Optical Skin For Robots: Tactile Sensing And Whole-Body Vision

Akihiko Yamaguchi and Christopher G. Atkeson
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, 15213
info@akihikoy.net, cga@cmu.edu

***Abstract*—We explore combining optical tracking of skin deformation to implement tactile sensing, and seeing through transparent skin to implement proximity sensing, as well as additional sensing modalities including vibration sensing (accelerometers and gyros) and force sensing (strain gages). Issues for discussion in a workshop on *Tactile Sensing For Manipulation* include the importance of proximity sensing and imaging surfaces of objects and obstacles for manipulation, in addition to tactile sensing of actual contact; the viability of actually processing tens of video streams for a hand or hundreds of video streams for whole-body vision; whether future lenses and optical paths will be short enough to make optical skin practical; what the desirable mechanical characteristics of skin should be; whether better sensors are more important than advances in AI and learning; and how we should evaluate proposed sensing systems.**

I. Motivation

One motivation of this work was the lack of contact sensing in the Atlas robots used in the DARPA Robotics Challenge. The tactile and force sensing in the three types of hands we worked with and the robot feet were poor. In a task like getting out of a vehicle with fewer degrees of freedom and less flexibility than a human, sensing contact location, force, and shape is critical. The Atlas robots had no skin sensing. We installed cameras on the robot wrists and knees, which helped greatly. If body-mounted cameras are so helpful, let's take this idea to its logical extreme of as many cameras (eyeballs) all over the body as possible.

A second motivation was our work on tactile sensing for soft inflatable robots. Rather than put sensors in the robot skin directly, why not use cameras inside the robot looking out to detect and track contact? For example, in the Disney movie *Big Hero 6,* there is a scene where the camera point of view is inside the inflatable robot Baymax (inspired by our work [3]) which shows how imaging a contact from the inside of a robot looking outward through transparent or translucent skin can reveal information about contacts.

We have been exploring using small cameras inside a robot looking out to track deformation of and see nearby objects through a transparent elastic skin (currently Smooth-On Clear Flex 30 with a protective layer of Saran Wrap). We started with developing sensing fingertips (FingerVision, Figures 1 and 2, [26]) with an eventual goal of developing full hand and then full body tactile and proximity sensing (Whole-Body Vision, Figure 3). One key idea is to use transparent skin, to enable proximity sensing. Another key idea is to separate electronics and wiring from the deformable part of the skin, so the electronics and wiring are not repeatedly flexed and thus break much less, and the skin can be easily replaced (humans replace the outer layer of their skin monthly, and more quickly during wound healing). This approach addresses the short lifetimes of currently available tactile sensing and

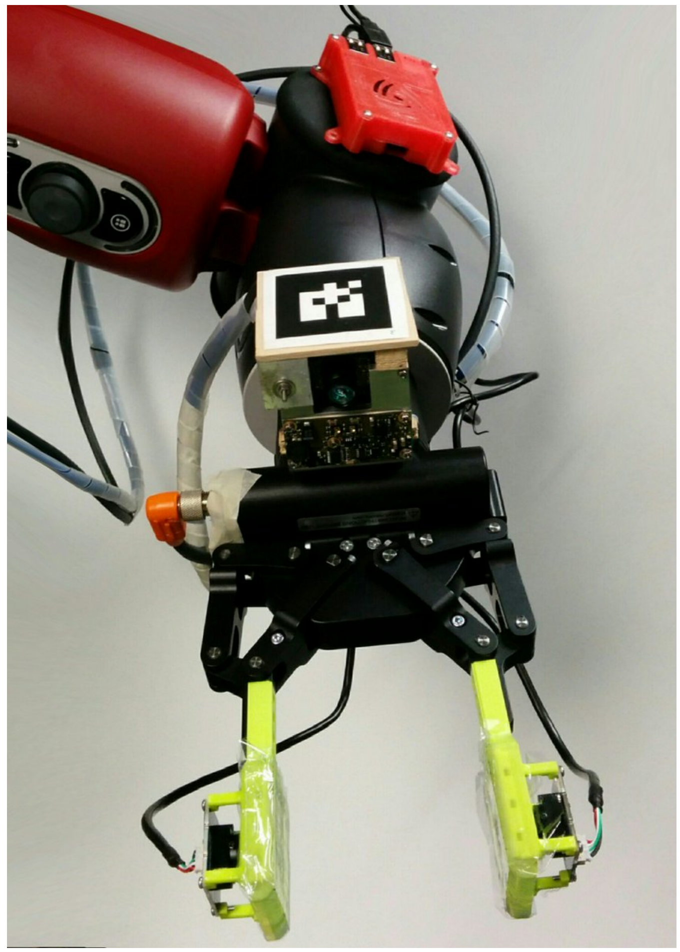

Fig. 1. FingerVision implementation using USB cameras (ELP-USBFHD01M-L180). Please look at https://youtu.be/L-YbxcyRghQ, https://youtu.be/TAA4YJqEOqg, https://youtu.be/uy32tO9e7O4, https://youtu.be/aaDUIZVNCDM, and https://youtu.be/FQbNV549BQU for videos of FingerVision in action.

the high cost of skin repair. A third key idea is to use high resolution imaging rather than single pixel (infrared range finder [18], for example) or low resolution imaging. The system cost is about the same, and we believe the increased processing requirements can be handled by low cost and low power GPUs that are becoming available. A fourth key idea is to put eyeballs all over the body, rather than just on the head. This greatly reduces occlusion during manipulation and physical interaction. Our approach also minimizes component and wiring counts and reduces the risk of electromagnetic interference. We can easily add internal (LED) lighting to compensate for a lack of external lighting, and also implement vision approaches based on structured lighting. Our eventual goal is to use large numbers of "lensless" or "flat" cameras, so that the skin can be quite thin and flexible [2, 6, 19].

II. WHAT HAVE WE LEARNED SO FAR, AND WHAT QUESTIONS REMAIN?

We have learned from our FingerVision implementation. Tangential positions, slip, and forces are all measured well. Normal forces are measured less well, and we are planning to build traditional strain gage based load sensors into the mounts of our fingertips and skin in general to improve normal force measurement (MS5803-14BA [23], a commercially available pressure sensor which is similar to but complements the Takktile tactile sensor based on a barometric (1.1 bar maximum) pressure sensor by measuring up to 14 bar (more range but less resolution and sensitivity)). The normal force sensing is done at the skin mounting points out of view of the camera. The bandwidth of optical sensing using video cameras is limited by the frame rate. It is useful to use vibration sensors (microphones, piezoelectric pickups, accelerometers, or gyros) to capture higher frequency information.

We are still searching for transparent materials with good mechanical properties (friction, elasticity, damping, plasticity, ...) for the outer skin layer. Currently we are experimenting with different transparent urethanes, silicones, and sorbothanes. We initially used 33x33mm cameras with M12 fisheye ($> 180°$ FOV) lenses (Figure 1). These cameras include an image processor for on-chip image compression and provide an MJPEG output video stream on a USB network connection. We are searching for more compact camera modules that have similar fisheye lenses. We will probably have to sacrifice the onboard image compression and switch to a cell phone camera using parallel MIPI or CSI-2 camera interfaces providing only raw image data to get a smaller camera (Figure 4). Our current camera, based on the Raspberry Pi camera (OV5467 chip), is less than a cubic centimeter including the M7 fisheye lens. We expect to shift to "flat" cameras as they become commercially available [2, 6, 19].

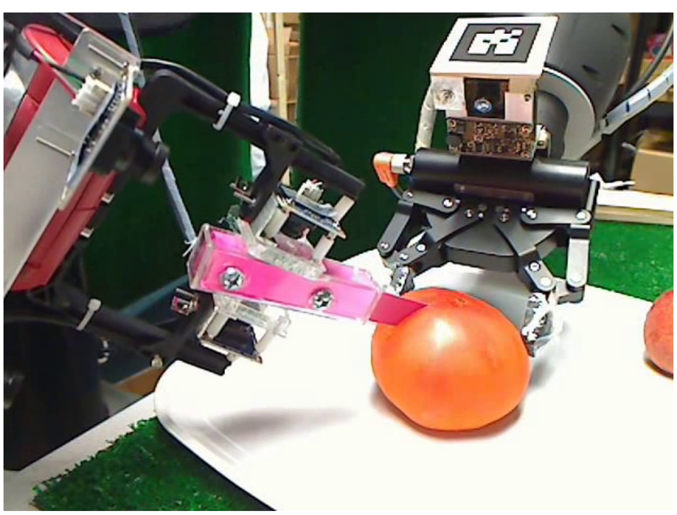

Fig. 2. FingerVision used in cutting a tomato.

Fig. 3. *Whole-Body Vision* in Japanese (Hyakume) and Greek (Argus Panoptes) mythology. Each figure has 100 eyes spread all over their body.

III. SIMPLE IMAGE PROCESSING TECHNIQUES

We are using OpenCV library routines to rapidly prototype vision algorithms. For tracking the markers placed on the surface of the soft skin, we use blob tracking (Figure 5). It consists of two processes: calibration to detect initial marker positions, and tracking the marker displacements. In both processes, the camera image is rectified to compensate for the

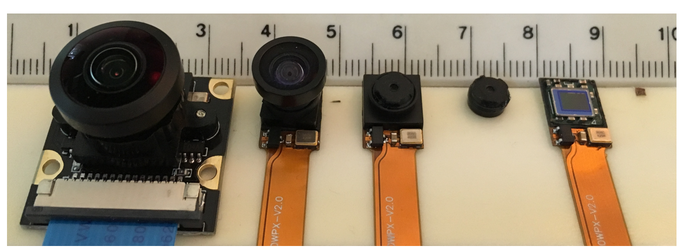

Fig. 4. Parallel (CSI-2) output cameras based on the OV5647 imaging chip. From left to right: 1) M12 fisheye lens, 2) M7 fisheye lens, 3) M6 low profile lens mounted (approximately a 50° FOV), 4) M6 low profile lens, 5) the imaging portion of the optical system, which is about 8x8x1mm. The small square at 9.5mm is the approximate size of the NanEye camera, which costs about $2000.

distortion caused by the fisheye lens, and then converted to a grey-scale image. During calibration, a blob detection method implemented in OpenCV (`cv::SimpleBlobDetector`) is used with a white sheet covering the sensor to remove the background. The calibration time is less than 1 second. Marker tracking is done independently per marker. We assume the marker is in a small region around its previous position, and apply the same blob detection method. If the marker movement is unexpectedly large, we reject the result. We also track the size of the blob to reject errors.

For detecting slip, we use a background subtraction method. We also considered optical flow, but background subtraction was better in some cases where the object did not have sufficient texture. Since the background subtraction perceives both the object movement and the background shift caused by gripper movement, we need to distinguish the object from the background. First we build a background model, and then we adaptively construct an object model. Finally we use the object model as a mask to extract the object movement. Both the background and the object models are represented as color histograms. We use the OpenCV routine `cv::BackgroundSubtractorMOG2` for background subtraction.

IV. Historical context

The idea of *Whole-Body Vision* is thousands of years old (Figure 3). We want to go beyond palm, wrist, crotch, and knee cameras to cover the entire robot with cameras. The idea of using imaging sensors for tactile sensing is decades old. Many have proposed the combination of structured light, markers, and multiple imagers to estimate deformation in internally transparent skin, including [21, 16, 20, 1, 10, 13, 4, 11, 15, 14, 27, 17]. Our method is close to these approaches. An important difference is the total transparency of our skin including the outer surface, which gives us vision of external surfaces not in contact as well. Previous work used an opaque top layer on the skin to block external light as it would affect marker tracking and other measurements internal to the skin. We solve the marker tracking problem under natural external scenes using computer vision in order to make use of totally transparent skin.

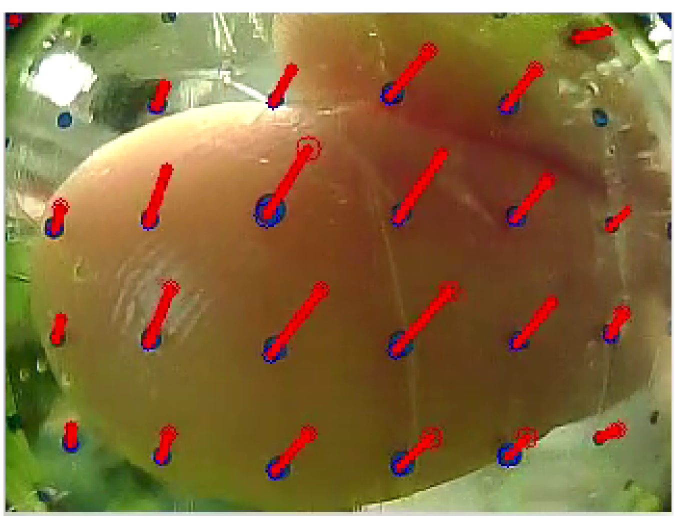

Fig. 5. The image shows what the sensor sees when a human finger presses against it. The red lines show (exaggerated) marker displacements.

Embedding optical range finders (single pixel depth cameras) in the skin of a robot to provide whole-body distance sensing was proposed more than 40 years ago and continues to be developed [12, 24]. Another sensor with fully transparent skin is proposed in [18] where single pixel infrared range finding sensors are used instead of a camera. The idea is to measure distances between the sensors and an object, and estimate the deformation of the transparent skin from the distance. Vertical contact forces are estimated from the deformation. If there is no contact with an object, this sensor gives the distance to the first object in the path of its ray. We believe full imaging can be used instead of just single pixel range finding for great benefit and not much added cost. Although this sensor and ours have different sensing modalities and ranges, we can share ideas; e.g. we could embed distance sensors around the cameras.

V. Multimodal sensing

Previous skin and tactile sensing projects typically focused on one or only a few types of sensors. We propose combining many types of sensors. in addition to using visible light optical sensors to measure skin deformation. Electrical properties of the skin including resistance, capacitance, and inductance can be measured. Capacitance sensors are often used on mouse pads, touch screens, and other touch-based interfaces. Printed antennas and inductive coils similar to what are used in wireless RFID anti-shoplifting devices may also be usefully placed on the skin surface or embedded in the skin to measure static and dynamic electric fields, as well as magnetometers or Hall effect sensors to measure magnetic fields. Radar chips are being developed for monitoring respiration and hand gestures at a distance [8]. Pressure sensors can be used to measure skin forces. Given the low cost and small size of far infrared (thermal) imaging sensors, there is no longer a need to restrict sensing to just visible light, or just near infrared. For robots that work with people or processes involving changing temperature (e.g. cooking) imaging in the infrared spectrum is useful (Figure 6), as well as skin temperature sensors. Small time of flight depth cameras are now available (e.g. DepthSense 541A of SoftKinetic Inc.). Ultrasound transducers can be built into robot skin to image objects and human tissue that are in contact to avoid damage, injury, and pain. Accelerometers, gyros, IMUs, piezoelectric sensors, and microphones are useful to detect vibrations, localize contacts, and recognize texture and material properties [22, 9]. Accelerometers are also useful to measure orientation relative to vertical (given by the direction of the gravity vector). High speed imaging used in optical mice (essentially using very high frame rate cameras with low angle of incidence illumination (Avago ADNS9800, for example)) can detect horizontal skin, object, and environment movement. Hairs or whiskers glued to piezoelectric sensors or optically

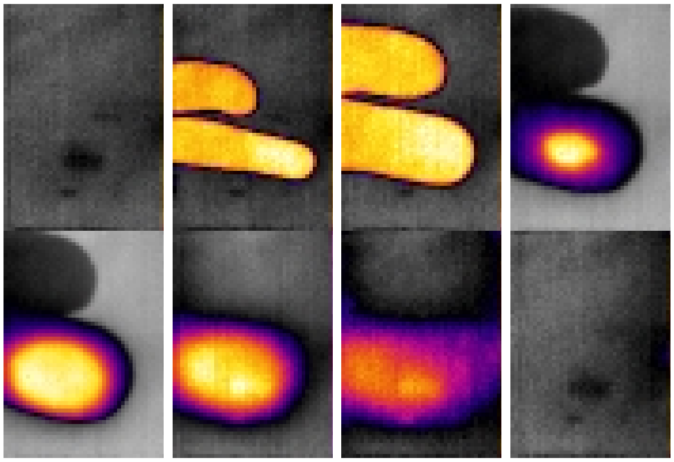

Fig. 6. A time series of far infrared (thermal) images from a camera looking through a skin at a finger touching the skin. The skin is transparent in the far infrared spectrum, as well as for visible light. Due to the large dynamic range of the sensor, each picture is scaled so the hottest value is yellow. Note that it is easy to tell which finger actually touches the skin, that the skin is heated up by contact (very quickly), and there is an afterimage as the skin cools off. The camera used was a Lepton LWIR module [7].

tracked provide mechanical sensing at a (short) distance. It may also be possible to embed mechanical elements in the skin that click or rasp when deformed, and use microphones to track skin deformation. We will explore deliberately creating air and liquid (sweat) flows (both inwards and outwards) for better sensing (measuring variables such as pressure, conductivity, and temperature) and controlling adhesion. We will explore humidifying the air for better airflow sensing, contact management, adhesion control, and ultrasound sensing.

VI. Issues to discuss at workshop

1) We believe that mechanical robustness, time to failure, and lifetime cost, rather than sensor accuracy, quality, or even what is measured, will determine what approaches to tactile sensing are actually adopted.

2) We believe non-contact proximity sensing is very useful in manipulation, perhaps more useful than contact or tactile sensing.

3) We believe it is useful to turn tactile sensing into a computer vision problem, and take advantage of the recent progress in computer vision.

4) We believe that using techniques such as change detection, surprise detection, regions of interest, and foci of attention can make processing tens or hundreds of video streams possible. We believe that inexpensive and low power GPUs are available now (for example consider the Raspberry Pi Zero family and the NVIDIA Tegra family).

5) We believe useful skin for hands and feet has high friction, high damping or energy loss, softness (approximately Shore A 30 durometer), some stretchability (less than 50%), and is hairless for greater friction and adherence. Skin for the rest of the body is similar but can be lower friction, and use hairs to detect obstacles at a distance (and parasites). This is similar to the distinction between glabrous and non-glabrous skin in biology.

6) We believe flat cameras will reduce the thickness of skin needed to implement optical sensing by reducing the length of the optical path. Utilizing many small short focal length lenses will also decrease the minimum focal distance, shortening the optical path, while maintaining a large depth of field.

7) We believe better sensors are more important than advances in AI and learning to achieve useful robots. We believe that the breakthrough in self-driving cars is due to sensors such as LIDAR and GPS and aggressive mapping, and not advances in AI or learning. Thermal imaging makes detecting and tracking humans, and human activity recognition, for example, much easier. Different aspects of contact (for example low vs. high frequency or normal vs. tangential force) should be measured by different types of sensors, rather than trying to do everything with a single type of tactile sensor.

8) We believe tactile sensing for manipulation should be evaluated by the quality of the resulting manipulation, and not the accuracy of measuring forces, skin deformation, contact or object shape, or incipient or actual slip. It is not clear which features matter for successful manipulation, but it is unlikely that accurate force sensing or high resolution shape measurements are necessary. Consider how biological disparity detection works. Neurons are tuned to near, far, and zero (at fixation) binocular disparity, rather than accurately representing the full range of disparities [5]. The near and far disparities are represented at a low resolution. Consider how camera autofocus often works by measuring and maximizing contrast [25]. There is no accurate measurement of focus. Consider how orientation to sound or a light source can work in simple robots by simply estimating whether the left or right sensor is louder or brighter and turning in that direction. Crude thermal imaging makes human detection and tracking easy. Evaluating tactile sensors based on how well they measure idealized engineering features is unlikely to optimize manipulation quality or system cost and robustness.

References

[1] B. Ali, M.A. Ayub, and H. Yussof. Characteristics of a new optical tactile sensor for interactive robot fingers. *Intl. J. Soc. Robot*, 4:85–91, 2012.

[2] M. Salman Asif, Ali Ayremlou, Aswin C. Sankaranarayanan, Ashok Veeraraghavan, and Richard G. Baraniuk. Flatcam: Thin, bare-sensor cameras using coded aperture and computation. *CoRR*, abs/1509.00116, 2015. URL http://arxiv.org/abs/1509.00116.

[3] C. G. Atkeson. Build baymax web page. build-baymax.org, 2017.

[4] M. Blank and J.F. Antaki. Low-cost quantitative palpation device for breast lesion tracking and telehealth. In *IEEE Health Innovations and Point-of-care Technologies Conference*, pages 280–283, 2014.

[5] Neil D. B. Bruce and John K. Tsotsos. Attention in stereo vision: Implications for computational models of attention. In Marc Pomplun and Junichi Suzuki, editors,

Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts, pages 65–89, 2013.

[6] Reza Fatemi, Behrooz Abiri, and Ali Hajimiri. An 8x8 heterodyne lens-less OPA camera. In *Conference on Lasers and Electro-Optics (CLEO): QELS Fundamental Science*, 2017. URL http://chic.caltech.edu/wp-content/uploads/2017/03/Cleo_2017_2D_OPA_V7.pdf.

[7] FLIR. The power of micro thermal imaging. www.flir.com/cores/lepton, 2016.

[8] Google. Project Soli. https://atap.google.com/soli/, 2016.

[9] D. Hughes and N. Correll. Texture recognition and localization in amorphous robotic skin. *Bioinspiration & Biomimetics*, 10(5):1–20, 2015.

[10] N. Igo and K. Hoshino. Small optical tactile sensor for robots. In *IEEE/SICE International Symposium on System Integration (SII)*, pages 746–751, 2012.

[11] Y. Ito, Y. Kim, and G. Obinata. Contact region estimation based on a vision-based tactile sensor using a deformable touchpad. *Sensors*, 14:5805–5822, 2014.

[12] A.R. Johnston. Optical proximity sensors for manipulators, 1973. NASA Technical Memorandum 33-612.

[13] E. Knoop and J. Rossiter. Dual-mode compliant optical tactile sensor. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1006–1011, 2013.

[14] N.F. Lepora and B. Ward-Cherrier. Superresolution with an optical tactile sensor. In *IEEE Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 2686–2691, 2015.

[15] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson. Localization and manipulation of small parts using GelSight tactile sensing. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3988–3993, 2014.

[16] M. Ohka, Y. Mitsuya, K. Hattori, and I. Higashioka. Data conversion capability of optical tactile sensor featuring an array of pyramidal projections. In *IEEE Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 1996.

[17] Optoforce. 3D Force Sensor, 2014. www.optoforce.com.

[18] Radhen Patel and Nikolaus Correll. Integrated force and distance sensing using elastomer-embedded commodity proximity sensors. In *Proceedings of Robotics: Science and Systems*, 2016.

[19] PetaPixel. Hitachi's lensless camera uses moire and math, not glass, to take photos. https://petapixel.com/2016/11/16/hitachis-lensless-camera-uses-moire-math-not-glass-take-photos/, 2016.

[20] K. Sato, K. Kamiyama, N. Kawakami, and S. Tachi. Finger-shaped GelForce: sensor for measuring surface traction fields for robotic hand. *IEEE Transactions on Haptics*, 3(1):37–47, 2010.

[21] J. L. Schneiter and T. B. Sheridan. An optical tactile sensor for manipulators. *Robotics & Computer Integrated Manufacturing*, 1(1):65–71, 1984.

[22] H. Shinoda and S. Ando. Ultrasonic emission tactile sensor for contact localization and characterization. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2536–2543, 1994.

[23] TE Connectivity (TE). Board level pressure sensors. http://www.te.com/usa-en/product-CAT-BLPS0013.html, 2017.

[24] D. Um, B. Stankovic, K. Giles, T. Hammond, and V. Lumelsky. A modularized sensitive skin for motion planning in uncertain environments. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 1, pages 7–12, 1998.

[25] Wikipedia. Autofocus. https://en.wikipedia.org/wiki/Autofocus, 2017.

[26] Akihiko Yamaguchi. FingerVision. http://akihikoy.net/p/fv.html, 2017.

[27] W. Yuan, R. Li, M.A. Srinivasan, and E.H. Adelson. Measurement of shear and slip with a GelSight tactile sensor. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 304–311, 2015.