Mine the Gap: Gap Estimation and Contact Detection Information via Adjacent Surface Observation

Yazdan Jamshidi
University of Central Florida
Department of Computer Science
Orlando, FL, USA
yjam@knights.ucf.edu

Greg Welch
University of Central Florida
Department of Computer Science
Orlando, FL, USA
welch@ucf.edu

ABSTRACT
In general, conventional computer vision techniques suffer from an inability to detect hidden surface contacts due to line-of-sight visibility problems. Rather than fitting models to scene objects and estimating inter-object gaps, our approach is to leverage the fact that light passing between and reflecting off the surfaces can offer valuable information as it alters the appearance of nearby surfaces. For a proof of concept demonstration, we employed a machine learning approach to classifying adjacent surface imagery to estimate hidden surface distances and contact locations in a controlled setting under ambient lighting conditions. Our proof-of-concept results demonstrate relatively high accuracy for the estimation of gap size and the detection of contact between hidden surfaces. We envision such measures could someday provide complementary information to be combined with traditional visible-surface methods, to obtain more precise and robust estimates of hidden surface relationships.

CSS Concept
•Computing methodologies→ Scene understanding

Keywords
Contact detection; Gap estimation; Occlusion, Depth sensing; Indirect observation; SVM

1. INTRODUCTION
We define *contact detection* as the process of determination that distance between two surfaces has effectively become zero. In many applications including user interface, safety and security systems it is crucial to detect physical contact between objects. In some applications it is feasible to instrument objects with electrical signals, to directly detect when the object comes in contact with the sensing area. Such interactive displays typically utilize capacitive, resistance, and surface wave sensors for touch sensing. Examples of applications include human-machine interfaces such as touch tablets, smart phones, and even car door lock systems [1]-[7]. Robotics and industrial automation are another application area where contact detection is a fundamental aspect of physical manipulations [8]-[11].

PRAI 2018, August 15–17, 2018, Union, NJ, USA

ACM ISBN 978-1-4503-6482-9/18/08…$15.00
https://doi.org/10.1145/3243250.3243260

However, in some applications it is not possible nor desirable to instrument the surfaces. In such cases contact detection can be thought of as a special case of gap estimation, where gap is defined as the distance between two surfaces of interest. One can continually sense and estimate the size of the gap between surfaces, e.g., using RGB or RGB-D (depth) sensors, and declare contact when the gap falls below a threshold. An accurate and robust method for the real time characterization of un-instrumented surface relationships could someday enable a range of applications such as human-computer interaction in future workspaces, human-machine safety in industrial settings (e.g contact avoidance between moving robotic components or manufactured objects), and the detection of hidden breaches of sterile fields in medical procedures. In particular, undetected contact with a contaminated object during a sterile medical procedure can introduce a healthcare-associated infection (HAI).

From consumer-level human-computer interaction to medical care, there exist a large number of circumstances where make robust, reliable, precise, and accurate assessments of the distances between the objects impractical or impossible.

Gap estimation is difficult when the gap is not directly observable. Such conditions arise for example when there is no direct view of the object surfaces or the gap, or the views are occluded by objects in the scene, including the objects of interest themselves (self-occlusion). In situations where the relationship between the sensors and the gap of interest remains relatively constant, the sensors can likely be positioned to directly observe the gaps formed between surfaces. In such circumstances, an instantaneous discontinuity in color or depth in a gap region could indicate a new contact between the surfaces.

Here we present the idea and some preliminary proof-of-concept results for a novel vision-based approach for relatively direct (non-inference based) measures of inter-surface distances and contact for surfaces that are hidden (e.g., occluded) with respect to conventional line-of-sight optical sensing. As opposed to directly observing or sensing the dynamic objects themselves, our idea is to look for evidence of the hidden surface relationships by observing the environmental effects of signals (e.g., ambient light) that propagate between and reflect off the hidden surfaces, spilling into observable areas of the environment. In some sense this is similar to the *transit approach* astronomers use to find planets because distant planets are not directly visible they look for evidence of the planet in the illumination patterns of nearby stars. We remark that recently Bouman et. al[12] leveraged the subtle spatio-temporal radiance variations that arise on the ground at the base of a wall's edge to construct a one-dimensional video of the hidden scene behind the wall. More specifically they could identify the number and location of people in a hidden scene. They assume the observation plane(ground) is planar and Lambertian, the visible and hidden scene are modelled as light

emitted from a large celestial sphere, and the people in the hidden scene are modelled as cylinder.

In our proof-of-concept experiments we use machine learning (SVM) to classify the visible environmental effects of these observable emerging light signals as a function of the physical gap size, one could estimate the gap size from such observations. Here we present the basic idea, specific methods, and results from controlled experiments. In the future we expect adjacent surface observations will offer complementary information that can be combined with other conventional methods to obtain more precise contact detection and gap estimation, in particular under conditions where gaps are not directly observable.

The remainder of this article is organized as follows. We present an overview of the model representation in Section II. Our experimental results are described in Section III. Finally, we present some conclusions and future work in Section IV.

2. EXPERIMENTAL METHOD

The purpose of this section is to demonstrate the effect of presence of and object on the adjacent surfaces. We use a regular camera to capture images from the scene, however, depth or touch information can not be obtained using a single camera. The key idea is that rather than relying on direct observations of an object, surface, or gap, we observe the visible surfaces *adjacent* to the object. In particular, as one object approaches another object, and the hidden gap shrinks or grows, the light scatter on the adjacent surfaces also changes. This change in the adjacent surface lighting can provide additional information about the geometric relationship between the two objects. By visualizing the information in the surroundings of the surfaces adjacent to the object of interest we illustrate how presence of the object in the scene can affect its surroundings. Figure 1 presents a conceptual illustration of the idea.

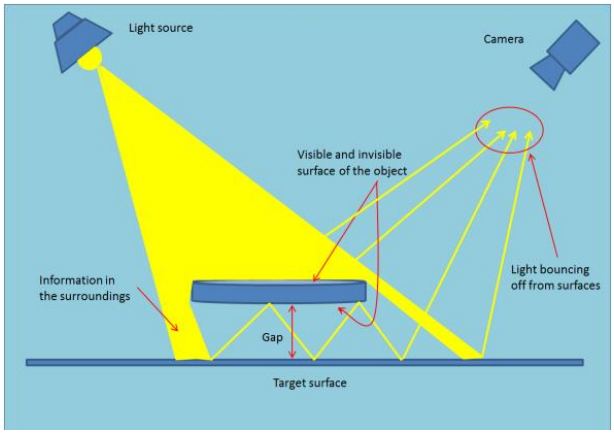


Figure 1. Seeing around the object.

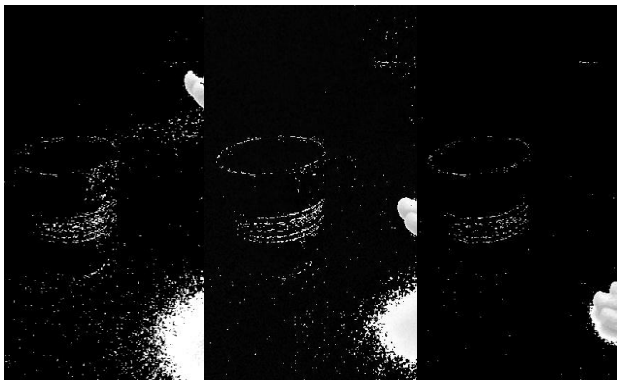
2.1 Light Scatter Visualization

To illustrate the effects of gap-induced illumination on the adjacent surfaces and provide some intuition for how measurement of the effect can contribute to the sensing of the gap, we set up a simple experiment. We illuminated a simple environment with a regular incandescent light bulb placed above a table and imaged the scene with a flat hand positioned at three different heights. As shown in Figure 2, light scattering off the *underside* of the hand is visible on the table and glass surfaces. Note that we subtracted a static background image from each dynamic image, according to the following equation:

$$|I(x, y, h) - I(x, y)|,\ h = 0, 4, 8$$

where $I(x, y, h)$ and $I(x, y)$ represent scene including hand at height h inches and scene without presence of hand, respectively. To offer a better sense of the hand's position, we made sure the finger tips were visible in the images. We remark that for the purpose of a better visualization we applied a threshold to pixel values to remove noise, but we did not used thresholding to train the model. As shown in the images, as the gap shrinks the adjacent scattered light changes accordingly.

(a)



(b) (c) (d)

Figure 2. (a) the environment without presence of hand. (b, c and d) The palm of the hand is placed at the height of 6, 3, and 0 inches.

2.2 Contact Detection and Gap Estimation Classification

To exploit the adjacent light scatter phenomena, we form vision-based measurements of the surrounding scene and train a support vector machine (SVM) using manually labelled images, to classify the conditions as corresponding to different gap sizes or contact conditions. SVM performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors that define the hyperplane are called support vectors. In other words, given labelled d-dimensional training data $\{x_1, x_2, \dots, x_n\}$ and $y_i \in \{-1,1\}$ be the class label of x_i, The decision boundary should classify all points correctly s.t $y_i(w^T x_i + b) \geq 1, \forall\, i \leq n$, and it can be found by solving the following constrained optimization problem:

$$Minimize\ \frac{1}{2}w^2$$

overall $w \in R^d$ and $b \in R$ subject to $y_i(w^T x_i + b) \geq 1, \forall\, i \leq n$.

The data preprocessing step consists of converting each image to grayscale, background subtraction, eliminating the fixed rectilinear region that contains the hand, then down sampling and

normalizing the imagery. We do not extract any sort of features from the images. We use the preprocessed images as direct input data to the SVM.

3. PROOF-OF-CONCEPT EXPERIMENTS

In this section we present quantitative results from three different experiments including (a) hidden contact detection, (b) estimation of the hidden gap between an object and a table, and (c) and hidden contact detection and localization. The experimental set up consists of a single ordinary camera placed next to a table, viewing a surface from a height of one meter above the table, and a regular incandescent light bulb placed near the table. To avoid utilizing the information about the object itself, in every image we mask out (remove) a fixed-dimension rectilinear region that would otherwise include the object. The result is a set of masked images that appear virtually identical—they contain the scattered adjacent lighting, but it is imperceptible. We use machine learning to extract the imperceptible adjacent light patterns from these masked images, and to classify the gap size and contact state of the hidden surfaces.

3.1 Contact Detection

The goal of this experiment is to develop a model to classify contact and none contact conditions. For this purpose, we set up a binary classification problem. We place an object at the height of 6 inches from the table surface and while the object is moving up, down, left and right with different orientation we collect 60 images by capturing one per second. The object is continuously approaching the surface until the distance between the object and surface reaches 2 inches. These are labelled as non-contact situations. For the contact conditions, while the object is moving around at the distance of less than 2 inches including touching the surface, 30 images are captured. The procedure is repeated for different hand shapes including, palm, fist, vertical palm, and vertical fist. As a result, the data set contains 360 images which 240 images represent none contact and the remaining denote contact conditions. Figure 3 shows the fist and palm of the hand for both contact and none contact conditions. The translucent white region on the right in each image is intended to show the fixed-dimension region that was removed from the camera image before processing by the SVM, while also showing what *was* in that region before removal. In other word we *hid* the hand by completely omitting the right half of the training and test images.

After training the model, 10-fold cross validation and f1-score were used to evaluate the model performance. The results are shown in table I where Label 0 and 1 represent contact and non-contact events, respectively. Table I reveals that although the objects were completely removed from the images and different hand shapes were used in the experiment, the model could effectively classify contact versus non-contact conditions with the accuracy of almost 98 percent for each class. The obtained result is compelling and the accuracy as high as 98 percent demonstrates that our method provides precise information for the purpose of contact detection.

Table I. The Mean and Standard Deviation(STD) of the classification accuracy for contact(0) and non-contact(1) situations.

Contact/Non-Contact	Mean	STD
0	97.99	0.02
1	98.93	0.01

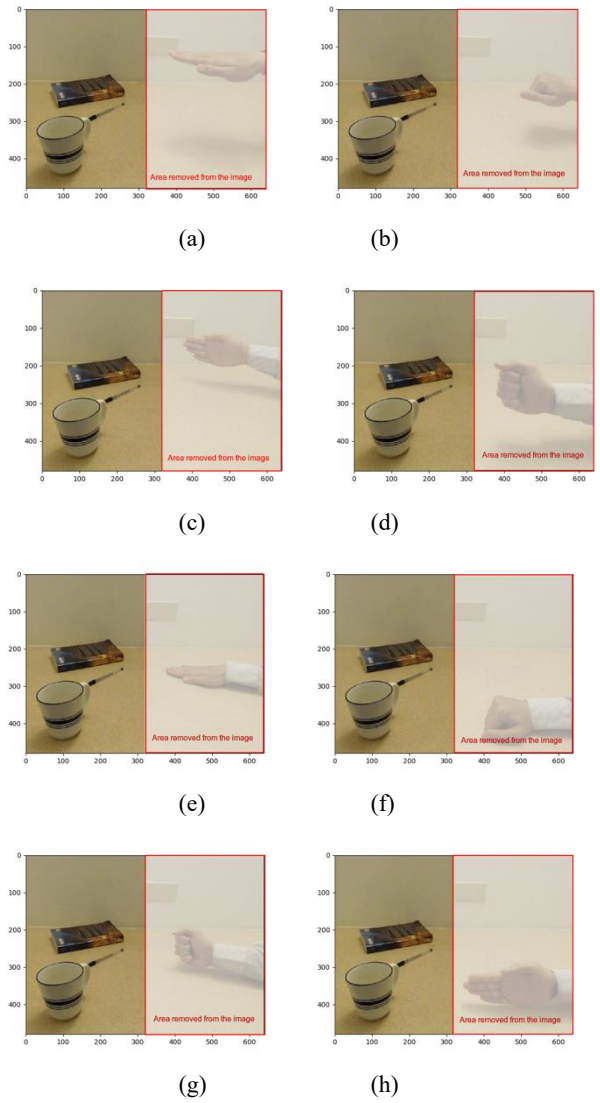


Figure3: (a), (b), (c), and (d) represent non-contact situations, and (e), (f), (g), and (h) indicate contact situations for palm of the hand, fist, vertical palm, and vertical fist, respectively. The translucent white area indicates the portion removed from the images given to the SVM.

3.2 Gap Estimation

In this experiment we measure the distance between two surfaces of interest. We place an object at the height of 6 inches from the surface and capture 30 images while the object is moving to the left and right with different orientations. We repeat this procedure with two-inch reductions in the gap until the object reaches the surface. We also repeat this for different objects, including four different hand shapes. Our final data set contains 480 images, each labelled with the object distance to the surface.

The results are shown in Table II. It turns out that the changes in the light scattering on the table surface accounts for the shrinking gap. This means that the light is dispersed on the adjacent surfaces to the object as it hits the object which causes subtle variations that can not be detected with the human eyes. However, SVM could successfully extract the subtle patterns in the scattered light and estimate the gap accordingly with the accuracy ranging from

95 to 98 percent. This high accuracy indicates our method performs well on the task of gap estimation.

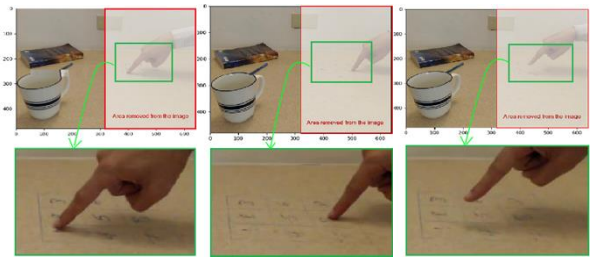
Table II. The Mean and Standard Deviation(STD) of the classification accuracy at different gap sizes. Labels 0, 1, 2, 3 indicates gap sizes 0, 2, 4, and 6 inches, respectively.

Labels	0	1	2	3
Mean	96.93	98.32	96.96	95.18
STD	0.02	0.01	0.02	0.02

3.3 Contact Detection and Localization

In this experiment we detect and localize contact. For this purpose, we drew a 6 × 6-inch square on the table surface and divide the square into nine equally-sized sub-squares containing the numbers 1 through 9. For each sub square while the pointing finger is touching the surface and moving around, we capture 30 images. We repeat the same procedure for all the sub squares and label the images accordingly. Moreover, 60 images are captured as the fingertip is moving above the entire square at the height of 1 inch. As a consequence, we have a multi-class classification. The results shown in Table III demonstrate the effectiveness of our method in terms of both contact detection and localization. The experiment was performed on a laptop equipped with Intel 2.66 GHz Core 2 Duo CPU and 4G of RAM, and took only 0.014 seconds to classify an image, lending support for real-time on-line applications. Our method achieved a low standard deviation, and a high classification accuracy of roughly 98 percent for almost all classes.

problem where labels 1–9 represent the contact zones and label 10 indicates no-contact situations. Figure 4 shows two contact and one no-contact situations.



(a) (b) (c)

Figure 4: (a), (b) represent fingertip touching sub squares number 1 and 8, respectively. (c) denotes a non-contact situation. The translucent white area indicates the portion removed from the images given to the SVM.

TABLE III: The mean and standard deviation (STD) classification accuracy for different contact zones (labels 1-9) and no-contact conditions (label 10).

labels	1	2	3	4	5	6	7	8	9	10
Mean	98.67	98.95	98.18	98.46	98.67	98.67	92.73	98.82	87.62	99.20
STD	0.03	0.02	0.03	0.03	0.03	0.02	0.1	0.02	0.1	0.01

4. Discussion and Future Work

We have presented proof-of-concept experiments for a novel vision-based method for contact detection and gap estimation. Common approaches such as camera-based computer vision and acoustic ranging are thwarted by line-of-sight issues including partial and full occlusions of the surfaces of interest, often by the objects themselves (self-occlusion). Unlike the existing approaches we do not rely on models for the objects of interest. Instead we observe the surfaces adjacent to the object for evidence of the hidden surface relationships. Our proof-of-concept experiment employed a single commercial off-the-shelf web camera and machine learning methods to detect subtle patterns in the light scattering on the adjacent surfaces. The results demonstrate the potential of our approach, encouraging further investigation and consideration of possible uses in a variety of applications.

One of the primary challenges in our method is that while the light signals could provide useful information about a gap, they will also be affected by other scene geometry and objects—any changes in lightning or other geometry could affect the SNR. In the future we aim to overcome the limitations and further explore solutions for conditions where the camera view and light source are not static. For example, a similar approach to [13] can be used to make the method robust in dynamic scenes. They used a standard 2D camera, and a laser pointer to detect motion and track a moving object hidden around a corner or behind a wall even in unknown rooms. Indeed, to obtain a measured image containing only light from the laser, they took the difference of images captured with and without laser illumination. Additionally, they subtracted a measurement image containing light reflected by the background that was smooth and well approximated by a linear function.

Furthermore, we aim to identify methods for possibly increasing the signal-to-noise ratio (SNR) and create additive or destructive patterns by combining two or more sources of propagating signals including time division, color/spectral multiplexing, and pseudo-random spread spectrum approaches.

In addition, the other approach we will take is to train the system with massive amount of data to facilitate more advanced methods (e.g., deep learning). To generate desired amount of data we will leverage the precise continuous measurement systems including magnetic or optical sensors on the objects with models of the objects and configures the space with precision imagery of the gaps. Similar to the light signals, it is possible the spectral properties of audio passing through hidden gaps will depend on the surface and other nearby materials. Therefore, as an alternative to light signals it is promising to leverage pseudo-random signals with relatively wide band spectral characteristics [14] and to learn the spectral profiles that correspond to various gap sizes, in effect measuring the dynamic acoustic impulse response function associated with the surfaces. However, in this case, another challenge arises from poor source and sensor choices and/or positions that should be taken into consideration.

In general, we consider our approach as a complement to visible surface approaches to provide more precise and robust estimates of hidden geometric relationships

5. REFERENCES

[1] Chan, R. and Wang, F. and You. P. A. 2010. Survey on the development of multi- touch technology. *Asia-Pacific Conference on Wearable Computing Systems.*

[2] Bau, O. Poupyrev, I. Israr, A. and Harrison. C. 2010. Teslatouch: Electrovibration for touch surfaces. *UIST'10, New York, USA.*

[3] Kim, H.-R. and Choi, Y.-K. and Byun, S.-H. 2010. A mobile-display-driver IC embedding a capacitive-touch-

screen controller system. *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*

[4] Schoenleben, O. and Oulasvirta, A. 2000. Sandwich keyboard: fast ten-finger typing on a mobile device with adaptive touch sensing on the back side. *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 175–178.

[5] Kim, H.-K. and Lee, S. and Yun, K.-S. 2011. Capacitive tactile sensor array for touch screen application. *Applied Artificial Intelligence*, 165:2–7.

[6] Schoning, J. et al. 2008. Multi-touch surfaces: a technical guide. *technical Report TUM-I0833*, 2008.

[7] Bhalla, M and Bhalla, A. 2010. Comparative study of various touchscreen technologies. *International Journal of Computer Applications* (0975 – 8887) Volume 6– No.8, September 2010.

[8] Backus, S. B. and Dollar, A. M. 2014. Robust resonant frequency-based contact detection with applications in robotic reaching and grasping. *IEEE/ASME Transactions on Mechatronics*, 19, 2014.

[9] Bartolozzi, C. Natale, L. Nori, F. and Metta, G. 2016. Robots with a sense of touch. *Nature Materials*, pages 921–925.

[10] Dahiya, R. S. and Valle. M. 2013. Robotic tactile sensing: Technologies and system. *Springer-Verlag, New York*.

[11] Mittendorfer, P. and Cheng, G. 2011. Humanoid multimodal tactile-sensing modules. *IEEE Transactions on Robotics, 27.*

[12] Bouman, K. L. and Ye, V. and Yedidia, A. B. and Durand, F. and Wornell, G. W. and Torralba, A. and Freeman, W. T. 2017. Turning corners into cameras: Principles and methods. *ICCV*.

[13] lei n, . and e ters, . and art n, . and Laurenzis, M. and Hullin, M. B. 2016. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports 6*.

[14] Vallidis. N. 2002. Whisper: A spread spectrum approach to occlusion in acoustic tracking. ph.d., University of North Carolina at Chapel Hill.