June 2022

Depth Estimation Using a Single Near-Infrared Camera and Dot Illuminator

Kuntal Sengupta

Hart Levy

Adarsh Kowdle

Andrey Zhmoginov

Depth Estimation Using a Single Near-Infrared Camera and Dot Illuminator

ABSTRACT

This disclosure describes techniques to estimate the depth map of a scene by illuminating the scene with a near-infrared (NIR) dot projector and by using a single NIR camera that captures the resulting dot-matrix image. The use of a single camera and a single dot projector to estimate the depth map of a scene can significantly reduce the complexity, bill of materials, and calibration procedures for various face-based authentication applications.

KEYWORDS

- Near-Infrared (NIR) imaging
- Dot projector
- Dot illumination
- Depth map
- Presentation attack
- Face-spoofing attack
- Face recognition
- Time-of-flight (ToF) depth estimation
- Structured-light depth estimation
- Stereo reconstruction
- Deep learning
- Autoencoder
- Generative adversarial network (GAN)
- Face-based authentication

BACKGROUND

Face-based recognition technology has been implemented to enable users to unlock smartphones and other devices by merely looking at the device. Presentation attacks attempt to spoof face-based recognition technology using a photo of the true user. The attacks are forestalled by the use of a camera that can obtain a depth map.

A depth map can be computed by solving for the stereo match between NIR stereo images taken by two cameras of a scene illuminated by a dot projector. While stereo matching under active illumination is a well-studied problem, a key challenge is obtaining accurate calibration parameters for the cameras at any instant of time. When the cameras lose calibration, as they do periodically, the depth calculation becomes erroneous. Depth-quality metrics to trigger auto-calibration partially addresses calibration problems, but not in all situations.

DESCRIPTION

This disclosure describes techniques to estimate the depth map of a scene by illuminating the scene with a near-infrared (NIR) dot projector (also known as dot emitter) and by using a single NIR camera that captures the resulting image, known as the dot image. The use of a single camera and a single dot projector to estimate the depth map of a scene significantly reduces the complexity, the bill of materials, and the calibration procedures for various face-based authentication applications, e.g., smartphone unlock, laptop unlock, door unlock, driver monitoring, automotive entertainment, etc.

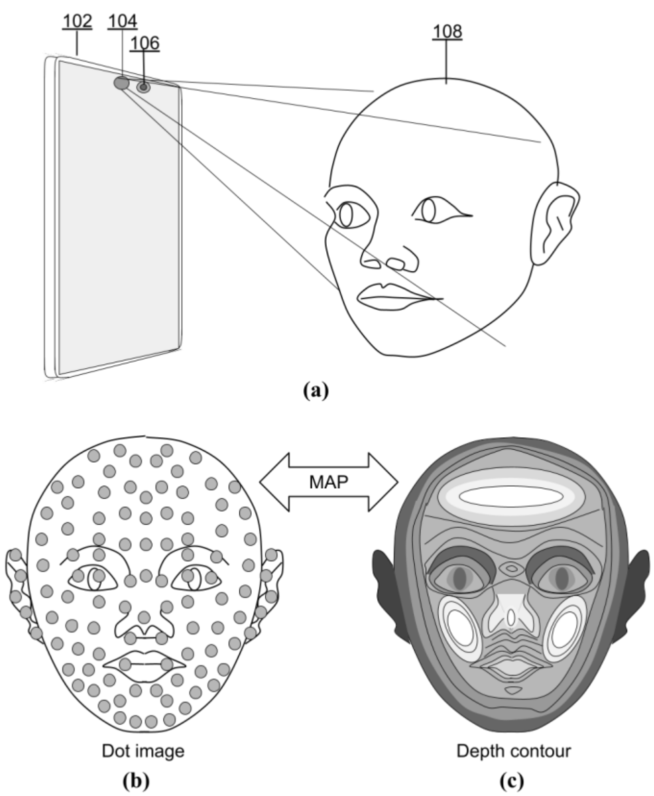

Fig. 1: Finding depth within a scene using a single NIR camera and dot illuminator

Fig. 1 illustrates finding depth within a scene using a single NIR camera and dot illuminator. A dot illuminator (104) mounted on a device (102) such as a smartphone emits a uniform dot matrix of invisible near-infrared light to strike a subject (108), e.g., a user's face. An NIR camera (106) captures the resulting dot image (Fig. 1b). The three-dimensional curves of the subject reflect the NIR dots such that they lose the initial uniformity of their spatial distribution.

The particular, non-uniform spatial dot distribution as captured in the dot image is a nearly unique representation of the subject. The dot image can be used to recognize the subject, e.g., for the purposes of authentication, by training a machine learning model to map the dot image to a depth contour plot (Fig. 1c).

During training, the machine learning model can be supplied with depth contour plots obtained using conventional depth-finding techniques, e.g., time-of-flight (ToF) depth estimation, structured-light depth estimation, stereo reconstruction, etc. During operation (inference), a single camera suffices to map the dot image to a depth contour plot. Calibration of the transmit and receive dot matrices, e.g., between the dot illuminator and the camera, is avoided, effectively being absorbed into the training procedure.

The techniques can be made robust to calibration drift by using training data from several devices of the same family. Examples of machine learning models suitable for the mapping task include generative adversarial networks (GAN), autoencoder networks, etc. A presentation (face-spoofing) attack is detected because a flat, printed photograph of the user lacks depth and is unable to produce the uniquely non-uniform dot image characteristic of the user's true three-dimensional face.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's face, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to estimate the depth map of a scene by illuminating the scene with a near-infrared (NIR) dot projector and by using a single NIR camera that captures the resulting dot-matrix image. The use of a single camera and a single dot projector to estimate the depth map of a scene can significantly reduce the complexity, bill of materials, and calibration procedures for various face-based authentication applications.

REFERENCES

[1] Goldman, Daniel, Harris Nover, and Supreeth Achar. "Real-time spacetime stereo using spacetime descriptors." U.S. Patent 11,190,746, issued November 30, 2021.

[2] Mokalla, Suha Reddy. "Deep Learning Based Face Detection and Recognition in MWIR and Visible Bands." Master's thesis, West Virginia University, 2020.

[3] Jonasson, Rebecca, and Anna Kollberg. "Structured light based depth and pose estimation." Master's thesis, Chalmers University of Technology, 2019.

[4] Yoon, Youngjin, Gyeongmin Choe, Namil Kim, Joon-Young Lee, and In So Kweon. "Fine-scale Surface Normal Estimation using a Single NIR Image." *arXiv preprint arXiv:1603.07475* (2016).

[5] Levy, Hart, "Determining Local Depth From Structured Light Using A Regular Dot Grid," Technical Disclosure Commons, (October 01, 2019) https://www.tdcommons.org/dpubs_series/2536

[6] Barry, Colin, Jessica De Souza, Yinan Xuan, Jason Holden, Eric Granholm, and Edward Jay Wang. "At-Home Pupillometry using Smartphone Facial Identification Cameras." In *CHI Conference on Human Factors in Computing Systems*, pp. 1-12. 2022.

[7] Agrawal, Amit, Ramesh Raskar, and Rama Chellappa. "What is the range of surface reconstructions from a gradient field?" In *European conference on computer visio*n, pp. 578-591. Springer, Berlin, Heidelberg, 2006.

[8] Anandkumar, Anima. "Neural operator: A new paradigm for learning PDEs" Video available online at https://youtu.be/Bd4KvlmGbY4