amount of data to be used by monitoring the evolution of errors (see Sect. 4.2.5). These proportions are varied by taking into consideration the fact that a good representation is desired specifically in the deformed area and therefore more samples are desired for regions in which the deformation is larger. But this stratified sampling is not sufficient, as the fine differences around the deformed zone might not be appropriately represented, which is the reason why a neural gas-tuned mesh simplification is also applied. The latter is important in order to ensure that fine differences around the deformed zone can be captured in the model. This fitting allows a redistribution of triangles over the mesh such that the fine details are accurately reproduced. The type of model obtained is denser in the region of the deformation (i.e. an average of 97% perceptual similarity with the collected data in the deformed area), while still preserving the object overall shape (i.e. average of 71% similarity over the entire surface) and only using on average 30% of the number of vertices in the mesh. If desired, a feedforward neural network can then be trained to predict the position of the vertices in the neural gas fitted mesh representing each deformed shape of an object based on an applied force magnitude at a given angle.

4.2.5 3D Model Quality Assessment

The quality of a 3D geometrical model can be evaluated from the quantitative and qualitative points of view. In terms of quantitative approaches, Metro [40] allows comparing two models based on the computation of the Hausdorff distance and returns the maximum and mean distance as well as the variance. The second category of quantitative errors can be a form of perceptual error, such as the normalized Laplacian pyramid-based image quality assessment error [41] that takes into account human perceptual quality judgments. As this error is meant to be used on images, images have to be collected over the models of objects from multiple viewpoints and these images can be used pairwise to compute the error. The error measures for each object are then to be reported as an average over the viewpoints. A qualitative evaluation of the results is obtained using Cloud Compare [42] that allows visualizing in an intuitive, color-coded manner the regions most affected by error with respect to its original version.

If the model is a predictive one, in the sense that it is able to predict the deformation of an object for unknown force measurements, the model can be validated qualitatively and quantitatively by using the same metrics as above, but by comparing the predicted mesh with a limited number of real measurements, in which the prescribed forces are exerted on the object.

4.3 Control Schemes for Object Interaction

Combining vision sensors with touch sensing has been explored in recent research to determine the appropriate forces to be applied by the fingers of a robotic hand on a deformable object under manipulation. The goal is to ensure that the hand adapts its behavior to the type of object and to the interaction scenario to achieve an intelligent autonomous manipulation. The stereoscopic vision system depicted in Fig. 2b provides global information by detecting and tracking the deformation of the object in three dimensions. Force and tactile sensors embedded in a Barrett robotic hand are used to provide local information about the deformation at contact points. This knowledge is then used to estimate an object's elastic characteristics and a corresponding control law is defined to maintain a stable and stationary grasp.

From a control system point of view, hand grasping and manipulation processes are carried out by controlling interaction forces at the contact points with an object. Most of the developed control algorithms follow either one of two classical control strategies to solve the force control problem. These are respectively the hybrid position/force control scheme [43] and the impedance control scheme [44]. However, there still exist only limited solutions to the control of robotic manipulation of deformable objects, as the classical approaches require in-depth a priori knowledge of the manipulated object dynamics. The various vision and touch sensing strategies explored in the previous sections can better support the control process by providing live and more comprehensive information about the behavior of soft objects under manipulation.

In general, the typical sequence to grasp and manipulate an object with a robotic hand involves a sequence of logical steps: (i) estimate the object's pose and geometry using 2D or 3D vision sensors; (ii) safely approach the hand with position control to perform the grasp; (iii) determine the contact points and required forces to ensure stable grasp and prevent damages; and (iv) perform manipulation process under force feedback. Vision sensors generally remain involved but mainly to monitor the overall process and detect possible failures. In the initial research that we performed, the focus has been placed on testing the object with a robotic hand immediately after the first contact and the initial grasp is established in order to automatically determine its elasticity characteristics and fine tune the grasping and manipulation process.

Closed-loop manipulation with a robotic hand typically involves a certain form of compliance. Conventional PID controllers have proved successful to ensure stable contact against a compliant surface. However, such an approach requires the

modeling of the hand dynamics through which torques, or forces, at joints are determined to drive the fingers' motion during the manipulation. In a real implementation, estimating the dynamic system's parameters with full accuracy proves very difficult, even impossible, which compromises the tuning of a PID controller. Consequently, adaptive control [45] offers a promising alternative to tackle the modeling and control issues for robotic hand control. Considering parameter estimation and adaptive dynamic control simultaneously copes with varying operating conditions, non-nonlinearities, or un-modeled dynamics, as those characterizing deformable objects. Therefore an adaptive feedback control algorithm is being developed for the purpose of dexterous robotic hand manipulation. While the detailed development of the control scheme remains beyond the scope of this chapter, the multisensory systems presented in Fig. 2b, c, that combine stereoscopic and 2.5D RGB-D imaging with force and tactile sensing, along with the characterization process for deformable objects reported earlier, are being put to advantage to control in real-time the motion of a robotic hand with specific determination of the amount of force to be applied at the fingertips. Fine tuning the interaction at each contact point is critical when manipulating deformable objects in order to ensure stable grasp, integrity of the object, or achieving desired shape forming.

5 Conclusion

The chapter discussed some recent research achievements related to sensing issues and interfacing techniques to enable safe interaction of commercial-grade robot manipulators with objects exhibiting rigid or soft surfaces, as developed over years in the authors' research groups. The main challenges of such systems are described, and recent trends in the literature are presented. The main objective being the development of autonomous robotic systems able to handle a wide variety of objects, the solution is decomposed in a series of interconnected challenges to be solved, starting with the data acquisition, data modelling and simulation of objects, and the definition and tuning of control schemes to handle safely the manipulation or the interaction with the object. Within each of these problems potential solutions are proposed and exemplified in the context of practical applications such as: adaptive surface and contour following, object characteristics identification from video and RGB-D data and dexterous robot hand manipulation of soft objects using the Barrett hand.

Acknowledgements The authors wish to acknowledge the contribution of numerous graduate students and collaborators to the research projects summarized in this chapter, and more specifically: A. Chavez-Aragon, F. Hui, A. Huot, F.F. Khalil, P. Laferrière, R. Macknojia, D. Nakhaeinia, E.M. Petriu, B. Tawbe, and R. Toledo. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI), the Ontario Innovation Trust (OIT), the Ontario Centres of Excellence (OCE), and the Fonds de recherche du Québec - Nature et Technologies (FRQNT).

References

1. M. Seitz, Towards autonomous robotic servicing: using an integrated hand-arm-eye system for manipulating unknown objects. J. Robot. Auton. Syst. **26**, 23–42 (1999)
2. D. Henrich, H. Worn (eds.), *Robot Manipulation of Deformable Objects* (Springer, London, 2000)
3. M. Saadat, P. Nan, Industrial applications of automatic manipulation of flexible materials. Int. J. Ind. Robots **29**, 434–442 (2002)
4. F.F. Khalil, P. Payeur, Dexterous robotic manipulation of deformable objects with multi-sensory feedback—a review. Robot Manipulators, Trends and Development, ed. by A. Jimenez, B.M. Al Hadithi (Vukovar, Croatia, In-Tech, 2010), pp 587–619
5. L. Zaidi, B. Bouzgarrou, L. Sabourin, Y. Menzouar, Interaction modeling in the grasping and manipulation of 3D deformable objects, in *Proceedings of IEEE International Conference on Advanced Robotics, Istanbul, Turkey* (2015), pp. 504–509
6. A.-M. Cretu, P. Payeur, E.M. Petriu, Learning and prediction of soft object deformation using visual analysis of robot interactions, in *Proceedings of International Symposium on Visual Computing, Las Vegas, Nevada, US*, ed. by G. Bebis et al. LNCS 6454 (2010), pp. 232–241
7. M. Krainin, P. Henry, X. Ren, D. Fox, Manipulator and object tracking for in-hand 3D object modeling. Int. J. Robot. Res. 1311–1327 (2011)
8. D. Navarro-Alarcon, H.M. Yip, Z. Wang, Y.-H. Liu, F. Zhong, T. Zhang, P. Li, Automatic 3-D manipulation of soft object by robotic arms with an adaptive deformation model. IEEE Trans. Robot. **32**(2), 429–441 (2016)
9. M.-H. Choi, S.C. Wilber, M. Hong, Estimating material properties of deformable objects by considering global object behavior in video streams. Multimed. Tools Appl. **74**, 3361–3375 (2015)
10. A.S. Prabuwono, S. Said, R. Sulaiman, Performance evaluation of autonomous contour following algorithms for industrial robot, in *Robot Manipulators Trends and Development*, ed. by A. Jimenez, B.M. Al Hadithi. InTech (2010), pp 377–398
11. A. Petit, V. Lippiello, B. Siciliano, Real-time tracking of 3D elastic objects with an RGB-D sensor, in *Proceedings of IEEE International Conference on Intelligent Robots and Systems, Hamburg, Germany* (2015), pp. 3914–3921
12. L. Zaidi, B. Bouzgarrou, L. Sabourin, Y. Menzouar, Modeling and analysis of 3D deformable object grasping, in *Proceedings of International Conference on Robotics in Alpe-Adria-Danube Region, Smolenice Castle, Slovakia* (2014), pp. 504–509
13. S. Hirai, T. Tsuboi, T. Wada, Robust grasping manipulation of deformable objects, in *Proceedings of IEEE Symposium on Assembly and Task Planning* (2001), pp 411–416
14. J. Stuckler, S. Behnke, Perception of deformable objects and compliant manipulation for service robots, in *Soft Robotics*, ed. by A. Verl et al. (Springer, 2015), pp. 69–80
15. C.M. Mateo, P. Gil, D. Mira, F. Torres, Analysis of shapes to measure surfaces, in *Proceedings of IEEE Conference on Informatics in Control, Automation and Robotics, Colmar* (2012) pp. 60–65
16. C. Elbrechter, R. Haschke, H. Ritter, Folding paper with anthropomorphic robot hands using real-time physics-based modeling, in *Proceedings of IEEE International Conference on Humanoid Robots, Osaka* (2012), pp. 210–215
17. Q. Pan, G. Reitmayr, T. Drummond, ProFORMA: probabilistic feature based on-line rapid model acquisition, in *Proceedings of 20th British Machine Vision Conference (BMVC), London* (2009)
18. D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragic, S. Kalkan, F. Worgotter, N. Kruger, Birth of the object: detection of objectness and extraction of object shape through object action complexes. Int. J. Hum. Robots **5**(2), 247–265 (2008)
19. J. Schulman, A. Lee, J. Ho, P. Abbeel, Tracking deformable objects with point clouds, in *Proceedings of IEEE International Conference on Robotics and Automation* (2013)

20. J. Hur, H. Lim, S.C. Ahn, 3D deformable spatial pyramid for dense 3D motion flow of deformable object, in *Proceedings of International Symposium on Visual Computing, Las Vegas, Nevada, US*, ed. by G. Bebis et al. LNCS 8887 (2014), pp. 118–127
21. D. Nakhaeinia, R. Fareh, P. Payeur, R. Laganière, Trajectory planning for surface following with manipulator under RGB-D visual guidance, in *Proceedings of Safety, Security and Rescue Robotics, Linkoping* (2013), pp. 1–6
22. F.F. Khalil, P. Payeur, A.-M. Cretu, Integrated multisensory robotic hand system for deformable object manipulation, in *Proceedings of IASTED International Conference Robotics and Applications, Cambridge, Massachusetts, USA* (2010), pp. 159–166
23. D. Nakhaeinia, P. Payeur, R. Laganière, Adaptive robotic contour following from low accuracy RGB-D surface profiling and visual servoing, in *Proceedings of Canadian Conference on Computer and Robot Vision* (2014), pp. 48–55
24. M. Zollhofer, et al., Real-time non-rigid reconstruction using and RGB-D camera. ACM Trans. Graph. **33**(4), 156:1–156:12 (2014)
25. M. Dou, J. Taylor, H. Fuchs, A. Fitzgibbon, S. Izadi, 3D scanning deformable object with a single RGBD sensor, in *Proceedings of IEEE Computer Vision and Pattern Recognition* (2015), pp. 493–501
26. Microsoft Kinect Fusion. https://msdn.microsoft.com/en-us/library/dn188670.aspx
27. Skanect. http://skanect.occipital.com/
28. R. Macknojia, A. Chavez-Aragon, P. Payeur, R. Laganière, Calibration of a network of Kinect sensors for robotic inspection over a large workspace, in: *Proceedings of IEEE Workshop on Robot Vision* (2013), pp. 184–190
29. D. Nehab, P. Shilane, Stratified point sampling of 3D models, in *Proceedings of Eurographics Symposium Point-Based Graphics*, ed. by M. Alexa, S. Rusinkiewicz (2004), pp. 49–56
30. H. Song, H.-Y. Feng, A point cloud simplification algorithm for mechanical part inspection, in *Information technology for balanced manufacturing systems*, ed. by W. Shen (Springer, 2006), pp. 461–468
31. A.M. Cretu, Experimental data acquisition and modeling of 3D deformable objects, Ph.D. thesis, University of Ottawa (2009)
32. M. Chagnon-Forget, G. Rouhafzay, A.-M. Cretu, S. Bouchard, Enhanced visual-attention model for perceptually-improved 3d object modeling in virtual environments. 3D Res. **7**(4), 1–18 (2016)
33. K. Drewing, A. Ramisch, F. Bayer, Haptic, visual and visuo-haptic softness judgements for objects with deformable surfaces, in *Proceedings of Eurohaptics Conference on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (2009), pp 640–645
34. Meshmixer. http://www.meshmixer.com/
35. Y. Shi, W.C. Karl, A real-time algorithm for the approximation of level-set-based curve evolution. IEEE Trans. Image Process. **17**(5), 645–656 (2008)
36. M. Muller, *Information Retrieval for Music and Motion* (Springer, 2007)
37. J.M. Ehrich, M. Flanders, J.F. Soechting, Factors influencing haptic perception of complex shapes. IEEE Trans. Haptics **1**(1), 19–26 (2008)
38. A.-M. Cretu, P. Payeur, E.M. Petriu, Soft object deformation monitoring and learning for model-based robotic hand manipulation. IEEE Trans. Syst. Man Cybern. Part B **42**(3), 740–753 (2012)
39. B. Tawbe, A.-M. Cretu, Data-driven representation of soft deformable eobjects based on force-torque data and 3D vision measurements, in *Proceedings of 3rd International Conference on Sensors and Applications* (2016) (in press)
40. P. Cignoni, C. Rocchini, R. Scopigno, Metro: measuring error on simplified surfaces. Comput. Graph. Forum **17**(2), 167–174 (1998)
41. V. Laparra, J. Balle, A. Berardino, E.P. Simoncelli, Perceptual image quality assessment using a normalized Laplacian pyramid, in *Proceedings of Human Vision and Electronic Imaging*, vol. 16 (2016)
42. CloudCompare—3D Point Cloud and Mesh Processing Software. http://www.danielgm.net/cc/. Accessed 1 Aug 2016

43. M.H. Raibert, J.J. Craig, Hybrid position/force control of manipulators. ASME J. Dyn. Syst. Meas. Control **102**, 126–132 (1981)
44. N. Hogan, Stable execution of contact tasks using impedance control, in *Proceedings of IEEE International Conference on Robotics and Automation* (1987), pp 1047–1054
45. G. Tao, *Adaptive Control Design and Analysis* (Wiley-IEEE Press, 2003)