Structured Light 3D Surface Sensing

Comprehensive Exam I Presentation

Ian J. Maquignaz

Electrical & Computer Engineering
Faculty of Engineering and Applied Science
Queen's University at Kingston, Canada

March 9, 2020

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Table of Contents

1 Intro – 3
2 Machine Vision – 4
  1 Passive Correspondence – 5
  2 Active Correspondence – 6
  3 Imperceptible Active Correspondence – 8
  4 Other Methods – 11
3 Machine Learning & Deep Learning – 16
4 Conclusion – 17

- The physical world is an enigma to a vast majority of computing systems
  - This is of little consequence to many applications
  - But a visual acclimation is an emerging and inherent requirement for a growing number of systems

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

- The physical world is an enigma to a vast majority of computing systems
  - This is of little consequence to many applications
  - But a visual acclimation is an emerging and inherent requirement for a growing number of systems

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Introduction

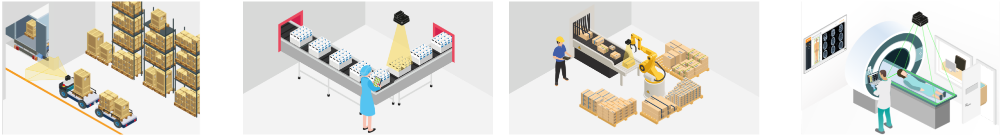
Why 3D Surface Sensing?

- The physical world is an enigma to a vast majority of computing systems
  - This is of little consequence to many applications
  - But a visual acclimation is an emerging and inherent requirement for a growing number of systems

PlayStation Camera [4]

iPhoneX [2]

Pixel 2 [3]



Basler AG [1]

Machine Vision

Correspondences:
Inference of spatial correspondences

- Stereovision

PlayStation Camera [4]

Active Correspondences:
Correspondences from active scene illumination

- PROCAM
- IR Depth Sensors

iPhoneX [2]

Other Methods:
Inference from active or passive analysis of the physical properties of light

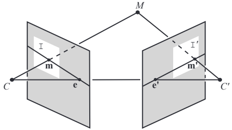
- Moving Lenses
- Plenoptic Lenses
- Time of Flight (Tof)

Pixel 2 [3]

Passive Correspondence Based Vision

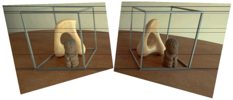
<u>*Characteristics:*</u>

- \+ Low cost
- \+ Real-time
- – High complexity
- – Dependent on scene morphology
- Generally produces a <u>*depth-map*</u>



Epipolar Geometry [5]

Stereo image pair [5]



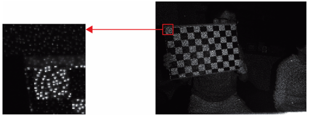
Rectifying Homography [5]

Active Correspondence Based Vision

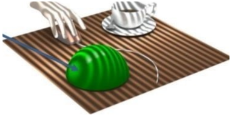
Structured Light

Structured light is the active illumination of a scene with specially designed spatially and/or temporally varying intensity patterns

Kinect V1 ©2012 IEEE [6]



Kinect V1 Sparse Dot Pattern ©2012 IEEE [6]

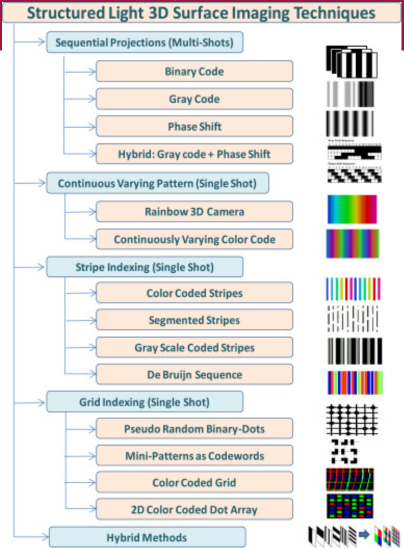


Zivid Time Multiplexed Stripes [8]

RealSense D435i [7]

Zivid [8]

Structured Light



Structured Light by Geng ©2011 Optical Society of America [9]

Category	Author	Year	Shots	Cameras	Axis	Pixel depth	Coding strategy	Subpixel acc.	Color
Discrete									
Spatial multiplexing									
De Bruijn	Boyer	1987	1	1	1	C	A	Y	N
	Salvi	1998	1	1	1	C	A	Y	Y
	Monks	1992	1	1	1	C	A	Y	N
	Pages	2004	1	1	1	C	A	Y	N
Non formal	Forster	2007	1	1	1	C	A	Y	N
	Fechteler	2008	1	1	1	C	A	Y	N
	Tehrani	2008	1	1	1	C	A	N	Y
	Maruyama	1993	1	1	2	B	A	N	Y
	Kawaski	2008	1	2	2	C	A	N	Y
	Ito	1995	1	1	2	G	A	N	Y
	Koninckx	2006	1	1	2	C	P	Y	Y
M-array	Griffin	1992	1	1	2	C	A	Y	Y
	Morano	1998	1	1	2	C	A	Y	Y
	Pages	2006	1	1	2	C	A	Y	N
	Albitar	2007	1	1	2	B	A	N	Y
Time multiplexing									
Binary codes	Posdamer	1982	> 2	1	1	B	A	N	Y
	Ishii	2007	> 2	1	1	B	A	N	N
	Sun	2006	> 2	2	1	B	A	Y	Y
N-ary codes	Caspi	1998	> 2	1	1	C	A	N	N
Shifting codes	Zhang	2002	> 2	1	1	C	A	Y	N
	Sansoni	2000	> 2	1	1	G	A	Y	Y
	Guhring	2001	> 2	1	1	G	A	Y	Y
Continuous									
Single phase Shifting (SPS)	Srinivasan	1985	> 2	1	1	G	P	Y	Y
	Ono	2004	> 2	1	1	G	P	Y	Y
	Wust	1991	1	1	1	C	P	Y	N
	Guan	2004	1	1	1	G	P	Y	Y
Multiple phase Shifting (MPS)	Gushov	1991	> 2	1	1	G	A	Y	Y
	Pribanić	2009	> 2	1	1	G	A	Y	Y
Frequency multiplexing									
Single coding frequency	Takeda	1983	1	1	1	G	P	Y	Y
	Cobelli	2009	1	1	1	G	A	Y	Y
	Su	1990	2	1	1	G	P	Y	Y
	Hu	2009	2	2	1	C	P	Y	Y
	Chen	2007	1	1	1	C	P	Y	N
	Yue	2006	1	1	1	G	P	Y	Y
	Chen	2005	2	1	1	G	P	Y	Y
	Berryman	2008	1	1	1	G	P	Y	Y
	Gdeisat	2006	1	1	1	G	P	Y	Y
	Zhang	2008	1	1	1	G	P	Y	Y
	Lin	1995	2	1	1	G	P	Y	Y
	Huang	2005	> 2	1	1	G	P	Y	Y
	Jia	2007	2	1	1	G	P	Y	Y
	Wu	2006	1	1	1	G	P	Y	Y
Spatial multiplexing									
Grading	Carrihill	1985	1	1	1	G	A	Y	N
	Tajima	1990	1	1	1	C	A	Y	N

Structured Light Classification by Salvi et al. [10]

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

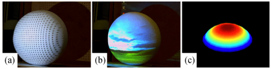
Imperceptible Active Correspondence Based Vision

Characteristics:

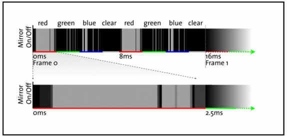
+ Dual-Function
− Hardware Driven
− Synchronized

Methods

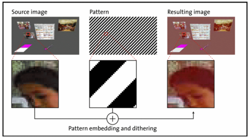
- Flicker Fusion
- Dithering
- High-speed projection
- Alternate Spectrums



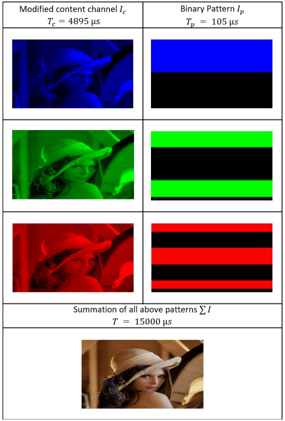
Flicker Fused Primitives ©2012 IEEE [11]



Mirror flip sequence for RGB (223,47,128) ©2004 IEEE [12]



Grey Code ©2004 IEEE [12]



High Speed Projection [13]

Evaluation of Throughput and Imperceptibility

Points Per Second (PPS):

$$PPS = Correspondences * Channels * framerate \tag{1}$$

Point Density:

$$PointDensity\left(\frac{PPS}{M * N * Channels}\right) = \frac{ChannelThroughput(bps)}{ChannelBandwidth(Hz)} \tag{2}$$

Mean Squared Error (MSE):

$$MSE = \frac{1}{M * N * D} \sum_{x=0}^{M} \sum_{y=0}^{N} (I_{Sample}(x, y) - I_{Ref}(x, y))^2 \tag{3}$$

Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 10 * log_{10}(\eta^2 / MSE) \tag{4}$$

Structural Similarity Index (SSIM):

$$SSIM(I_{Sample}, I_{Ref}) = \frac{(2 * \mu_{I_{Sample}} * \mu_{I_{Ref}} + (0.01 * \eta)^2)(\zeta_{I_{Sample}, I_{Ref}} + (0.03 * \eta)^2)}{(\mu^2_{I_{Sample}} + \mu^2_{I_{Ref}} + (0.01 * \eta)^2)(\zeta^2_{I_{Sample}} + \zeta^2_{I_{Ref}} + (0.03 * \eta)^2)} \tag{5}$$

Comparison of Methods

Table

Metrics of Proposed PROCAM Surface Imaging Sensors.

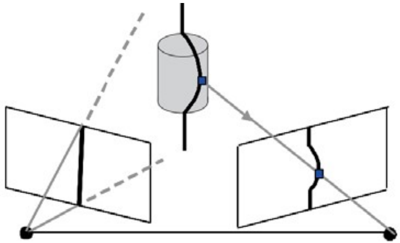
	Resolution	***FPS***	***Shots***	***PPS***	***DFP***	***FP***	***FPD***
Kinect V1 [14, 15]	320x240	30	1*	2.3M	76,800	76,800*	1.0*
Kinect V2 [14, 15]	512x424	30	1*	6.5M	217,088	217,088*	1.0*
RealSense D435 [7, 16]	1280 x 720	$90^{\star}$	1*	27.6M	921,600	921,600*	1.0*
Qiu et al. [17]	$1280x800^{\dagger}$	$2,720^{\ddagger}$	1	$11.7M^{\ddagger}$	4,315	4,315	0.004
Cole et al. [13]	1400x256	22.2	9	7.96M	358,400	39,822	0.019
Dai and Chung [18]	68x51	60	2	208,080	3,468	1,734	0.0022
Gong and Zhang [19]	480x480	4k	1	921.6M	230,400	230,400	1.0

Points Per Second (PPS), Depth Frame Points (DFP), Frame Points (FP), Frame Point Density (FPD)
*=missing, ⋆=conditional, †=wrong, ‡=*theoretical*

Other Methodologies

Examples:

- Moving Lenses
- Plenoptic Lenses
- Time of Flight (ToF)



Parallax [20]

Why seek alternatives?

- Parallax of triangulation-based vision
- Mitigate challenges with scene morphology
  - Textures absent or requiring radiometric compensation
  - Unfavorable surface diffusion patterns

Other Methodologies

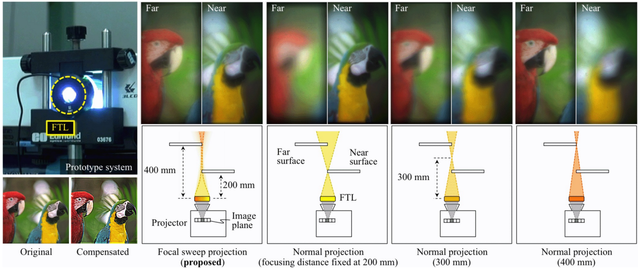
Moving Lenses

Characteristics:

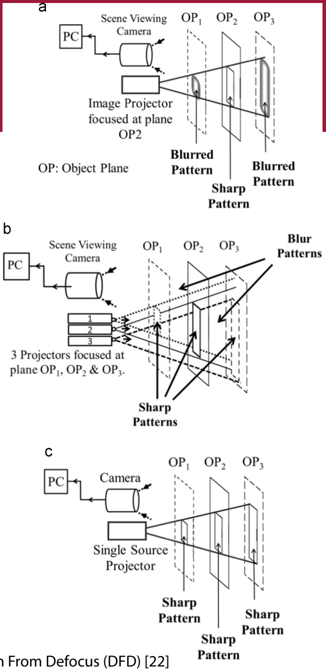
+ High depth of field

−/+ PCBV and ACBV possible

− Complex moving parts



Fast Focal Sweep Projection ©2015 IEEE [21]



Depth From Defocus (DFD) [22]

Intro

Machine Vision

PCBV

ACBV

IACBV

Evaluation

Other

ML & DL

Conclusion

References

Other Methodologies

Plenoptic Lenses

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Characteristics:

- \+ High depth of field
- – Specialized hardware
- – Dependent on scene illumination and texture

(a) Portion of the sensor image in the area of the wheel, captured by our camera.
(b) Normally rendered image with 2× zoom inset.
(c) Image rendered with focused plenoptic algorithm, with 2× zoom inset.

Plenoptic Image ©2009 IEEE [24]

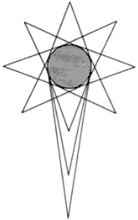


Fig. 1. Diagram from Leonardo's notebooks illustrating the fact that the light rays leaving an object's surface may be considered to form a collection of cones (which Leonardo calls "pyramids"), each cone constituting an image that would be seen by a pinhole camera at a given location.

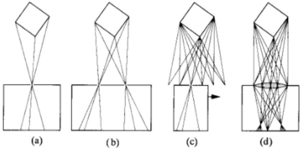


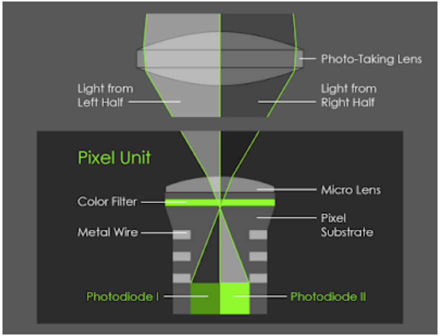
Fig. 2. (a) Pinhole camera forms an image from a single viewpoint; (b) in a stereo system, two images are formed from different viewpoints; (c) in a motion parallax system, a sequence of images are captured from many adjacent viewpoints; (d) a lens gathers light from a continuum of viewpoints; in an ordinary camera these images are averaged at the sensor plane.

Plenoptic Lense ©1992 IEEE [23]

Characteristics:

+ High depth of field
– Specialized hardware
– Dependent on scene illumination and texture

Pixel 2 [3]



Pixel 2 Lens [25]

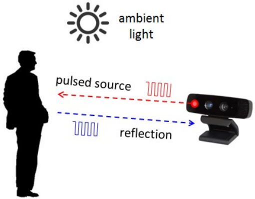
Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Other Methodologies

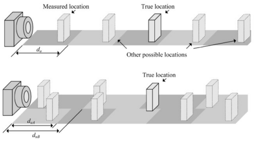
Time of Flight (ToF)

Characteristics:

+ High depth of field
+ Scene morphology invariant (mostly)
– Frequency inversely proportional to range



Time of Flight (ToF) [20]



Multi-Frequency ToF [20]

Kinect V2 [26]

Intel RealSense LiDAR Camera L515 [7]

Other Methodologies

ToF Comparison

CONSIDERATIONS	STEREO VISION	STRUCTURED-LIGHT	TIME-OF-FLIGHT (TOF)
Software Complexity	**High**	Medium	**Low**
Material Cost	**Low**	**High**	Medium
Compactness	**Low**	High	**Low**
Response Time	Medium	**Slow**	**Fast**
Depth Accuracy	**Low**	**High**	Medium
Low-Light Performance	**Weak**	**Good**	**Good**
Bright-Light Performance	**Good**	**Weak**	**Good**
Power Consumption	**Low**	Medium	**Scalable**
Range	**Limited**	**Scalable**	**Scalable**

Comparison of 3D Imaging Technologies. Copyright© 2014, Texas Instruments Incorporated [20]

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Machine Learning & Deep Learning

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Current implementations of ML & DL have been reserved to post-processing

–> Object Recognition, Image Segmentation, etc. . .

Proposed works:

- Stereo-vision by Sinz et al. [27]

  Demonstrated that Gaussian processes could be used to learn mappings from image to spatial coordinates
- Shape recognition by Dai and Chung [11, 18]

  Detected and recognized primitive shapes used to create codewords and correspondences
- Triangulation correspondence problem as a classification-regression by Fanello et al. [28]

  Demonstrated a 375Hz depth camera which used cascading random forests to infer depth from IR dot patterns.

Conclusion

3D Surface Imaging is area of growing interest and demand!

- Existing and proposed ACBV sensors offer static operation, often facing challenges with:
  - Changing environmental illumination
  - Radiometric compensation for textures
  - Diffraction and diffusion of light
  - Motion in scenes (multi-shot approaches)
  - Occlusion
- ML & DL offers an opportunity for greater versatility and performance
  - Specialized (and possibly dynamic) patterns
  - Training & testing for varied and dynamic environments
  - Dynamic operation for a greater depth of field, edge refinement, continuity and/or imperceptibility

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

Bibliography I

[1] Basler AG. Basler 3d cameras. Accessed 19/1/2020. [Online]. Available: https://www.baslerweb.com/en/products/cameras/3d-cameras/

[2] Apple Inc. Refurbished iphone x 64gb - space gray (unlocked). Accessed 05/3/2020. [Online]. Available: https://www.apple.com/shop/product/FQA52LL/A/Refurbished-iPhone-X-64GB-Space-Gray

[3] Google Inc. Google pixel 2 64gb unlocked gsm/cdma 4g lte octa-core phone w/ 12.2mp camera - just black. Accessed 05/3/2020. [Online]. Available: https://www.amazon.com/Google-Pixel-Unlocked-64gb-Black/dp/B0766GHWM6

[4] Sony Interactive Entertainment LLC. Playstation camera. Accessed 10/2/2020. [Online]. Available: https://www.playstation.com/en-ca/explore/accessories/playstation-camera-ps4/

[5] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," Tech. Rep. MSR-TR-99-21, April 1999. [Online]. Available: https://www.microsoft.com/en-us/research/publication/computing-rectifying-homographies-for-stereo-vision/

[6] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb 2012.

[7] Intel Corporation. Intel® realsense™ technology. Accessed 17/1/2020. [Online]. Available: https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html

[8] Zivid. Zivid. Accessed 17/1/2020. [Online]. Available: http://www.zivid.com/

[9] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Adv. Opt. Photon.*, vol. 3, no. 2, pp. 128–160, Jun 2011. [Online]. Available: http://aop.osa.org/abstract.cfm?URI=aop-3-2-128

[10] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognition*, vol. 43, no. 8, pp. 2666 – 2680, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031000124X

[11] J. Dai and R. Chung, "Embedding imperceptible codes into video projection and applications in robotics," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 4399–4404.

Bibliography II

[12] D. Cotting, M. Naef, M. Gross, and H. Fuchs, "Embedding imperceptible patterns into projected images for simultaneous acquisition and display," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nov 2004, pp. 100–109.

[13] A. Cole, S. Ziauddin, and M. Greenspan, "High-speed imperceptible structured light depth mapping," *Accepted in International Conference on Computer Vision Theory and Applications*, 2020.

[14] J. Park, H. Chao, H. Arabnia, and N. Y. Yen, "Advanced multimedia and ubiquitous engineering," *Future Information Technology*, vol. 2, 2015.

[15] M. Rahman, *Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interfaces*. Apress, 2017.

[16] Intel Corporation. Intel®realsensetmd400series product family. Accessed 14/2/2020. [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf

[17] Y. Qiu, J. Malcolm, A. Vatoo, S. Ziauddin, and M. Greenspan, "Inverse rectification for efficient procam pattern correspondence," *Accepted in Winter Conference on Applications of Computer Vision*, 2020.

[18] J. Dai and C. R. Chung, "Embedding invisible codes into normal video projection: Principle, evaluation, and applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 12, pp. 2054–2066, Dec 2013.

[19] Y. Gong and S. Zhang, "Ultrafast 3-d shape measurement with an off-the-shelf dlp projector," *Opt. Express*, vol. 18, no. 19, pp. 19 743–19 754, Sep 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-19-19743

[20] L. Li, "Time-of-flight camera–an introduction," *Technical white paper*, no. SLOA190B, 2014.

[21] D. Iwai, S. Mihara, and K. Sato, "Extended depth-of-field projector by fast focal sweep projection," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 4, pp. 462–470, 2015.

[22] M. J. Amin and N. A. Riza, "Active depth from defocus system using coherent illumination and a no moving parts camera," *Optics Communications*, vol. 359, pp. 135–145, 2016.

Bibliography III

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References

[23] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, Feb 1992.

[24] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *2009 IEEE International Conference on Computational Photography (ICCP)*, April 2009, pp. 1–8.

[25] M. Levoy and Y. Pritch, "Portrait mode on the pixel 2 and pixel 2 xl smartphones," *Google AI Blog*, October 2017, accessed 19/1/2020. [Online]. Available: https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html

[26] Depthkit, "Kinect for windows v2," https://docs.depthkit.tv/docs/kinect-for-windows-v2, (Accessed on 02/12/2020).

[27] F. H. Sinz, J. Q. Candela, G. H. Bakır, C. E. Rasmussen, and M. O. Franz, "Learning depth from stereo," in *Pattern Recognition*, C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 245–252.

[28] S. Ryan Fanello, C. Rhemann, V. Tankovich, A. Kowdle, S. Orts Escolano, D. Kim, and S. Izadi, "Hyperdepth: Learning depth from structured light without matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5441–5450.

This presentation has been prepared for the PhD Advisory Committee, per conformity with the requirements for PhD Comprehensive Examination by Queen's University's Department of Electrical and Computer Engineering

Queen's University
99 University Avenue
Kingston, Ontario K7L 3N6
Canada
www.queensu.ca

Intro
Machine Vision
PCBV
ACBV
IACBV
Evaluation
Other
ML & DL
Conclusion
References