Enhance Vision-based Tactile Sensors via Dynamic Illumination and Image Fusion

Artemii Redkin
LASR Lab
TU Dresden
Dresden, Germany
aredkin@lasr.org

Zdravko Dugonjic
LASR Lab
TU Dresden
Dresden, Germany
zdugonjic@lasr.org

Mike Lambeta
Meta AI
Menlo Park
CA, USA
lambetam@meta.com

Roberto Calandra
LASR Lab
TU Dresden
Dresden, Germany
rcalandra@lasr.org

Abstract

Vision-based tactile sensors use structured light to measure deformation in their elastomeric interface. Until now, vision-based tactile sensors such as DIGIT and GelSight have been using a single, static pattern of structured light tuned to the specific form factor of the sensor. In this work, we investigate the effectiveness of dynamic illumination patterns, in conjunction with image fusion techniques, to improve the quality of sensing of vision-based tactile sensors. Specifically, we propose to capture multiple measurements, each with a different illumination pattern, and then fuse them together to obtain a single, higher-quality measurement. Experimental results demonstrate that this type of dynamic illumination yields significant improvements in image contrast, sharpness, and background difference. This discovery opens the possibility of retroactively improving the sensing quality of existing vision-based tactile sensors with a simple software update, and for new hardware designs capable of fully exploiting dynamic illumination.

1 INTRODUCTION

In the realm of robotics, haptic exploration is a key to understanding the world through touch interactions. Tactile sensors empower robots to gather essential information about their surroundings, manipulate objects precisely, and ensure safe interactions within dynamic environments (1). By detecting physical contact, tactile sensing allows robots to avoid collisions, adjust movements, and handle objects delicately, especially in tasks requiring precision and interaction (2; 3).

Vision-based Tactile Sensors (VBTS) are a popular choice of tactile sensors (2). They enable robots to perceive their environment by capturing surface deformations upon contact with objects, thus facilitating the measurement of forces, textures, and shapes. VBTS typically incorporate LED lights in their construction, and currently, all such sensors use static illumination, meaning the lighting intensity and colors remain constant during measurements. Enhancing images from VBTS holds pivotal importance due to their widespread applicability across diverse robotic tasks. These sensors serve as crucial components in robotic systems, providing essential data for various operations. The state-of-the-art approach involves training deep neural networks using images from VBTS, where the quality of the input image significantly influences the model's performance and output. Improved imaging quality from VBTS could offer deeper insights into robotic interactions with objects,

NeurIPS 2024 Workshop on Touch Processing: a new Sensing Modality for AI .

ultimately enhancing problem-solving capabilities. Addressing this need, our study aims to explore the feasibility of image enhancement in VBTS and propose methodologies for achieving this enhancement.

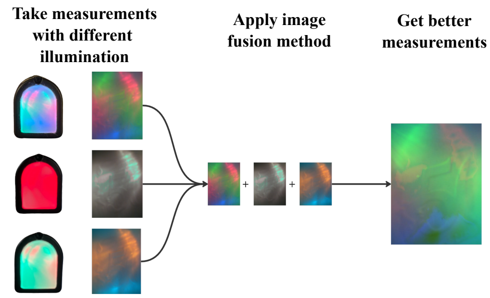

Figure 1: Current vision-based tactile sensors use static illumination patterns. In this work, we instead propose to collect several measurements under dynamic illumination conditions, and then fuse them together. Experimental results show that this approach yield significantly improved measurements.

In this study, we contribute to the field by establishing a framework for enhancing the measurement quality of vision-based tactile sensors through the application of dynamic lighting and image fusion techniques. Our investigation delves into the comprehensive evaluation and demonstration of diverse approaches tailored to enhance image quality. Specifically, our methodology integrates dynamic lighting schemes to enhance contrast and sharpness, while employing image fusion algorithms to combine multiple sensor outputs into cohesive images. We further validate the feasibility of enhancing sensor images and conduct a comparative analysis of various illumination variations and image fusion methods, assessing their applicability to vision-based tactile sensors. Through rigorous experimentation and analysis, we present a spectrum of effective techniques poised to enhance images acquired from VBTS.

2 RELATED WORK

Previous research in the field of vision-based tactile sensing focus on strategic design and positioning of lighting system, so that illuminated elastomer gives the optimal response for downstream tasks. It has been noted in (4) that more light sources improve tactile readings as it allows better distribution of light over the elastomer surface. In (5) they evaluate how three different illumination setups affect the performance of contact state estimation problem. Another work (6) compare how both positioning and combinations of monochrome red, green and blue lights impact the results of 3D reconstruction task. Unlike prior research, our study systematically evaluates the effect of combining images captured under dynamic illumination setups compared to the static ones.

While the aforementioned work focus on static lighting configurations, a broader field in computer vision demonstrates the advantages of active lighting. Based on the idea of photometric sampling (7), the authors in (8) proposed the method for obtaining object reflectance and surface normal by illuminating the scene with high-frequency pulsing LED light sources, placed around the object. Another work by (9) shows that it is possible to create a depth edge map by flashing the scene with lights positioned around the camera lens. More recently, one notable example of dynamic lighting is the quantitative differential phase contrast imaging technique introduced by Tian and Waller (10). This method utilizes different lighting conditions in an LED array microscope to enhance phase contrast, improving the visualization of transparent samples in biological research. Unlike traditional applications focused on visual imaging, microscopy, or medical diagnostics, applying these techniques to tactile sensing introduces innovative strategies for capturing and interpreting tactile information.

In the domain of image fusion, Wang and Chang proposed the Laplacian Pyramid method (11), addressing multi-focus image fusion by decomposing images into multiple levels and selectively incorporating focused elements from each level into the final image. This technique preserves the best-focused aspects of each original image, beneficial in fields where detailed texture is essential. Further advancements in image fusion include the discrete fractional wavelet transform method introduced by Xu, Wang, and Chen (12). This approach allows for the integration of multiple medical images into a single composite, retaining critical information from each source image for improved medical diagnosis and treatment planning. However, previous research has not explored enhancing the quality of images in the context of vision-based tactile sensors.

3 DYNAMIC ILLUMINATION FOR VISION-BASED TACTILE SENSORS

The objective of image fusion is to combine two or more images into a single output that enhances overall image quality, making the selection of an effective fusion method essential. With dynamic

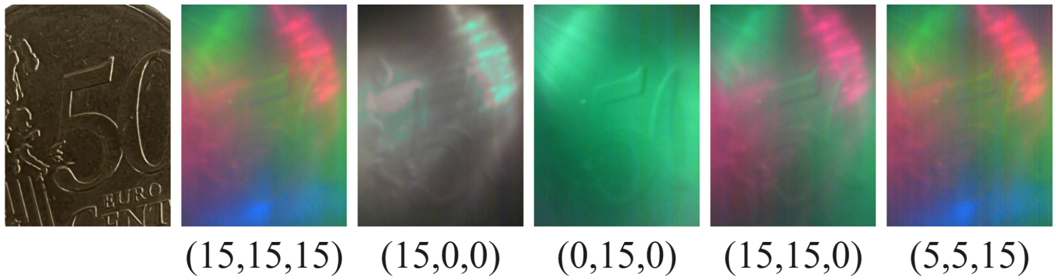

Figure 3: Images of the coin surface obtained with DIGIT with different illumination settings.

lighting approach, image fusion is closely linked with determining the optimal illumination patterns for the touch sensor. Both identification of optimal illumination patterns of the tactile sensor, number of images required and selection of the most effective image fusion fusion method are crucial in dynamic lighting. This problem can be considered as the following optimization task

$$\underset{\Theta,n,f}{\text{argmax}}\, \mathcal{P}\left(f\left(I_{\theta_1}, I_{\theta_2}, \ldots, I_{\theta_n}\right)\right)\,,$$

where $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ represents the set of illumination patterns in which the images I_{θ_i} were taken, n is the image budget - the number of images to be used for fusion, $f : (I_1, I_2, \ldots, I_n) \mapsto I^*$ is the image fusion method, $\mathcal{P} : I^* \mapsto \mathbb{R}$ is some image quality metric applied to the resulting fused image, and I_{Θ_i} is the image taken in illumination determined by θ_i.

In our experiments we use the following metrics for image quality evaluation: 1) Gradient-based Sharpness is defined as $S = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$, where I represents the image, N - the number of pixels of the image, $\frac{\partial I}{\partial x_i}$ and $\frac{\partial I}{\partial y_i}$ are the partial derivatives of the image intensity with respect to the spatial coordinates x_i and y_i. 2) Root Mean Squared Contrast is defined as $C_{\text{rms}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - \mu)^2}$, where I_i represents the intensity of the i-th pixel, μ is the mean intensity of all pixels, N is the total number of pixels in the image. 3) Difference with Background is defined as $D = \frac{1}{N}\sum_{i=1}^{N}|I_i - B_i|$, where I is the image of the elastomer's surface in contact with an object and B is the background image (i.e., the image obtained from the sensor without touching any object).

4 EXPERIMENTAL RESULTS

With our experiments, we aim to answer the question: Can we enhance the quality of the measurements of DIGIT sensor using dynamic lighting and image fusion techniques?

For our experiments, we employed a standard DIGIT (2) vision-based tactile sensor, which is equipped with three LED lights: red, green, and blue. The intensity of each light can be adjusted from 0 to 15, where 0 means no light and 15 - maximum intensity, enabling the creation of various illumination patterns represented by tuples (R, G, B), where R, G, and B denote the intensity values of the red, green, and blue LEDs, respectively. By default, the DIGIT sensor is set to (15, 15, 15), with all LEDs operating at maximum intensity.

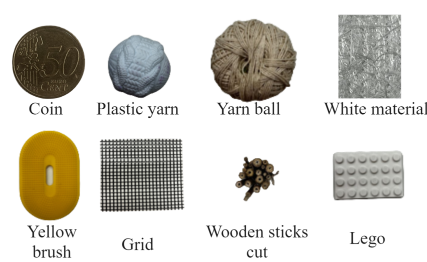

Figure 2: Objects used in the experiments.

4.1 Enhancing Image Quality

We ask: is it feasible to enhance image quality from the DIGIT sensor using dynamic lighting and image fusion techniques? To assess this, we captured images of various objects under different illuminations. Initially, we employed the channelwise summation method, alternating between

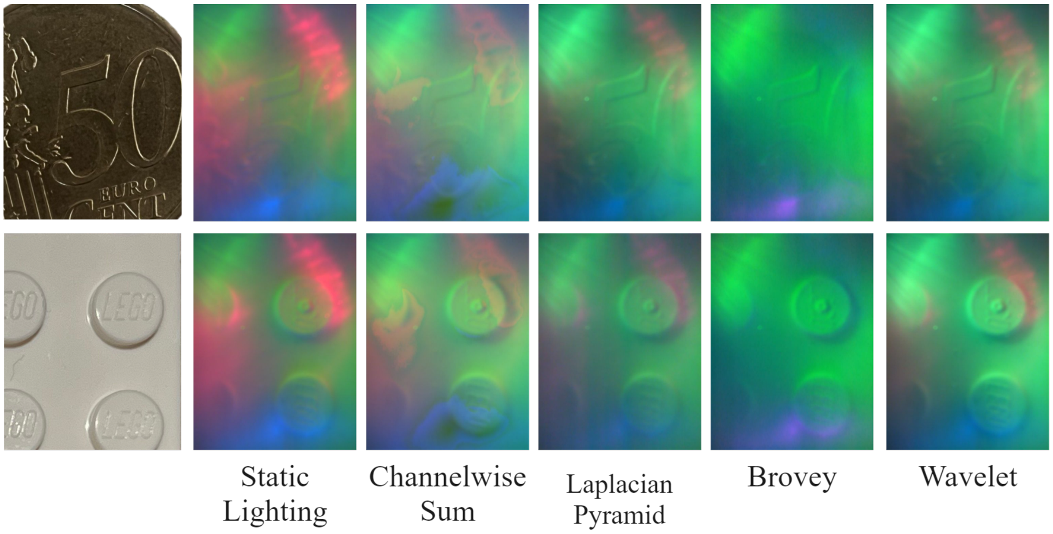

Figure 5: Images of the coin and Lego brick obtained with static and dynamic lighting using different image fusion methods.

illuminating the objects solely with red, green, and blue light intensities of (15, 0, 0), (0, 15, 0), and (0, 0, 15), respectively.

Subsequently, we expanded our measurements to include additional illumination settings such as (15, 15, 0), (0, 15, 15), (15, 10, 5), and so forth, to leverage the Laplacian pyramid image fusion method. This method was then applied to sets of 2 and 3 images. To demonstrate the effectiveness of the Laplacian pyramid method, we combined images taken with intensity settings (15, 15, 0) and (0, 15, 15).

Upon analysis, we found that the Laplacian pyramid method consistently improved image quality for all objects in terms of background difference, sharpness, and contrast compared to images taken with standard DIGIT illumination settings. Meanwhile, the channelwise sum method improved background difference and sharpness for all objects, as well as sharpness for most objects, with the exception of Lego bricks and wooden sticks. Then we extended the range of applied methods and used Brovey image fusion and Discrete Wavelet Image Transform Image Fusion, which showed high efficiency.

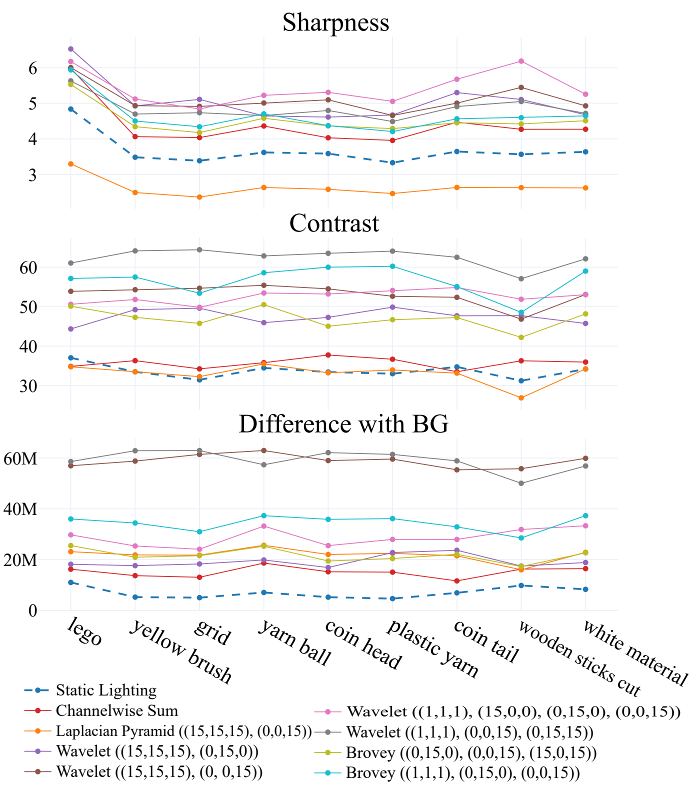

Figure 4: Metrics for all of the objects. Dynamic lighting with selected image fusion methods increased all of the metrics for all objects simultaneously. Wavelet methods provide the highest increase in image metrics.

Our findings reveal that applying dynamic illumination and image fusion techniques to images obtained from the sensor enhances image quality, improving contrast, sharpness, background difference, and human perception simultaneously. Particularly noteworthy is the discovery that the fusion method yielding the highest performance was the Wavelet Transform Image Fusion when applied to images acquired with (15,15,15) and (0,15,0) illumination intensity settings.

We conclude that the use of dynamic lighting and image fusion techniques is a promising direction to enhance image quality of vision-based tactile sensor.

References

[1] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. Adelson, and S. Levine, "The feeling of success: Does touch sensing help predict grasp outcomes?" in *Proceedings of (CoRL) Conference on Robot Learning*, November 2017, pp. 314 – 323.

[2] M. Lambeta and et al., "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[3] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik, "General in-hand object rotation with vision and touch," in *Conference on Robot Learning*. PMLR, 2023, pp. 2549–2564.

[4] W. K. Do, B. Jurewicz, and M. Kennedy, "Densetact 2.0: Optical tactile sensor for shape and force reconstruction," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 549–12 555.

[5] O. Azulay, N. Curtis, R. Sokolovsky, G. Levitski, D. Slomovik, G. Lilling, and A. Sintov, "Allsight: A low-cost and high-resolution round tactile sensor with zero-shot learning capability," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 483–490, 2023.

[6] S. Wang, Y. She, B. Romero, and E. Adelson, "Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6468–6475.

[7] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Determining shape and reflectance of lambertian, specular, and hybrid surfaces using extended sources," in *International Workshop on Industrial Applications of Machine Intelligence and Vision,*. IEEE, 1989, pp. 169–175.

[8] A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec, "Performance relighting and reflectance transformation with time-multiplexed illumination," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 756–764, 2005.

[9] R. Raskar, K.-H. Tan, R. Feris, J. Yu, and M. Turk, "Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging," *ACM transactions on graphics (TOG)*, vol. 23, no. 3, pp. 679–688, 2004.

[10] L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an led array microscope," *Optics Express*, vol. 23, no. 9, pp. 11 394–11 403, 2015.

[11] Y. Wang and L. Chang, "Laplacian pyramid multi-focus image fusion," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 3, pp. 268–280, 2011.

[12] X. Xu, Y. Wang, and S. Chen, "Medical image fusion using discrete fractional wavelet transform," *Biomedical signal processing and control*, vol. 27, pp. 103–111, 2016.