Visuo-haptic sensor for force measurement and contact shape estimation

Nicolas Alt and Eckehard Steinbach
Institute for Media Technology, Technische Universität München, Germany

Abstract—Systems interacting with objects in unstructured environments require both haptic and visual sensors to acquire sufficient scene knowledge. Typically, separate sensors and processing systems are used for the two modalities. We propose to acquire haptic and visual measurements simultaneously, with the same standard camera that already observes the scene. The compression of a passive, deformable foam rod mounted on the actuator is measured visually, yielding a 1D stress function sampled along the contour of the rod. Like that, visual and haptic measurements are naturally coherent, the system is passively compliant and the complexity of the sensor subsystems is reduced.

The approach is implemented and validated on a small robotic platform which is used for haptic exploration and interacts with objects by pushing them. In our experiments, friction forces and contact points are determined when the robot comes into contact with one or multiple obstacles. Furthermore, geometric primitives of their footprints are estimated. Using a simple exploration scheme, a haptic map of static and movable objects is built.

I. INTRODUCTION

Autonomous robotic systems typically use dedicated haptic or tactile sensors for the interaction with objects. These sensors are required to determine exact contact points and forces when grasping or pushing an object. However, before an object is manipulated by a robot, it is typically searched and tracked using a visual sensor, such as a camera or a laser scanner. Robots therefore often rely on two separate sets of sensors, based on the proximity of an object. This system design causes a number of problems: A handover point between the systems must be determined, ensuring coherent measurements between the haptic and the visual system. Incoherent measurements may lead to failures, such as incorrect grasps. Furthermore, system complexity and costs are higher due to the additional sensors – for example, haptic sensors require a lot of cabling, if they are attached to several parts of the robot’s surface. Finally, both the visual and haptic modality alone have their specific shortcomings: Visual methods, for instance, often fail for transparent or specular objects, and cannot provide any information about weight or deformability of an object. Haptic sensors only provide sparse information about an object, or require time-intensive exploration steps.

It is feasible to combine these two sets of sensors into a single visuo-haptic system. Cameras allow for remote sensing and are therefore essential for environment mapping, even on simple platforms. We propose to use visual sensors also to detect tactile events in the proximity of the robot, by attaching a deformable material to the robotic actuator. The material deformation can be determined visually with high precision, and since the deformation parameters are known, the acting force can be derived. For low-end robots, the approach provides visual and haptic information from one integrated system, reducing costs by removing additional haptic sensors. More complex systems, on the other hand, benefit from more accurate models of the environment obtained by fusion of visual and haptic data.

In this work, we present a passive sensor made of a plastic foam which allows to measure contact forces and object shape along a line or curve. The mechanical part is completely passive, since it only consists of the foam rod, and it is naturally compliant. Forces applied to the foam result in a deformation, which is measured by a camera mounted above the rod. The sensor takes advantage of an existing camera and requires mounting of an inexpensive piece of deformable material to the robot or actuator. Force is measured in a dense fashion along the entire length of the deformable material. The friction force of objects is determined by pushing them, allowing also to differentiate static and movable objects. Furthermore, the shape of obstacles is estimated by fitting geometric primitives. Shape information is used to create a “haptic map” of the room. The setup is demonstrated on a small robotic platform used to explore a room, see Fig. 1.

Related Work: A tactile sensor in the shape of a fingertip with optical readout has been proposed in [1]. The surface of its dome is made of an elastic material with dots drawn on the
inside. These dots are used as visual features, tracked from the inside by a specialized high-speed camera. Contact forces are calculated from the displacement of the dots. In [2], an optical haptic sensor is presented that measures the traction field – i.e. magnitude and direction of forces – applied to a block of transparent silicon rubber. A camera looks to the inside of this transparent block and tracks markers which are embedded in the deformable material. The cameras and illumination sources used in both works are an integral part of the sensors and do not watch the environment. Therefore, the sensors, as a system component, are still purely haptic sensors; they do not provide the advantages of a visuo-haptic system.

Image edges have been used for a long time as features for objects which are well-described by their contour. Snakes [3] are one popular approach for the detection and tracking of edge features. They consist of several connected points which move in the image to minimize an energy term. This term consists of an image part, attracting points to edge features, an internal part, ensuring smoothness, and a part for external constraints, which limit the motion of the points and allows for initialization. The energy term is minimized by local search, allowing to track edges as they move. If some edge points are very weak, the smoothness term ensures that the corresponding points are “dragged” by their neighbors. Initialization is done by a human or by another system which has some a priori knowledge of the estimated object position.

A number of approaches build on this idea: Active shape models (ASM) [4] learn shape priors for a certain object class (such as hands or faces) and integrate this prior in the energy term. Hence, the model allows only for realistic deformations. Dynamic shape priors, as proposed in [5], increase temporal stability and consistency.

Material properties of plastic foam have been studied intensively, see e.g. [6]. The strain-stress relation for these materials is highly non-linear, typically showing 3 regions: A region of linear elasticity for very low strains, a plateau region showing high sensitivity to stress, and a region of densification when very high stress is applied. This behavior is observed in the plot Fig. 2.

II. VISUO-HAPTIC SENSOR

The sensor consists of a deformable material, such as plastic foam, attached to a robotic platform which explores obstacles in the environment haptically, i.e. by driving towards and pushing them. Contact points and forces are determined by visually measuring the deformation of the foam. To that end, a standard camera is pointed to the foam material, as shown in Fig. 1. Visual processing yields the deformation along the foam in world coordinates as described in Sec. III.

The sensitive part of the sensor consists of a passive foam rod (orange part in Fig. 1) which is roughly 25cm in length (major axis) and exhibits a cross section of $w \times h = 2 \times 1 cm$. The cross section is chosen based on the deformability of the material and the required force range. One of the long sides is attached to a rigid mounting plate, which may be straight or curved, and the opposing side gets in contact with objects.

![Fig. 2: Experimentally determined strain-stress relation for the plastic foam. Data points are obtained using the camera as outlined in Sec. III while increasing (red) or decreasing stress (blue). The curve shows a third-order polynomial which has been fitted to the data points in the central region.](image)

The direction of exploration, denoted $x$ in Fig. 1, corresponds to the dominant (forward) motion of the platform on the floor. Forces act along vectors $d(s)$ and deform the foam rod along its width $w$. Deformation is expressed as a scalar function $\delta(s)$ of displacement [in m] which is densely sampled along the mounting line or curve of the sensor $s$. The normalized strain of the material is $\frac{\delta(s)}{w}$. Sheer forces parallel to the mounting curve are neglectable in this setup and are not measured.

For calibration, the deformation behavior of the foam material is tested by pushing a plate of $2 \times 1 cm$ towards it with increasing force, see data points in Fig. 2. As expected from literature, the curve shows three different regions, see discussion in Sec. I. Here, we mainly rely on the plateau region for normalized strains in $[0.1, 0.5]$, which corresponds to a reasonable range of forces for the application at hand. Additionally, this region allows for the most accurate measurements, since the sensitivity of the material to force changes is largest. Note also that the curve exhibits a strong hysteresis effect, depending on whether forces are increased (red curve) or decreased (blue points). Therefore, we only measure the displacement while forces are increased. The characteristic curve is repeatable for multiple experiments.

A third-order polynomial $f$ is fitted to the red data points, yielding a phenomenological model for the strain-stress relationship which is depicted as a red curve in Fig. 2. The elasticity region (for small strains) is not correctly modeled, since it cannot be measured exactly enough. Once the strain goes to the densification region, the sensor is in saturation. The total contact force is obtained by integration over the stress (or pressure), using model $f$, sensor width and height $w, h$, as well as the mounting curve $s$:

$$F = h \int_s f(\frac{\delta(s)}{w}) \, ds$$  \hspace{1cm} (1)

III. VISUAL PROCESSING

As depicted in Fig. 1, the camera provides a top view of the deformable foam rod and the mounting plate on the robot.
Additionally, it observes objects coming in contact with the robot. The mounting curve $s$ and the force vectors are roughly parallel to the image plane of the camera, providing optimal resolution of the measurement. The outer contour of the foam rod deforms when it comes into contact with an object, and the amount of deformation is determined using tracking of visual edges. Additionally, the inner edge between the foam rod and the mounting plate (e.g. the rigid surface of the robot) is tracked to obtain a reference (see below).

The use of image edges is most feasible for the application at hand since the foam has no stable inner visual features. Contour detection based on image edges is stable regardless of lightning conditions, except for complete darkness. The edge strength varies considerably depending on the visual appearance of objects touching the sensor, which must be accounted for by the algorithm. In the rare case where brightness, color and shading values of the foam rod equal exactly those of an object, the edge at the foam’s contour would disappear. To prevent this case, it is possible to work with a foam material that changes its color along the major axis.

Edges are tracked using the well-known concept of snakes, see Sec. I, which consist of connected points “snapping” to image edges. We track points along the contour of the foam spaced about $3\text{mm}$ apart, which allows for an accurate representation of possible deformations. After initialization (see below), snake points move within a certain local neighborhood to iteratively minimize an energy term which consists of several parts, see Eqn. (2): First of all, the negative amount of edge strength $E_{e}$ tries to keep the points on images edges. In order to avoid snake points snapping to strong internal edges within an object, $E_{e}$ is limited to the average edges strength $\bar{\sigma}$ of all snake points. A smoothness term $E_{s}$ accounts for the fact that the slope of $\delta(s)$ is limited. Furthermore, there are no edges within the foam rod, i.e. in between the inner and outer contour. Therefore, an energy term $E_{j}$ penalizes points jumping over edges along the path from the closest point of the reference snake $p_{k}'$ to the current point. Finally, $E_{c}$ constrains the motion to the vector of deformation $d$, which is perpendicular to the reference contour of the foam rod. As this is the major direction of deformation, snake points stay on the same physical point on the foam, and the sampling density remains constant. The total energy is evaluated and minimized in a local neighborhood $\xi = [x, y]^T$ and expressed as follows:

$$E(\xi) = w \cdot [E_{e}, E_{s}, E_{j}, E_{c}], \text{ with}$$

(2)

$$E_{e} = - \min(\lVert \nabla G \ast i \rVert, \bar{\sigma}),$$

$$E_{s} = (p_{k-1} - 2\xi + p_{k+1}) \cdot (p_{k-1} - 2\xi + p_{k+1}),$$

$$E_{j} = \frac{1}{\bar{\sigma}} \int_{x = p_{k}'}^{\xi} |\nabla G \ast i| dx,$$

$$E_{c} = \begin{cases} 0, & \text{if } \lVert \xi - p_{k}' \rVert \leq \delta(s) \text{ and } \bar{\sigma} \\ \infty, & \text{otherwise} \end{cases}$$

Where $i$ – image, $\nabla$ – gradient operator, $G$ – Gaussian blur operator for noise reduction, $P$ – projection operator, see Eqn. (3). Weights $w$ are set such that energy terms are in $[-1, 1]$ within the search space. A 1D constraint is imposed by $E_{c}$, so the search for the optimum is fast even for large neighborhoods. Processing at framerate (30 Hz) poses no problem to a mid-range Intel i5 platform. Note that it is not feasible to integrate shape priors in the energy term, as in more recent work using snakes. The contour of the foam is solely determined by the shape of the obstacles, and the correlation of close-by values of $\delta(s)$ is accounted for by $E_{e}$.

The approximate position of the foam rod in the camera image is typically known from a geometric robot model. Otherwise, it may be located using markers. First, the inner snake is initialized by adding points iteratively at a constant distance and having them snap to edge pixels. Points $p_{k}'$ of the inner snake serve as a reference position of the sensor base – which might change slightly due to movement of the mounting plate or the camera. Next, points of the outer snake are initialized slightly outside of the inner snake. To allow for varying rod widths, these points are pushed away from the inner snake by an additional energy term until they reach the stable outer edge. The idle state of the outer snake is used as the zero reference of displacement $\delta$.

Pixel positions on the snake are converted to real-world coordinates using the intrinsic matrix of the camera $K$ and a coordinate frame $(P_0, R)$ at the tip of the foam rod spanned by the exploration vector $x$ and with $y$ parallel to the floor. The mounting curve $s$ and deformation vectors $d$ lie on the $x-y$ plane of this frame. In a robotic system $(P_0, R)$ are determined from the extrinsic camera parameters and the geometric robot model. Otherwise, the pose can be determined by applying markers. Like that, it is enough to obtain 2D information from the camera. The projection operator $P$ converts a point in the image $p = (x, y, 1)^T$ [in pixel] via the camera-centered coordinate frame $P(X, Y, Z)$ [in meters] to a point $P^F(X', Y', 0)$ on the $x-y$ plane of the reference frame with normal $n = R_{x,y}$:

$$P = \frac{P_0 \cdot n}{p' \cdot n}, \quad p' = K \cdot p$$

(3)

$$P^F = R(P - P_0), \quad P : p \rightarrow P^F$$

Visual measurements are noisy due to image noise, shaking of the camera and location uncertainty of the image edge. For individual snake points noise with $\sigma_1 = 70\mu m$ is observed in the static case. Additionally, edge locations may be biased by large illumination changes or strong intensity discontinuities on objects close to the sensor. Bias effects depend on the surrounding scene and change with a much lower frequency than image noise. Taking into account these effects, at worst, an edge uncertainty of $\sigma_2 = 0.5\text{mm}$ is observed.

IV. FITTING OF GEOMETRIC PRIMITIVES

In typical household or office environments, a robot encounters obstacles such as walls, boxes, table legs, cylindrical trash bins or vases. As these artificial objects are aligned perpendicular to the floor and exhibit symmetry, a 2D “footprint” is
often sufficient for tasks such as mapping or navigation. In case of contact with the sensor, the impression in the foam corresponds to a partial 2D contour of the object. Typically, a connected set of deformed snake points (deformation set) corresponds to a single object.

Two-dimensional primitives are fitted to points \( P^F \) from the deformation set. They serve as a rough approximation of the shape of an object beyond the contact points. The following primitives, which can represent typical objects, are used: Lines (for walls, large boxes), line segments (small boxes), corners (boxes, table legs) and circles (trash bins, bottles). Line-based primitives are fitted into the point set using least-squares minimization. A line segment is used if the center points of the primitives are fitted into the point set using least-squares minimization. A line segment is used if the center points of the deformation lie on a line, but the end points do not. Corners are described by two lines which intersect at the point of maximal deformation. Circles are fitted using the algebraic approach described in [7].

Each candidate primitive is tested on how well it fits to the observed points \( P^F \). A score is calculated, based on the mean squared distance between observed points and closest model points as well as the number of points \( n \) used by the model:

\[
g = \exp \left( -\frac{1}{|D|} \sum_{k \in D} \frac{1}{\sigma^2} d^2_{M_k}(P^F_k) \right) \cdot \frac{n}{|P^F|} \tag{4}\]

Where \( d_{M_k} \) – distance of a point from model \( M_k \), \( D \) – deformation set. The primitive with the highest score and \( \gamma > 0.5 \) is chosen and used for the generation of a 2D map with fixed and movable obstacles (see below).

\[V. \text{ Experiments} \]

First, the mobile platform equipped with the proposed visual-haptic sensor is driven against several different obstacles in a room, such as boxes, bottles, tables, doors and walls. Contact with an obstacle is detected when the foam rod starts to deform. In that case, the speed of the platform is reduced to allow for more accurate measurements and to avoid damage to the object. The movement is stopped completely if one of the following conditions becomes true: a) the amount of deformation goes beyond an upper limit, i.e. the strain goes to the densification region, b) the total force, see Eqn. (1), goes beyond the pushing capabilities of the robot, c) the robot moves for a distance larger than the width of the sensor. In the latter case, a movable object is observed, and the measured force corresponds to the friction force of the object, see Fig. 3.

In the other two cases, the explored object is static – at least for the capabilities of the platform.

For both kinds of objects, during the halt of the platform the fitting of geometric primitives (Sec. IV) is triggered. Measurements are most stable in the static case and allow for the best possible fit. Examples of some explored objects are shown in Fig. 4, together with the fitted geometric primitive. Some of the objects are (partially) transparent bottles, which would cause problems to purely visual methods. Note that the fitted circles match the size of the bottles and the concrete column quite well.

In a second experiment, the platform is used for haptic exploration of an office environment. Location information is obtained from the robot odometry and a gyroscope in the platform. In larger workspaces, cues from a visual localization system should be added to reduce drift. A naive exploration strategy is used, having the platform drive along the boundary of the room: Whenever an obstacle is hit, the robot is stopped, moves back, rotates counterclockwise to drive in parallel to the obstacle for a while, rotates clockwise and drives straight again. As described above, it is determined whether the obstacle is movable or static by pushing it, and its geometric primitive is stored.

Haptic maps showing static and movable objects are generated during exploration and updated during each contact event. The maps show the confidence score of occupancy, where −1 corresponds to free space and 1 corresponds to occupied space. At each contact event, the stored geometric primitive is added to the map as follows: The binary occupancy map \( n_{\text{O}} \) is determined for the primitive – e.g. the area inside a circle is marked as occupied (1), and the outer parts are marked unoccupied (−1). Beyond the contact points, primitives represent just predictions of the environment, which become more inaccurate with increasing distance from the closest contact point. This is modeled as a normalized distance map \( n_d \in [0,1] \), which shows the confidence of the geometric primitive. Normalization is performed based on the extend of the contact area. Like that, a primitive – such as a line – can predict a larger geometry, yet the prediction is quickly updated once a more confident measurement is available. The current observation is integrated into the global map \( m \) as follows:

\[
n' = (1 - cT \cdot n_d) \cdot m + cT \cdot n_d \cdot n_{\text{O}} \tag{5}\]

\[n_d(x, y) = \exp \left( -\frac{1}{D} \min_k ||(x, y, 0) - P^F_k|| \right)\]

The factor \( c \) determines how quickly old measurements are replaced and is set to 0.5 here. Transformation \( T \) aligns the new observation with the map frame, using the current pose of the platform. The distance map \( n_d \) is calculated from the distance to the closest contact point \( P^F_k \), normalized by the diameter \( D \) of the contact area.
Fig. 4: Leftmost: Camera view from the platform while touching different objects. The remaining pictures are split, showing image gradients in the left half, and the original image in the right half. Additionally, the fitted geometric primitives are depicted – from left to right: Line, line segment, corner, multiple circles.

Fig. 5: An office scene is mapped by haptic exploration and fitting of geometric primitives. Occupied areas are depicted in red, free areas in green and movable objects in blue. Color intensity corresponds to mapping confidence. Black edges show strong borders between free and occupied areas.

Results are shown for a small office space in Fig. 5 which was explored using 22 contact events. The structure of the room with the walls, column and door is accurately represented, even though some shapes extend beyond the room. One of the reachable chair legs appears as a small blob left of the paper bin. The discovered movable objects (paper bin near the right wall and two bottles in the left part) are depicted in blue. The bottle in the upper left area was explored twice and moved in between by the platform. Erroernously, a corner primitive rather than a circle is fitted, as the object is small. Note that visual mapping systems would easily overlook this bottle, since it is completely transparent, apart from the cap.

VI. CONCLUSION

In this work a novel visuo-haptic sensor is presented for the exploration of obstacles in household and office environments. It consists of a deformable, passive material which is measured visually using a standard webcam. This low-cost sensor can be easily attached to mobile platforms to acquire haptic information about the environment, taking advantage of already existing cameras. We use the sensor to determine the friction forces of movable obstacles and to estimate their footprints by fitting geometric primitives. Using a simple exploration scheme in a small office space, a haptic map of static and movable obstacles is built.

In future work, additional cues will be derived from the visual data provided by the camera. Using for instance a structure-from-motion approach, 3D models of obstacles can be generated and fused with the haptic geometric primitives. Deformable obstacles can be modeled by tracking deformations in the 3D model while the pushing force increases. Finally, a visual mapping system, such as [8], will be integrated, enabling the use of a more intelligent exploration scheme. Fusion of visually and haptically acquired maps will allow for richer and more accurate environment representations.

REFERENCES