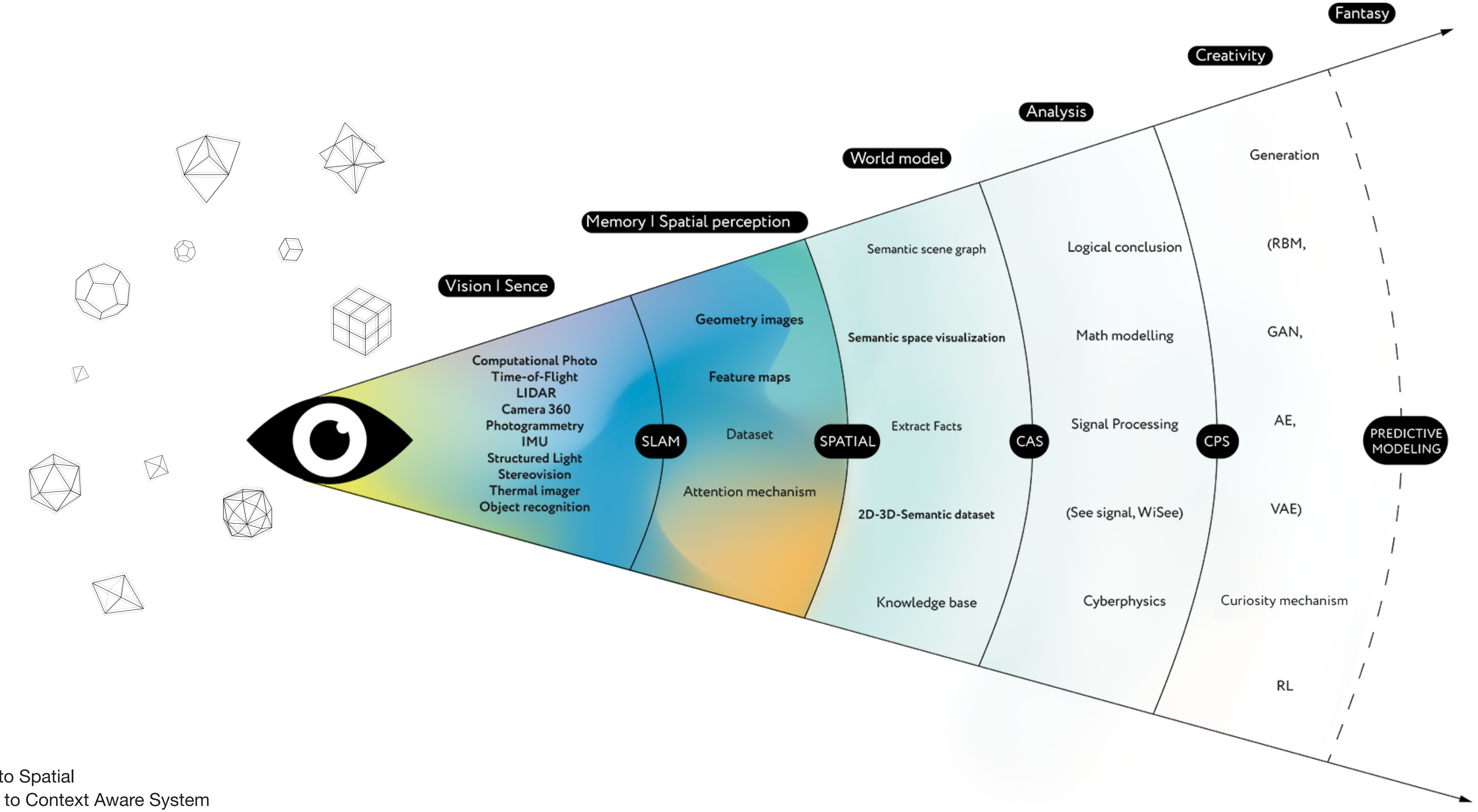
MACHINE VISION

PHYGITALISM

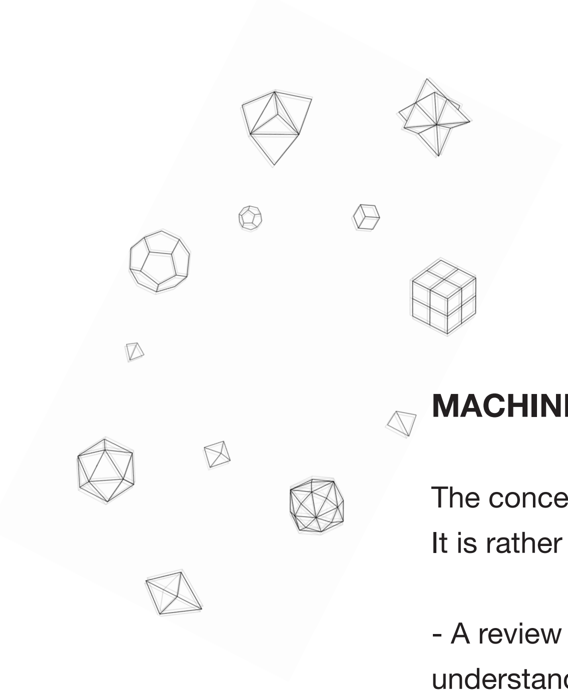
v. 01
2020

MACHINE VISION



- From SLAM to Spatial
- From Spatial to Context Aware System
- From Context Aware System to CPS
- From CPS to Predictive Modeling

INDEX



MACHINE VISION

The concept of **Vision** should not be taken literally - eyesight.
It is rather vision, foresight. This visual research has several objectives:

- A review of existing technologies and algorithms in the form of an understandable scheme, and an attempt to forecast future technologies.

- To present the algorythmic aesthetics of **Machine Vision** as art - phygital art - we call it Phygitalism, an understanding of art through the fusion of the physical and digital worlds. This concept, first of all, includes projects exploring the impact of the digital world on the environment, exploring the relationship of human and machine, issues of digital ethics. Many projects created in the areas of science art and tech art can be interpreted in the same way as digital art.

Machine vision, in computer vision, is the ability of a computer to see; it employs one or more video cameras, analog-to-digital conversion (ADC) and digital signal processing (DSP). The resulting data goes to a computer or robot controller. Machine vision is similar in complexity to voice recognition.

Two important specifications in any vision system are the sensitivity and the resolution. Sensitivity is the ability of a machine to see in dim light, or to detect weak impulses at invisible wavelengths. Resolution is the extent to which a machine can differentiate between objects. In general, the better the resolution, the more confined the field of vision. Sensitivity and resolution are interdependent. All other factors held constant, increasing the sensitivity reduces the resolution, and improving the resolution reduces the sensitivity.

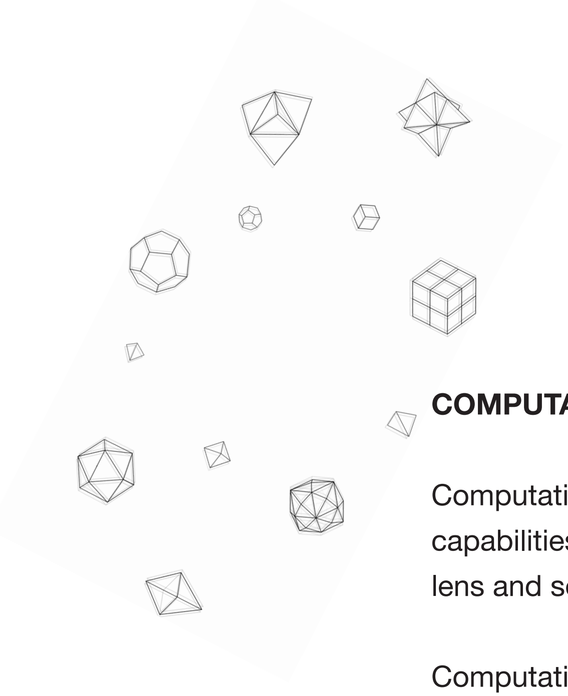
Human eyes are sensitive to electromagnetic wavelength is ranging from 390 to 770 nanometers (nm). Video cameras can be sensitive to a range of wavelengths much wider than this. Some machine-vision systems function at infrared (IR), ultraviolet (UV), or X-ray wavelengths.

Binocular (stereo) machine vision requires a computer with an advanced processor. In addition, high-resolution cameras, a large amount of random access memory (RAM), and artificial intelligence (AI) programming are required for depth perception.

Machine vision is used in various industrial and medical applications.

Examples include:

- Electronic component analysis
- Signature identification
- Optical character recognition
- Handwriting recognition
- Object recognition



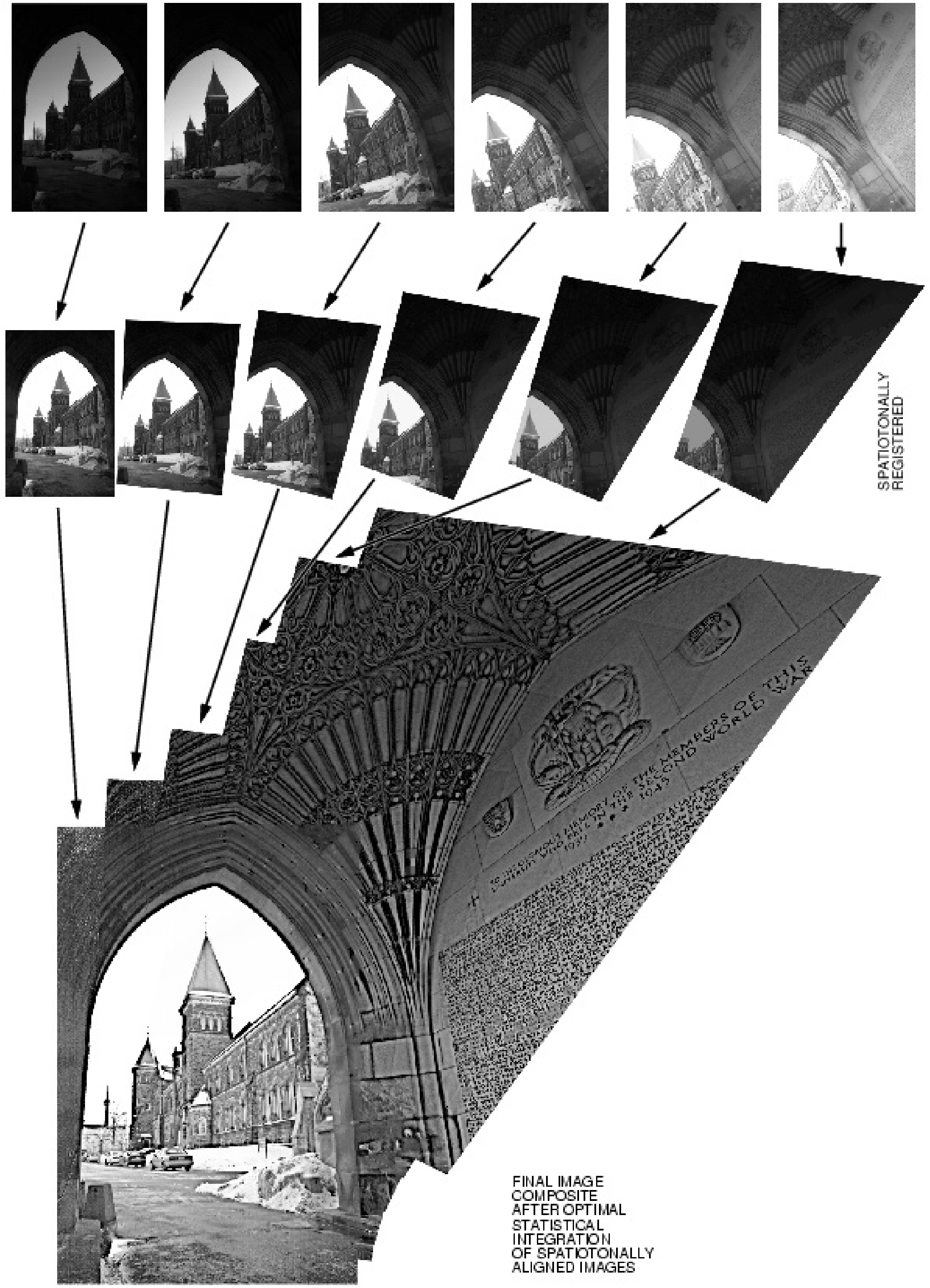
COMPUTATIONAL PHOTO

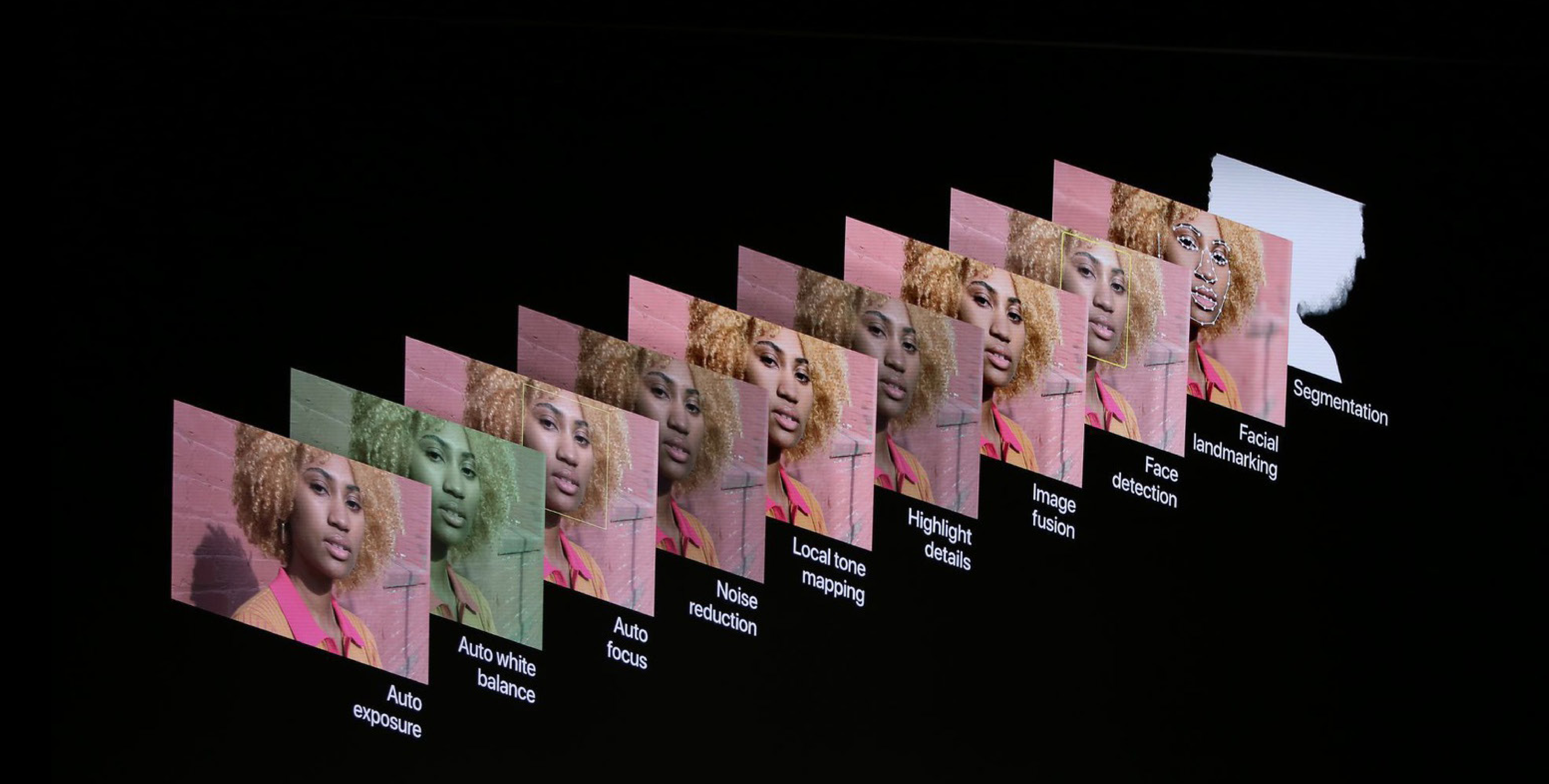
Computational photography — is the use of computer processing capabilities in cameras to produce an enhanced image beyond what the lens and sensor pics up in a single shot.

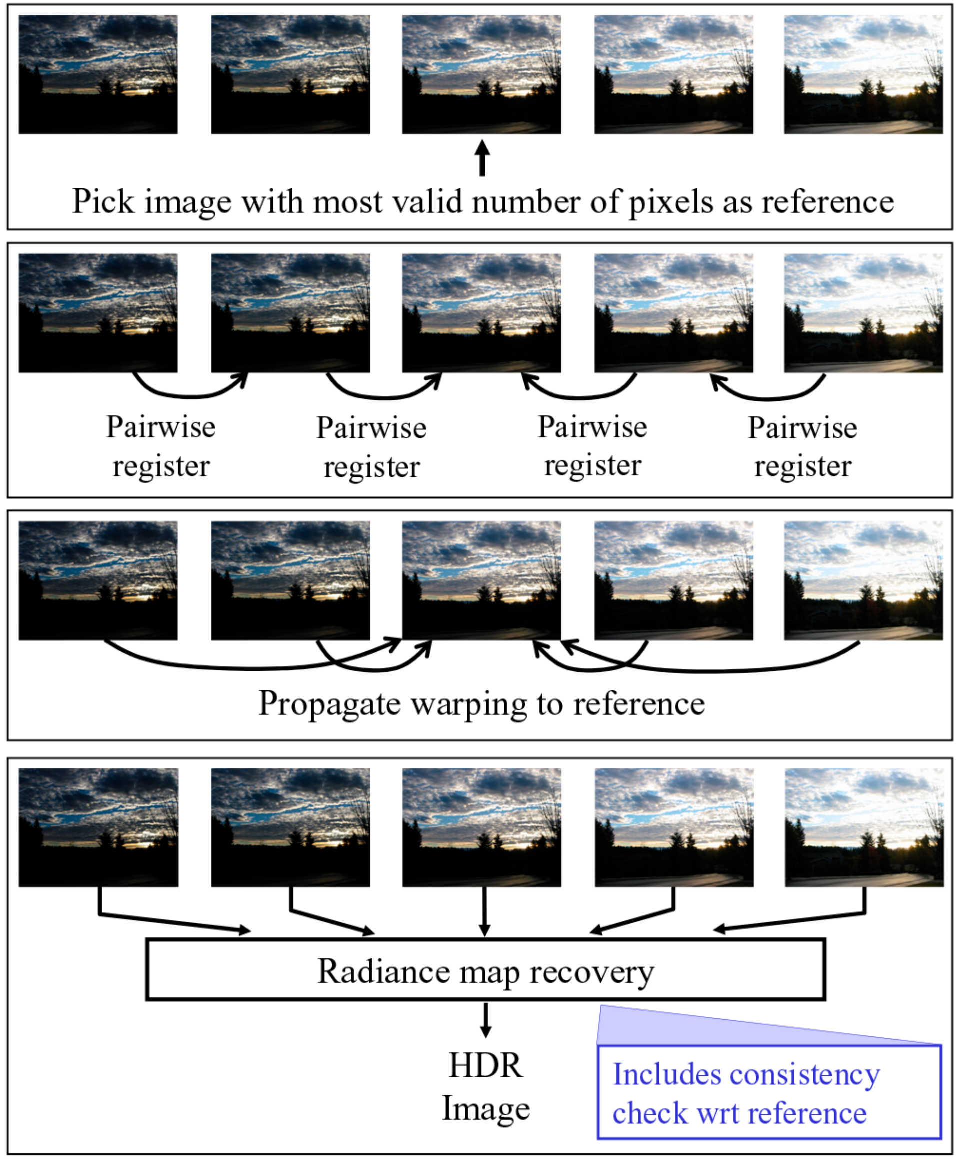
Computational photography is common in digital cameras and especially smartphones, automating many settings to make for better point-and-shoot abilities. Using image processing algorithms, computational photography improves images by reducing motion blur and adding simulated depth of field, as well as improving color, contrast and light range.

While enhancements are often added to digital cameras, they are even more frequently featured in smartphones, where there is less space for a big lens that might otherwise enhance pictures. Smartphones also have a relative abundance of processing power over the typical digital camera.

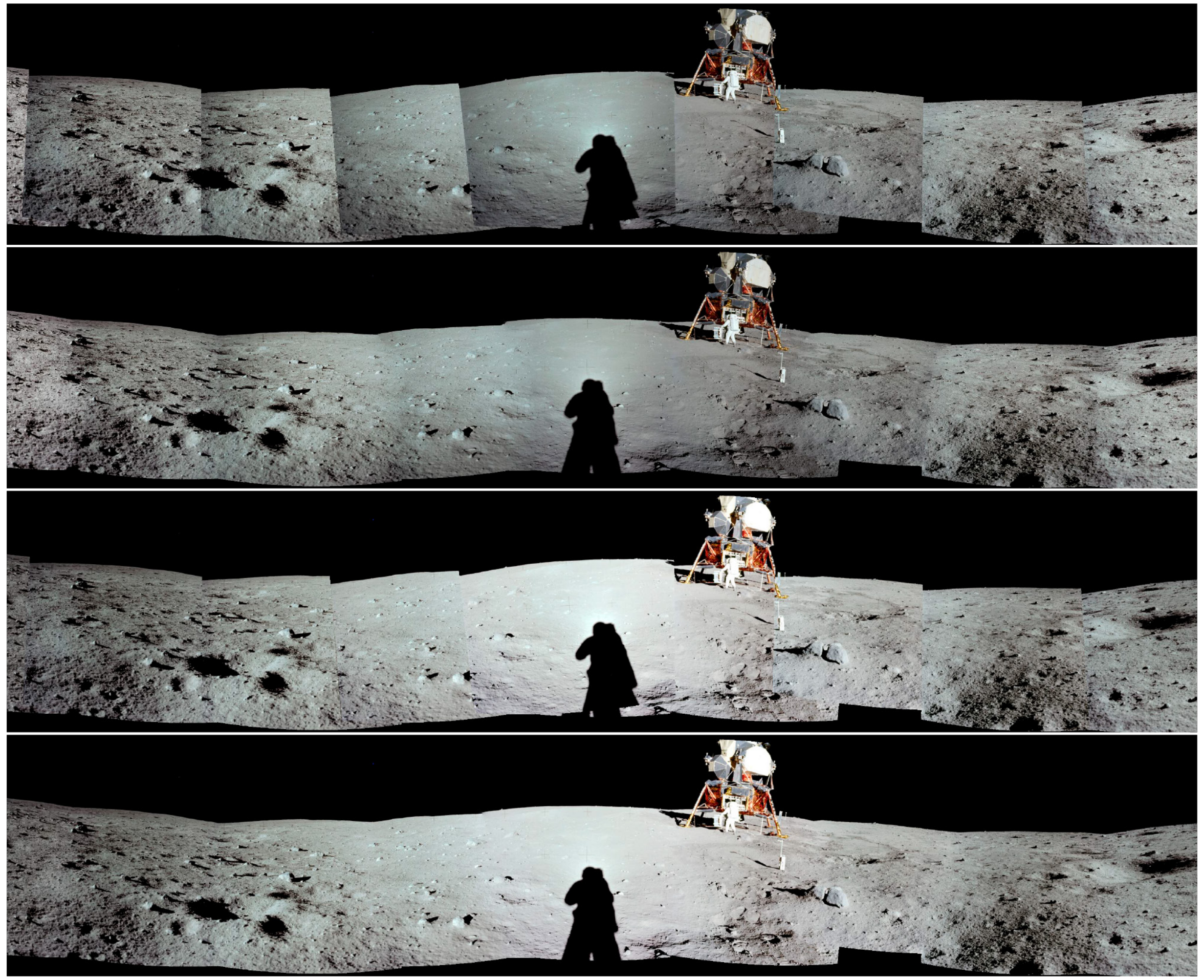
In computational photography, a number of pictures are often taken for cross-referencing. After cross-referencing images, the software automates many of the settings a photographer might carefully set to make artful shots. These images are sometimes cut into tiles and sections from some frames and might be blended, dropped or used to eliminate blur from random movement or get the best detail or light and dark balance. Some features in digital photography, like image stabilization, are both in hardware and in software.



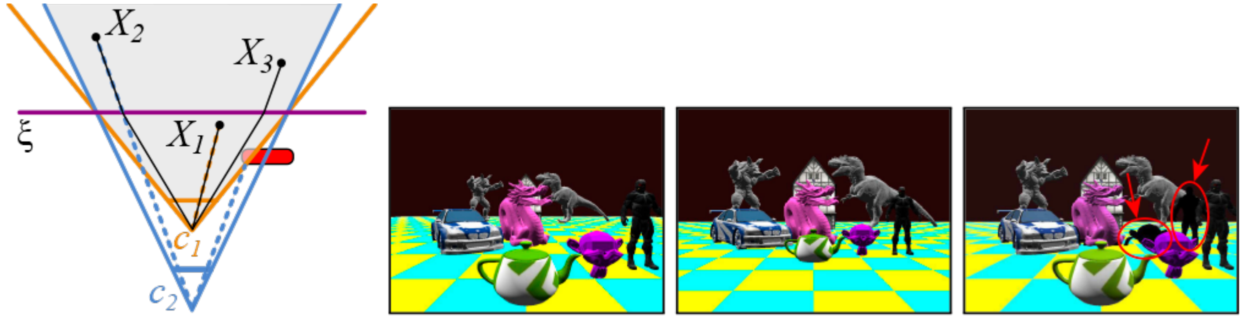

Auto exposure
Auto white balance
Auto focus
Noise reduction
Local tone mapping
Highlight details
Image fusion
Face detection
Facial landmarking
Segmentation



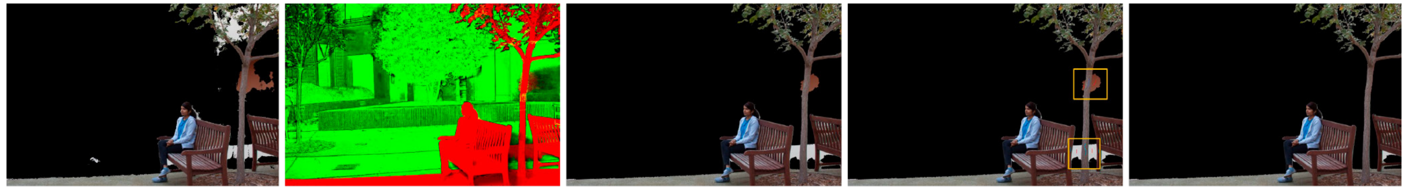
Steps in creating HDR image from image stills.
High Dynamic Range Video (Sing Bing Kang Matthew Uyttendaele Simon Winder Richard Szeliski)



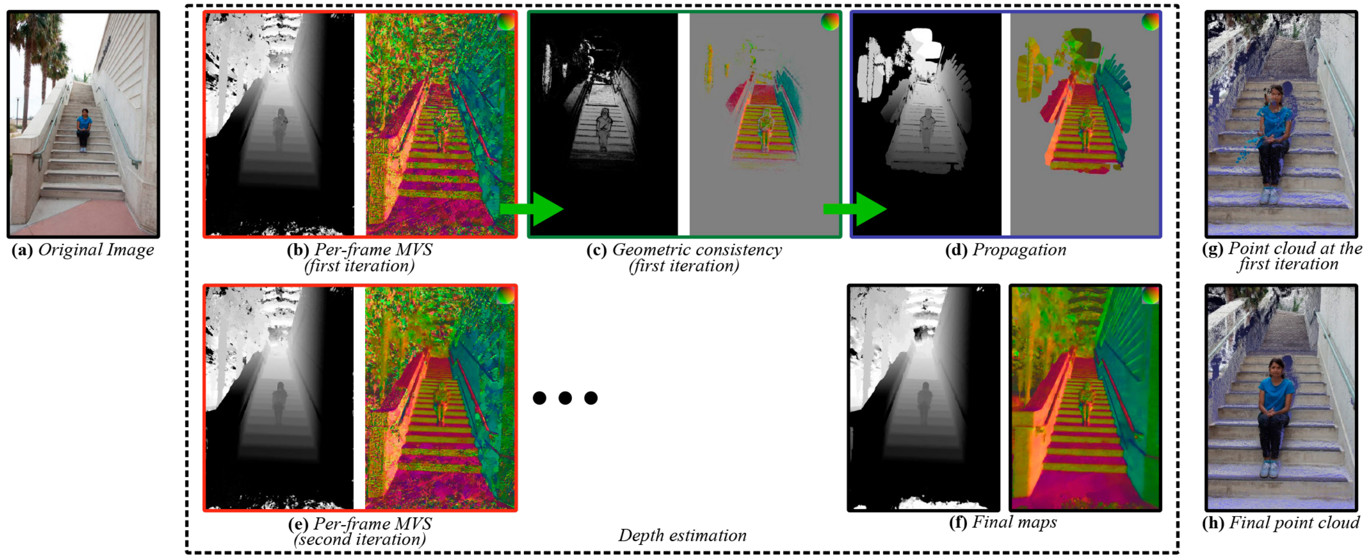
Apollo 11. From top: original images; original images after Photomontage blending; vignette compensation only; vignette compensation and Photomontage blending.
Vignette and exposure calibration and compensation (Dan B Goldman, 2005)



(a) (b) Short (c) Long (d) Short-long



(a) Depth thresholding (b) Confidence map (c) Refined segmentation (d) User scribbles (e) Final result

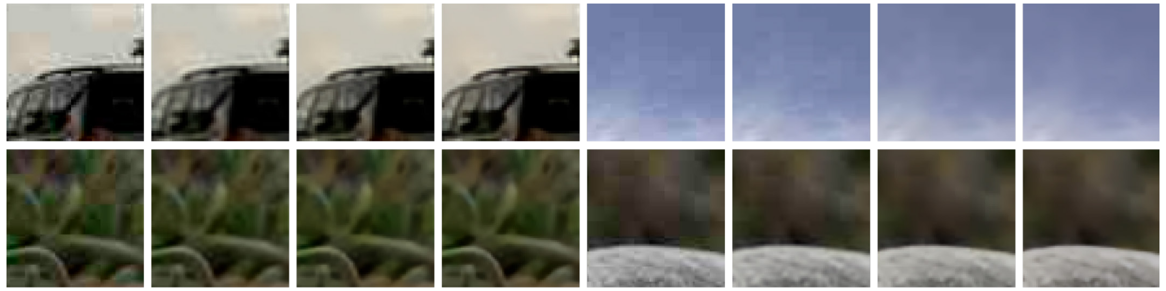


(a) *Original Image* **(b)** *Per-frame MVS (first iteration)* **(c)** *Geometric consistency (first iteration)* **(d)** *Propagation* **(g)** *Point cloud at the first iteration*

(e) *Per-frame MVS (second iteration)* *Depth estimation* **(f)** *Final maps* **(h)** *Final point cloud*

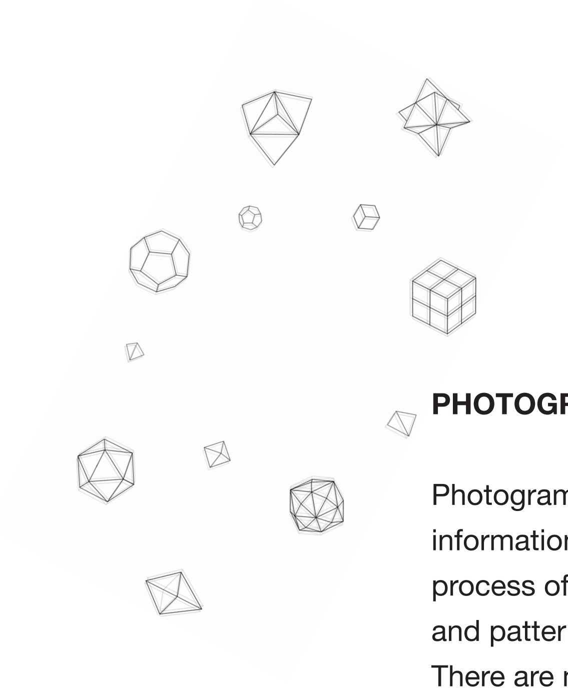
Computational Zoom: A Framework for Post-Capture Image Composition
(Abhishek Badki, Santa Barbara And Nvidia Orazio Gallo, Jan Kautz, Pradeep Sen, 2017)

(a) Clean (b) Noisy (c) BM3D (d) ℓ_2 (e) ℓ_1 (f) SSIM (g) MS-SSIM (h) Mix



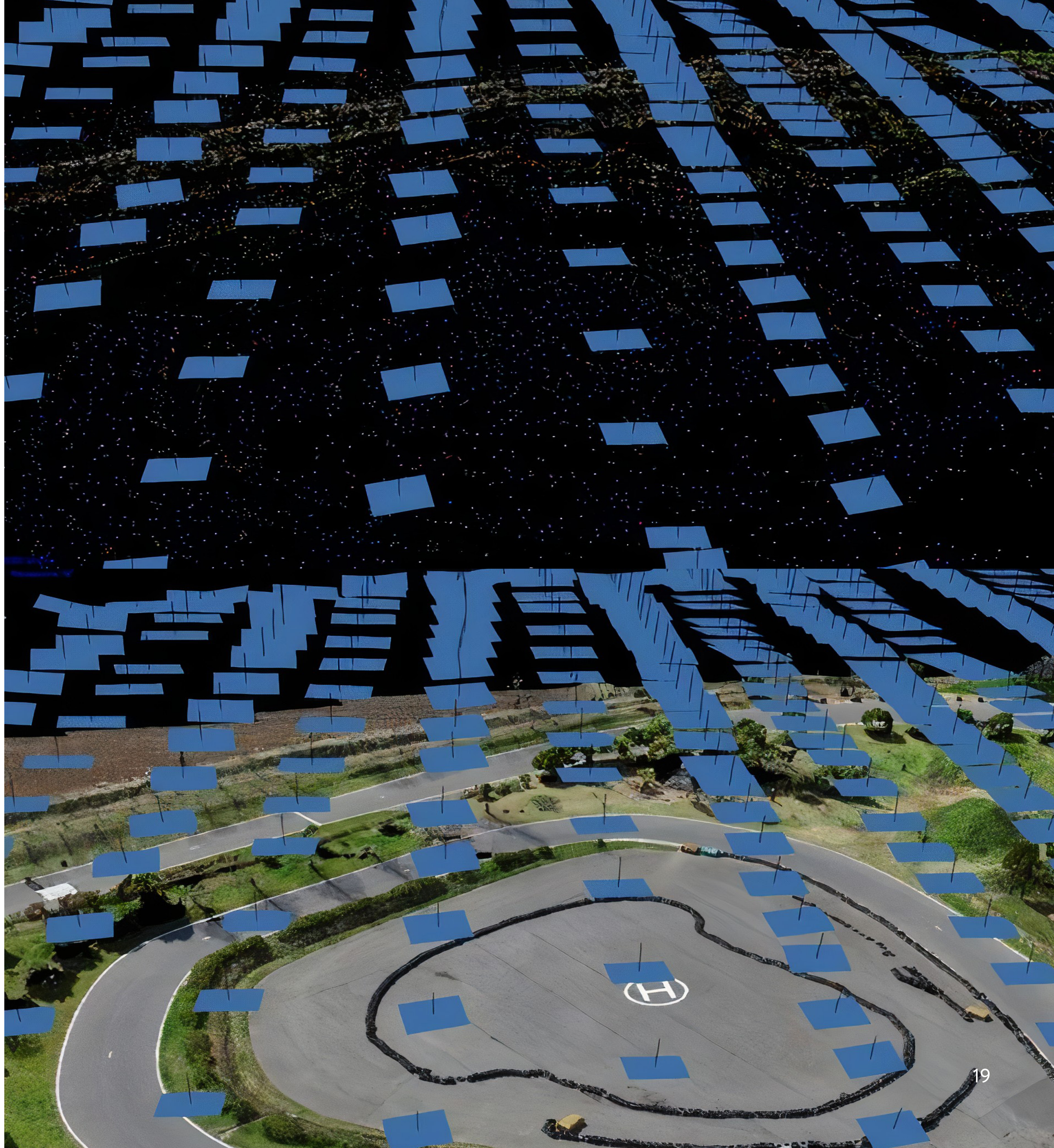
(a) JPEG (b) ℓ_2 (c) ℓ_1 (d) Mix (e) JPEG (f) ℓ_2 (g) ℓ_1 (h) Mix

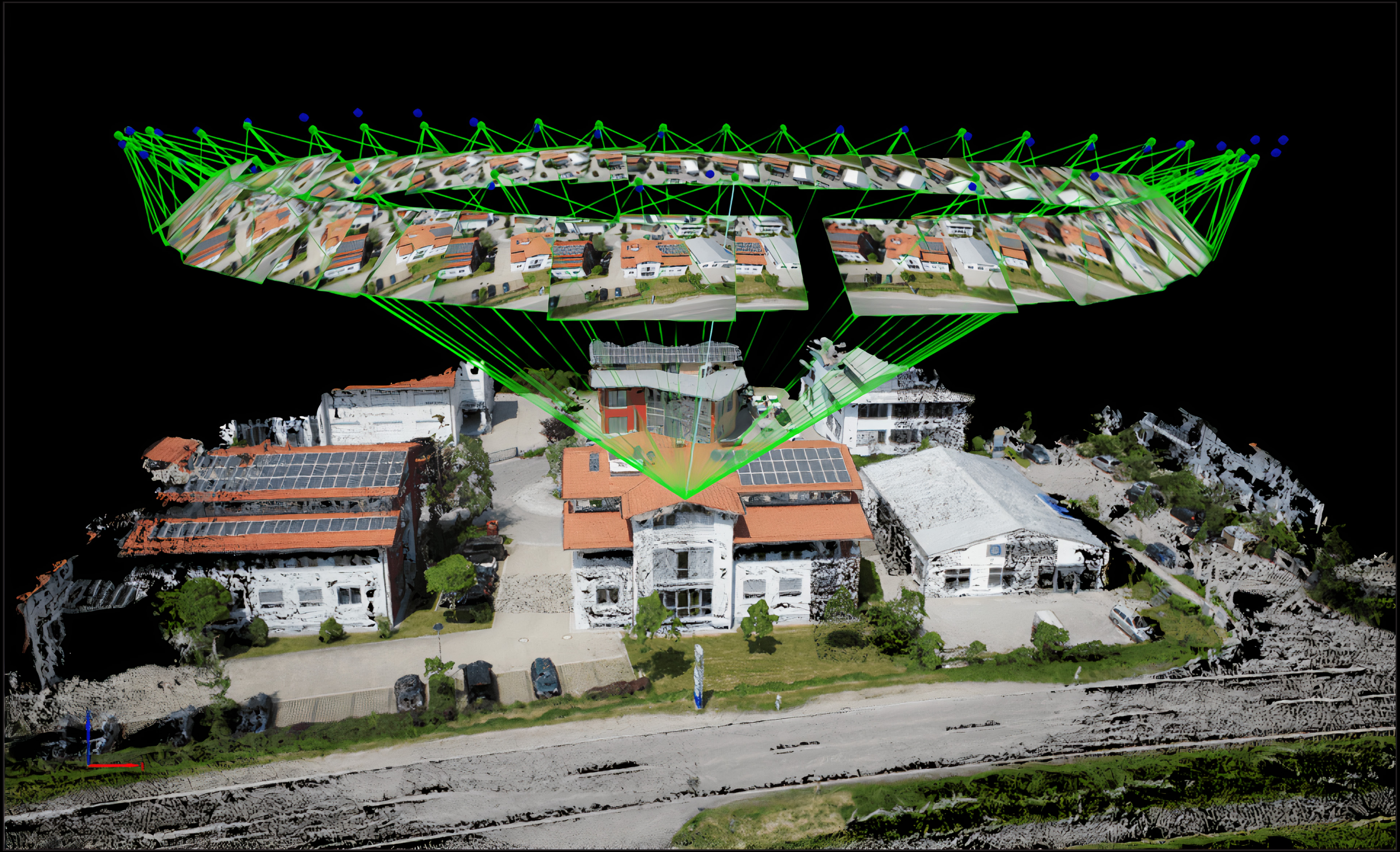
Loss Functions for Image Restoration with Neural Networks
(Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, 2017)

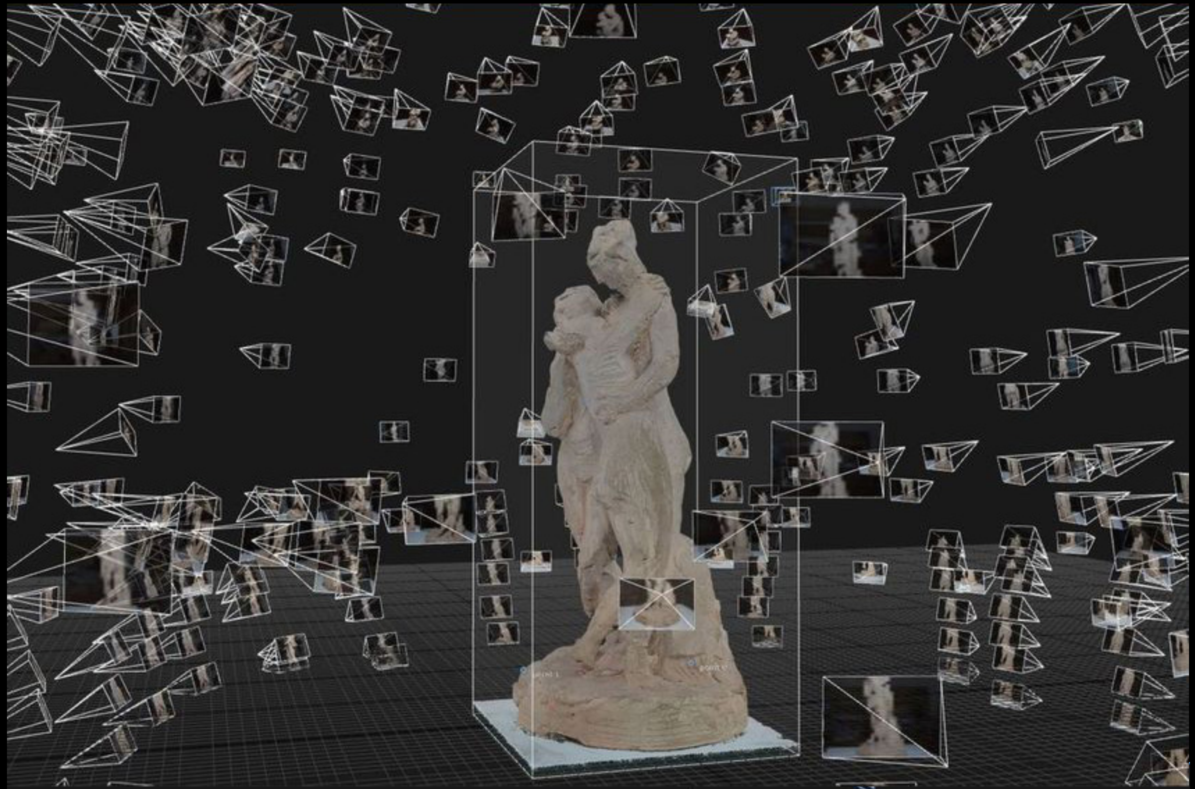


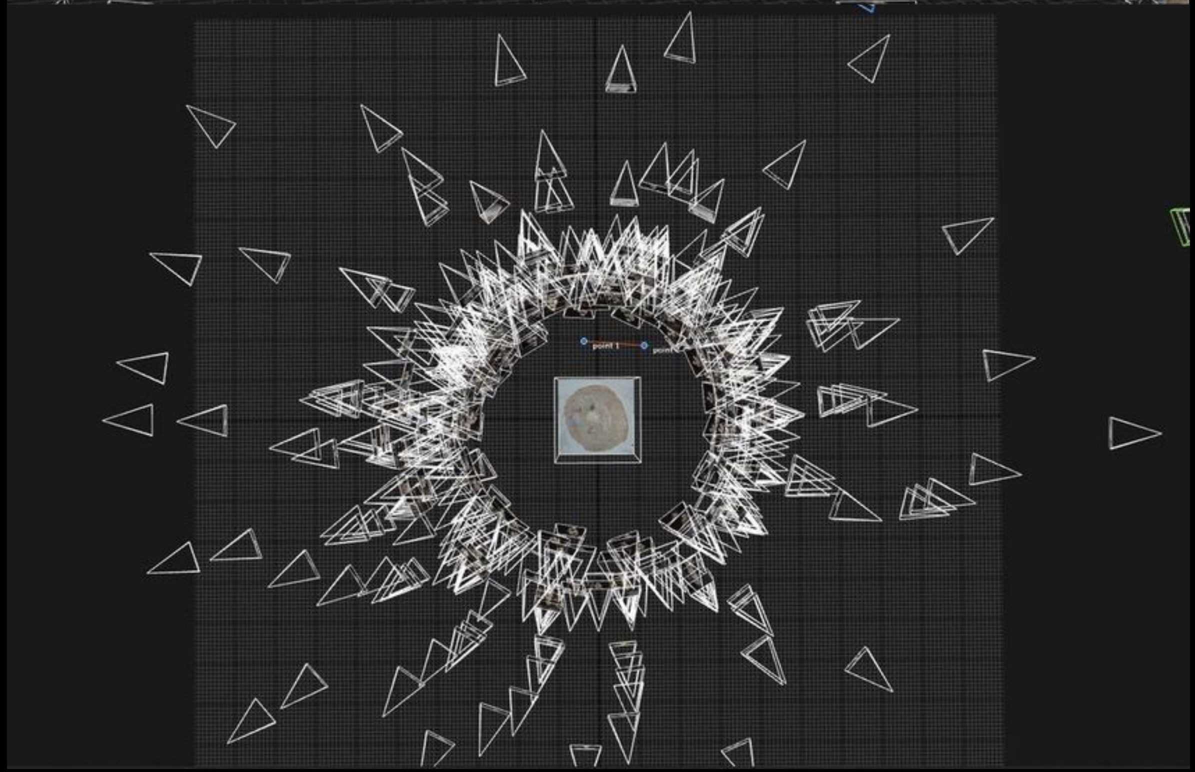
PHOTOGRAMMETRY

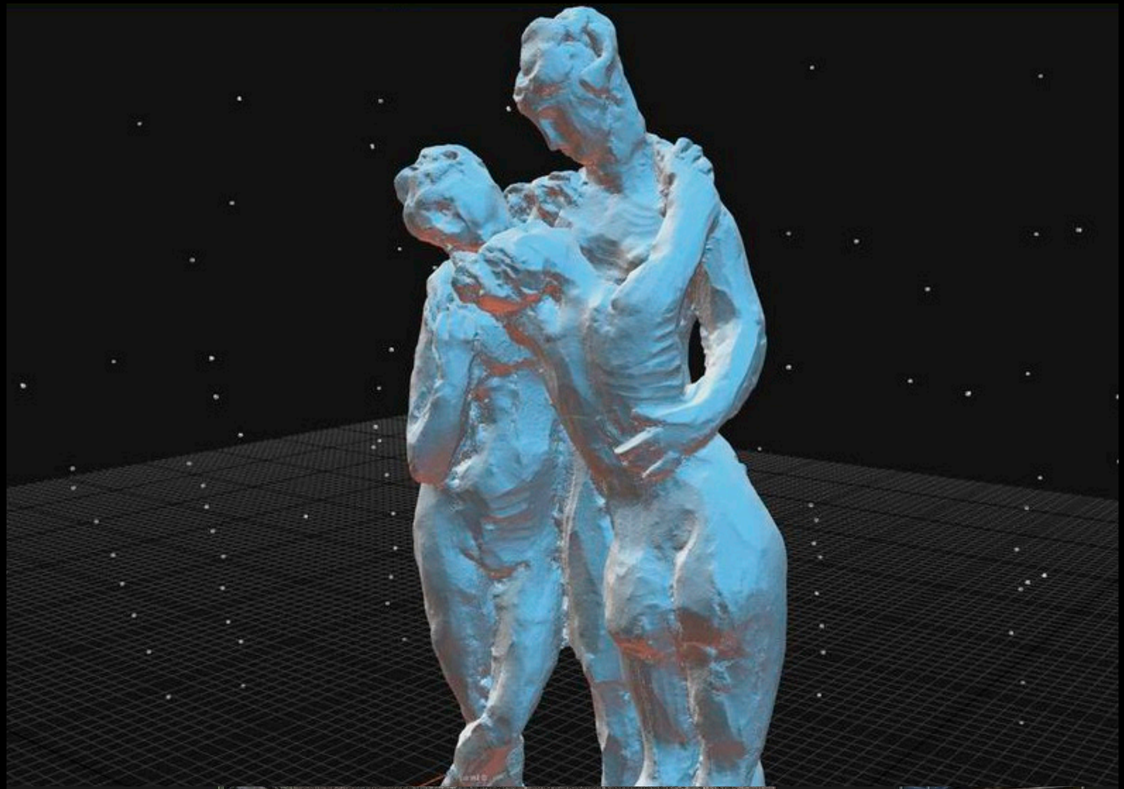
Photogrammetry — is the science and technology of obtaining reliable information about physical objects and the environment through the process of recording, measuring and interpreting photographic images and patterns of electromagnetic radiant imagery and other phenomena. There are many variants of photogrammetry. One example is the extraction of three-dimensional measurements from two-dimensional data (i.e. images); for example, the distance between two points that lie on a plane parallel to the photographic image plane can be determined by measuring their distance on the image, if the scale of the image is known. Another is the extraction of accurate color ranges and values representing such quantities as albedo, specular reflection, metallicity, or ambient occlusion from photographs of materials for the purposes of physically based rendering.
Close-range photogrammetry refers to the collection of photography from a lesser distance than traditional aerial (or orbital) photogrammetry. Photogrammetric analysis may be applied to one photograph, or may use high-speed photography and remote sensing to detect, measure and record complex 2D and 3D motion fields by feeding measurements and imagery analysis into computational models in an attempt to successively estimate, with increasing accuracy, the actual, 3D relative [motions.]







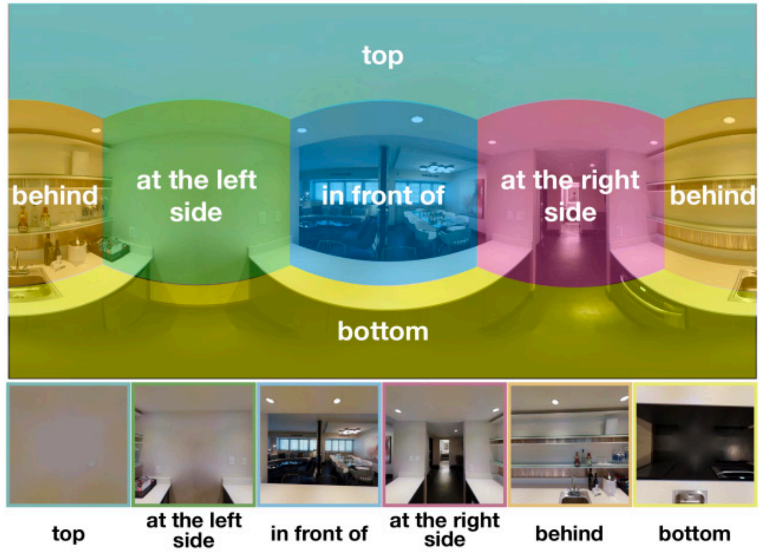


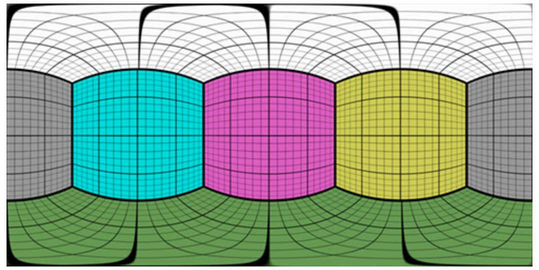


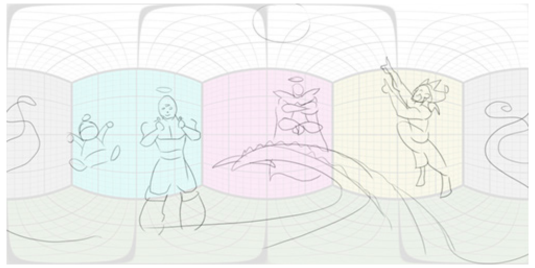
CAMERA 360

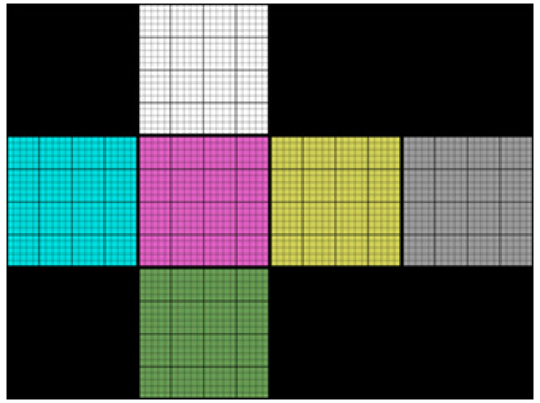
Camera 360 — in photography, an omnidirectional camera (from «omni», meaning all), also known as 360-degree camera, is a camera having a field of view that covers approximately the entire sphere or at least a full circle in the horizontal plane. Omnidirectional cameras are important in areas where large visual field coverage is needed, such as in panoramic photography and robotics.

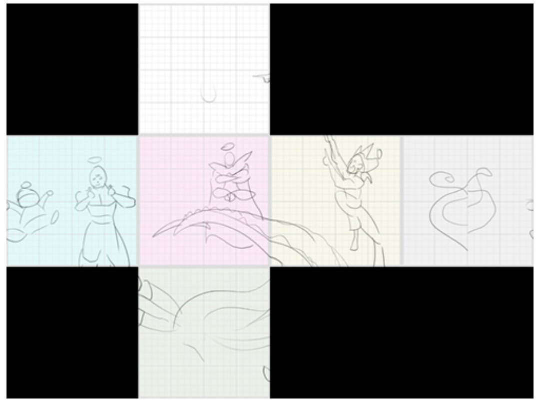
A camera normally has a field of view that ranges from a few degrees to, at most, 180°. This means that it captures, at most, light falling onto the camera focal point through a hemisphere. In contrast, an ideal omnidirectional camera captures light from all directions falling onto the focal point, covering a full sphere. In practice, however, most omnidirectional cameras cover only almost the full sphere and many cameras which are referred to as omnidirectional cover only approximately a hemisphere, or the full 360° along the equator of the sphere but excluding the top and bottom of the sphere. In the case that they cover the full sphere, the captured light rays do not intersect exactly in a single focal point.

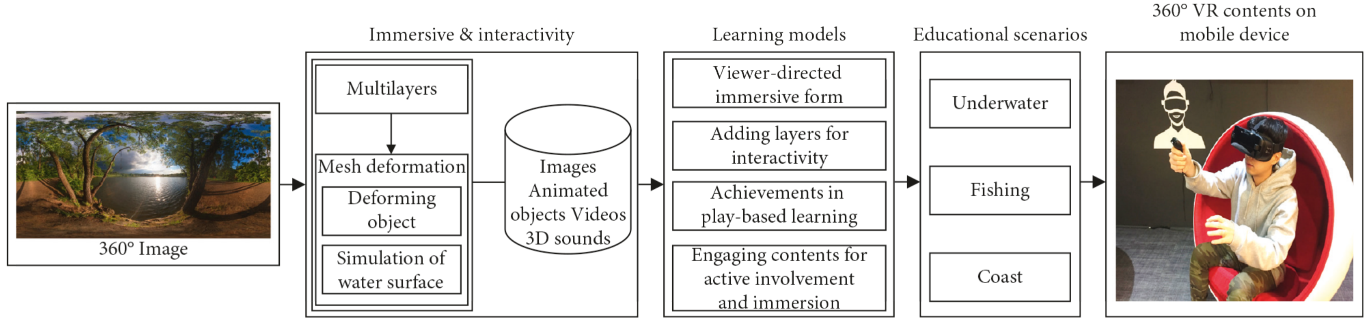

top
behind
at the left side
in front of
at the right side
behind
bottom
top
at the left side
in front of
at the right side
behind
bottom











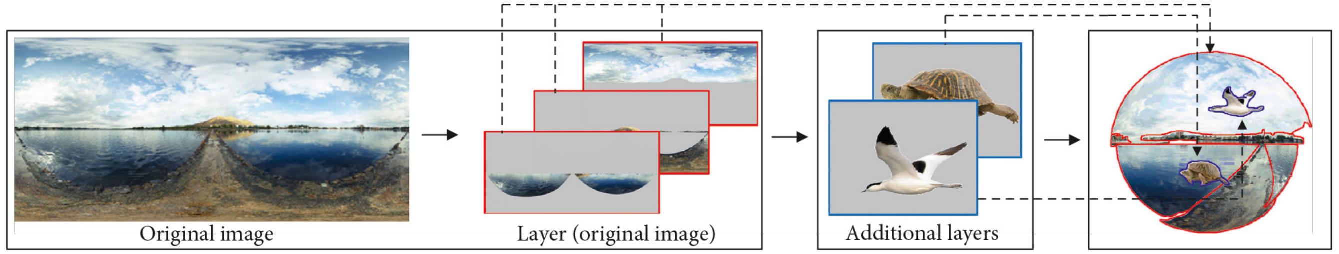


FIGURE 2: Interactive 360° VR contents with multiple layers.

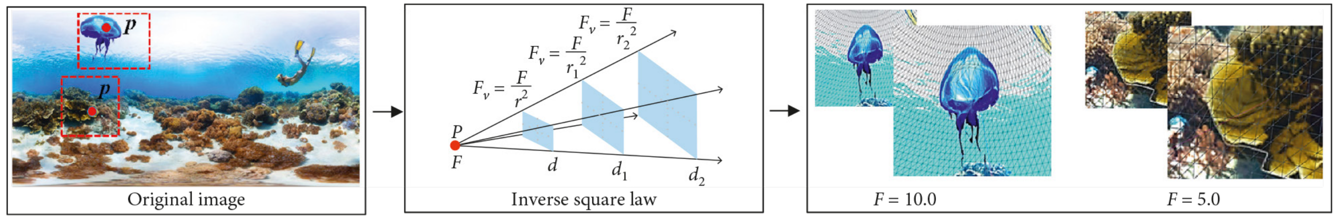
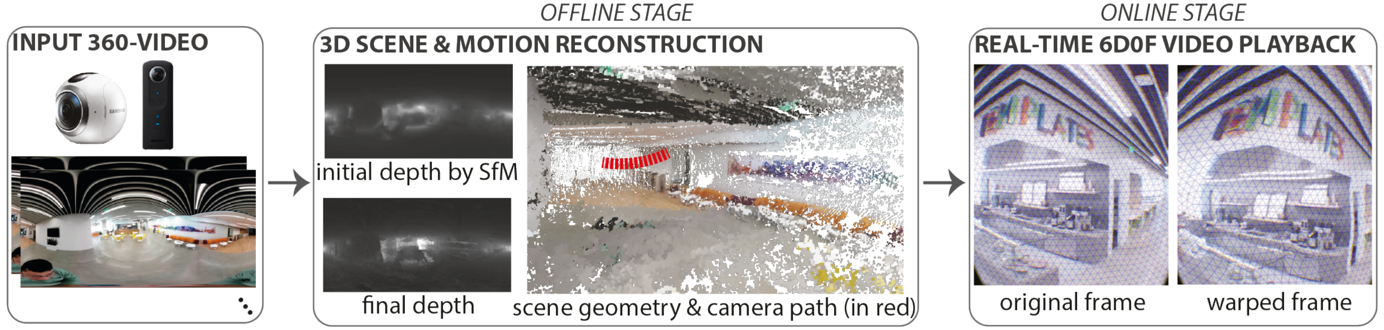
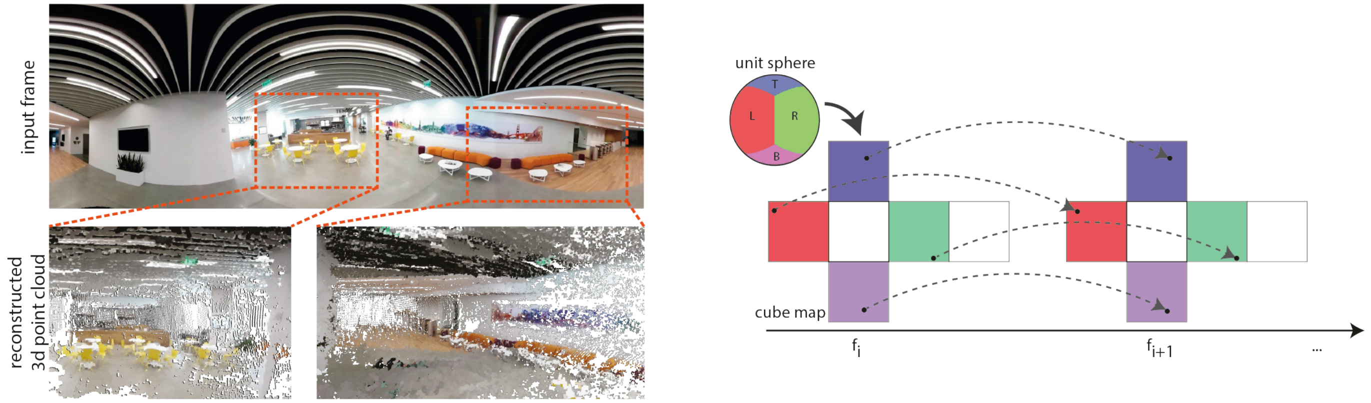
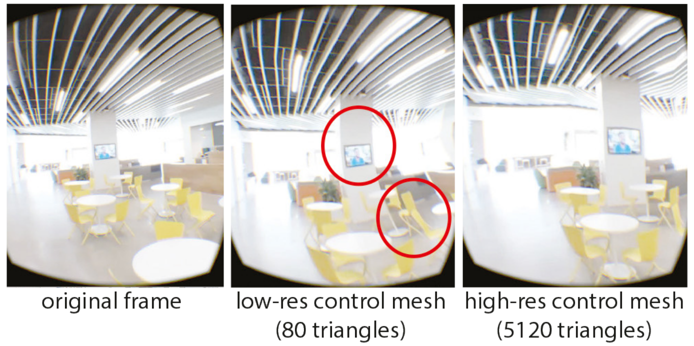


FIGURE 3: Force-based deformation of a mesh on a spherical surface.

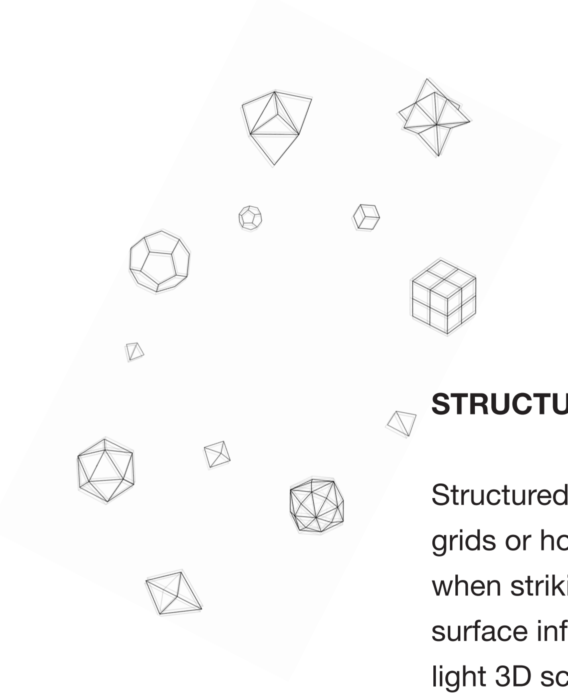
Interactive and Immersive Learning Using 360° Virtual Reality
Contents on Mobile Platforms (Kanghyun Choi, Yeo-Jin Yoon, Oh-Young Song and Soo-Mi Choi, 2018)







6-DOF VR Videos with a Single 360-Camera
(Jingwei Huang, Zhili Chen, Duygu Ceylan, Hailin Jin)

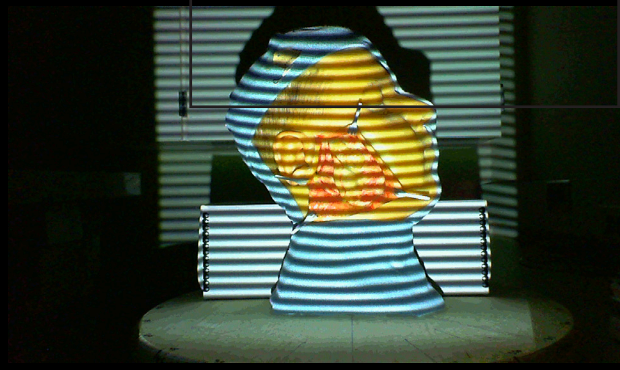


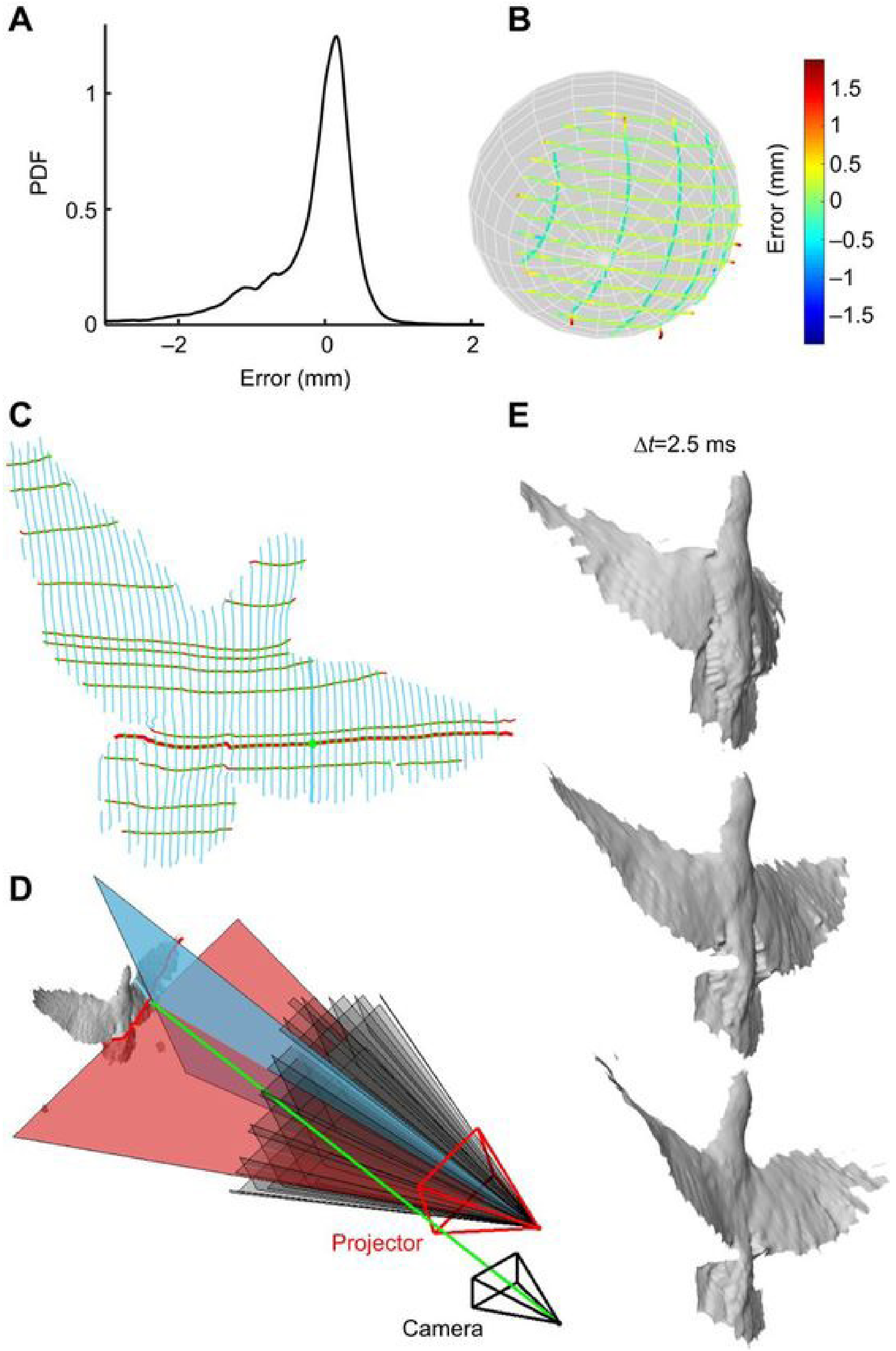
STRUCTURED LIGHT

Structured light is the process of projecting a known pattern (often grids or horizontal bars) on to a scene. The way that these deform when striking surfaces allows vision systems to calculate the depth and surface information of the objects in the scene, as used in structured light 3D scanners.

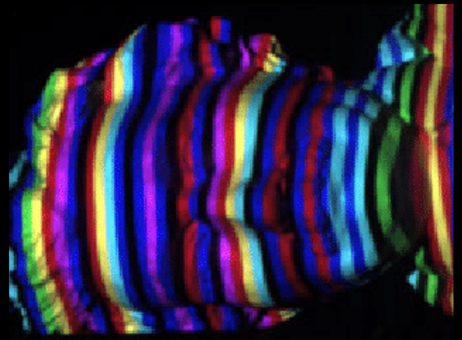
Invisible (or imperceptible) structured light uses structured light without interfering with other computer vision tasks for which the projected pattern will be confusing. Example methods include the use of infrared light or of extremely high frame rates alternating between two exact opposite patterns.

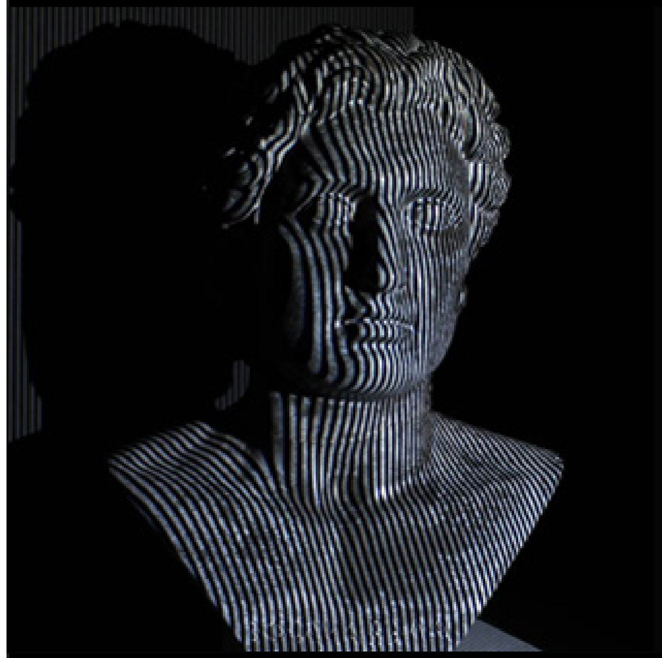
Structured light is used by a number of police forces for the purpose of photographing fingerprints in a 3D scene. Where previously they would use tape to extract the fingerprint and flatten it out, they can now use cameras and flatten the fingerprint digitally, which allows the process of identification to begin before the officer has even left the scene.

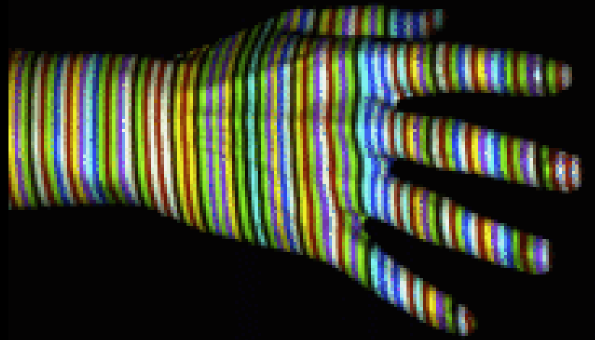


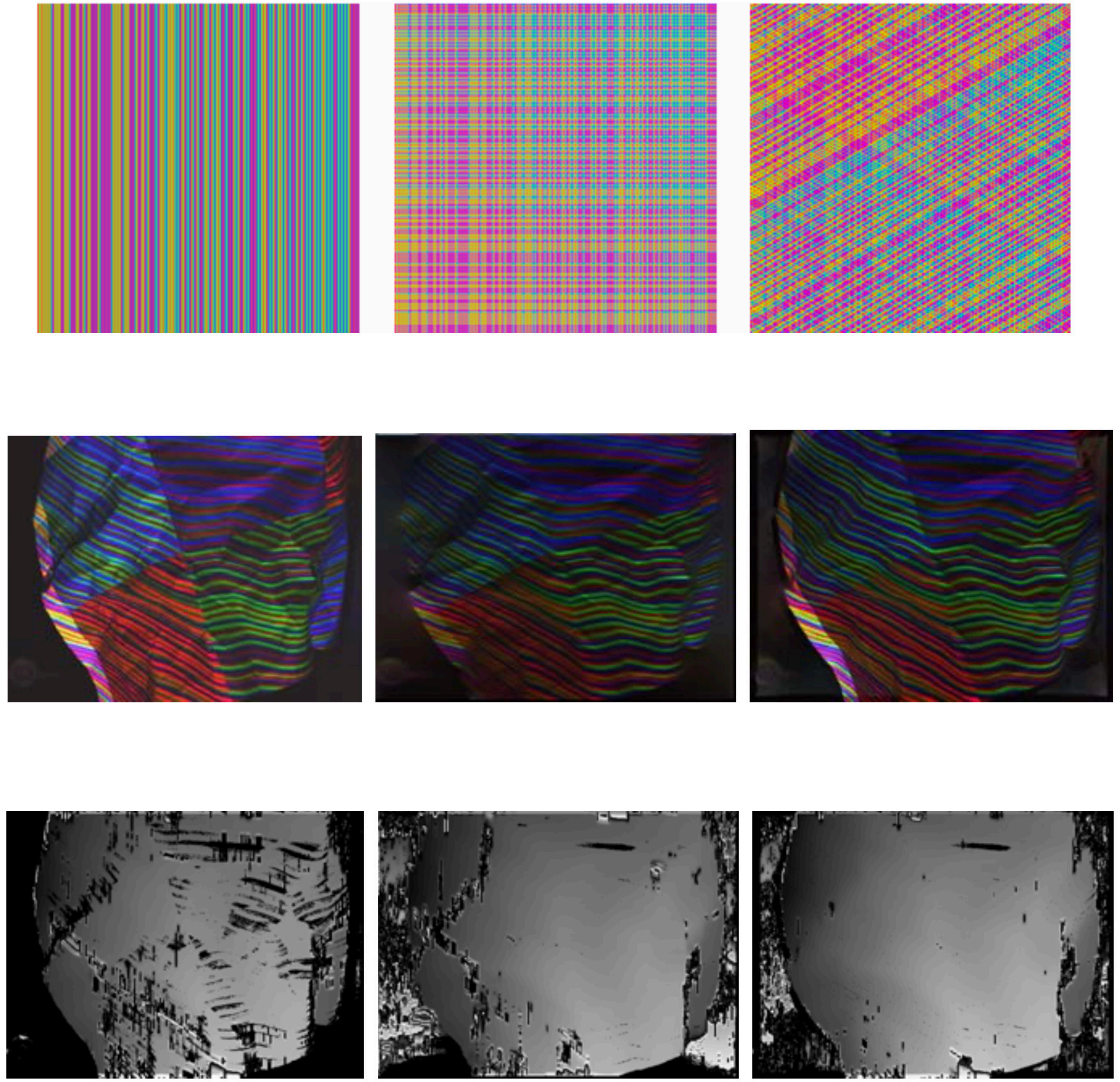


High-speed surface reconstruction of a flying bird using structured light
(Marc E. Deetjen, Andrew A. Biewener, David Lentink, 2017)

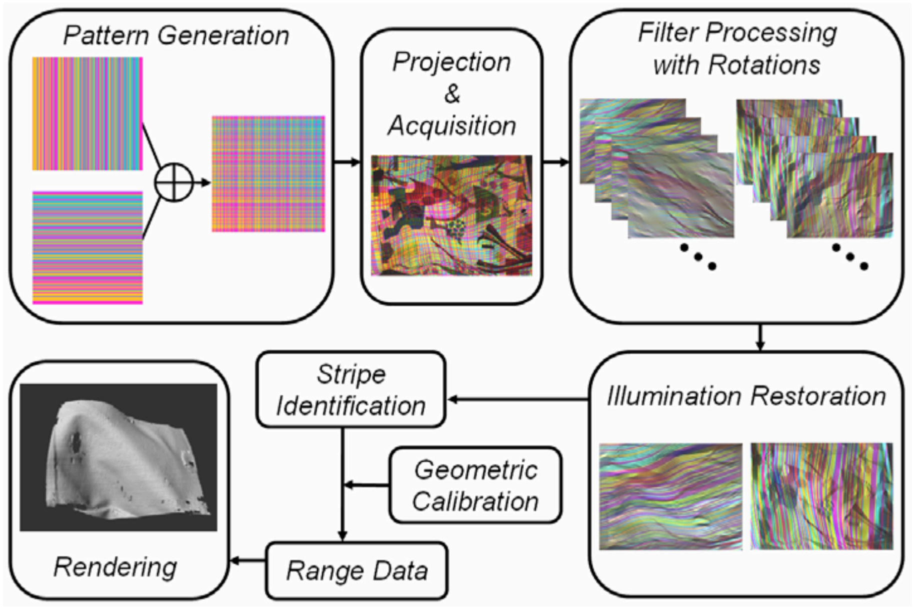


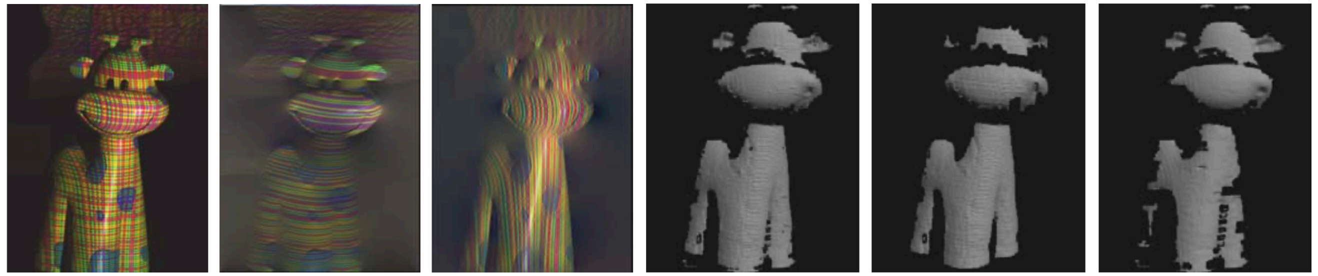






Color-Stripe Structured Light Robust to Surface Color and Discontinuity
(Kwang Hee Lee Changsoo Je Sang Wook Lee, 2007)





Color-Stripe Structured Light Robust to Surface Color and Discontinuity
(Kwang Hee Lee Changsoo Je Sang Wook Lee, 2007)

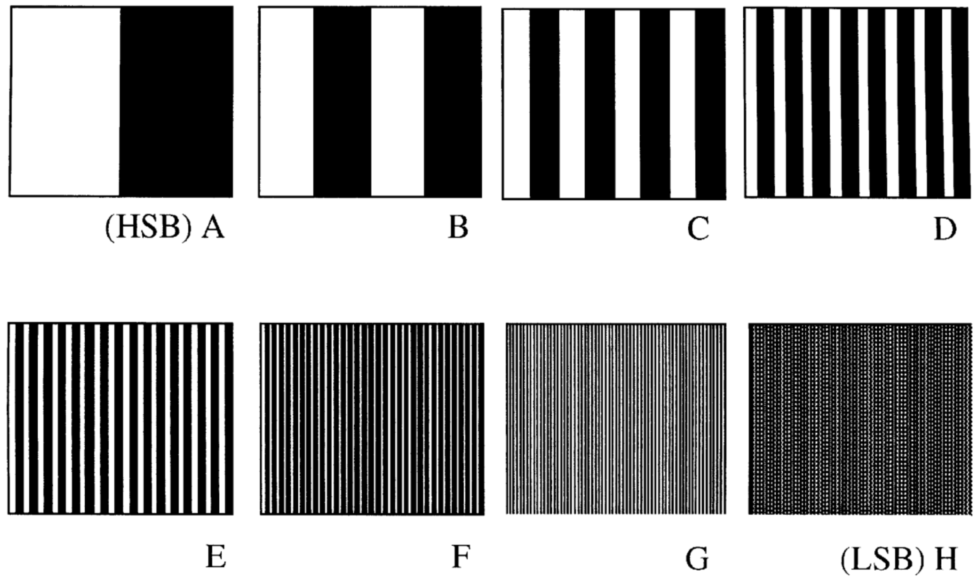


Fig. 5. 8 bits temporally binary-coded pattern projection.

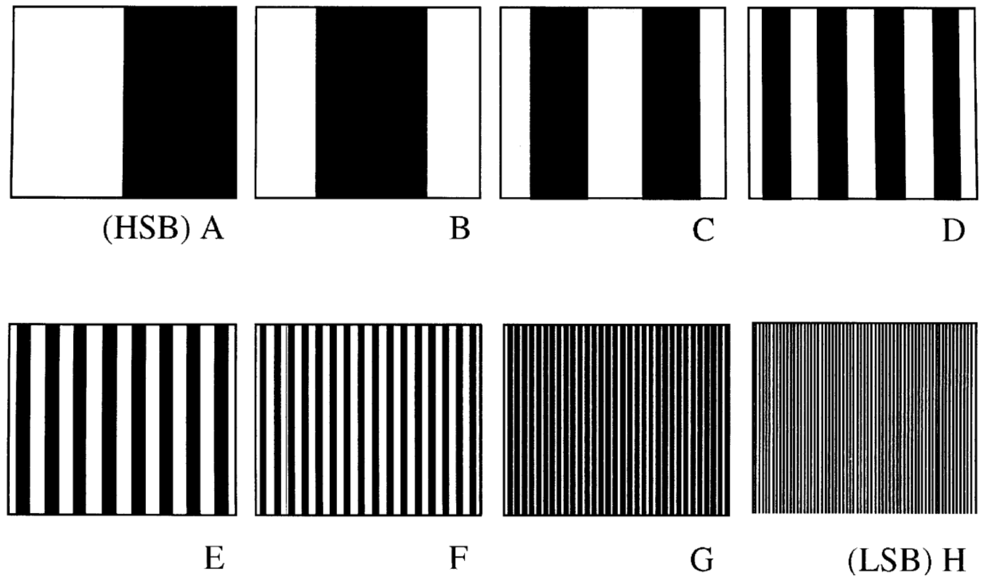
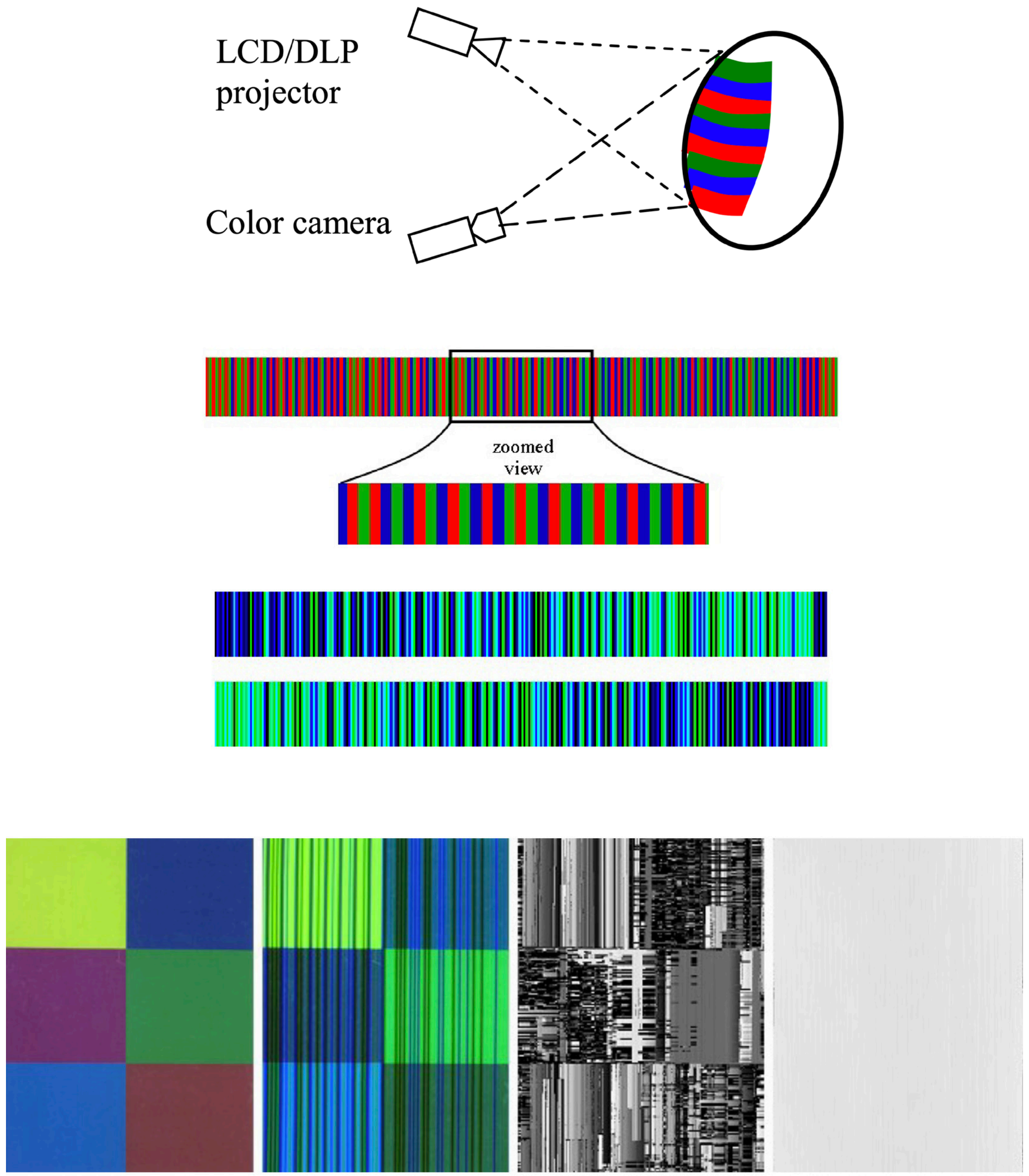
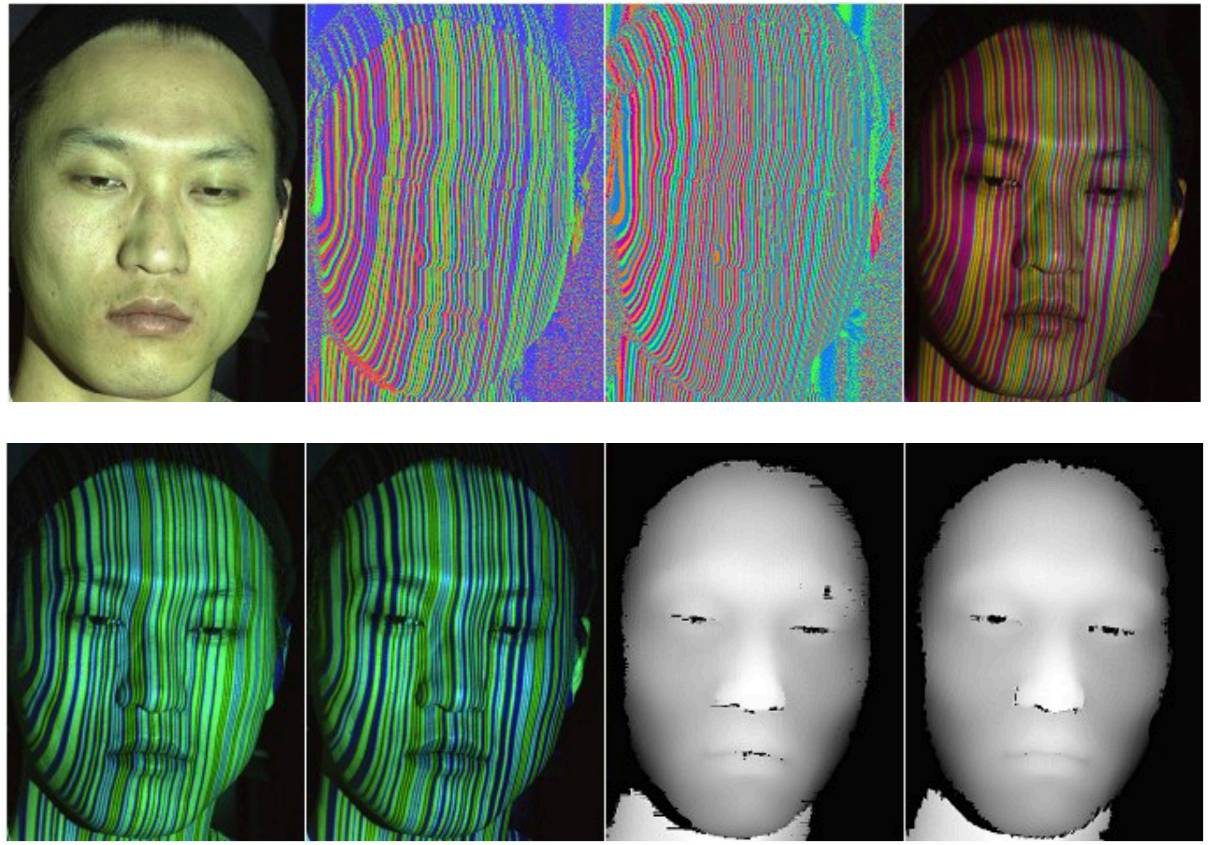


Fig. 6. 8 bits temporally Gray-coded pattern projection.

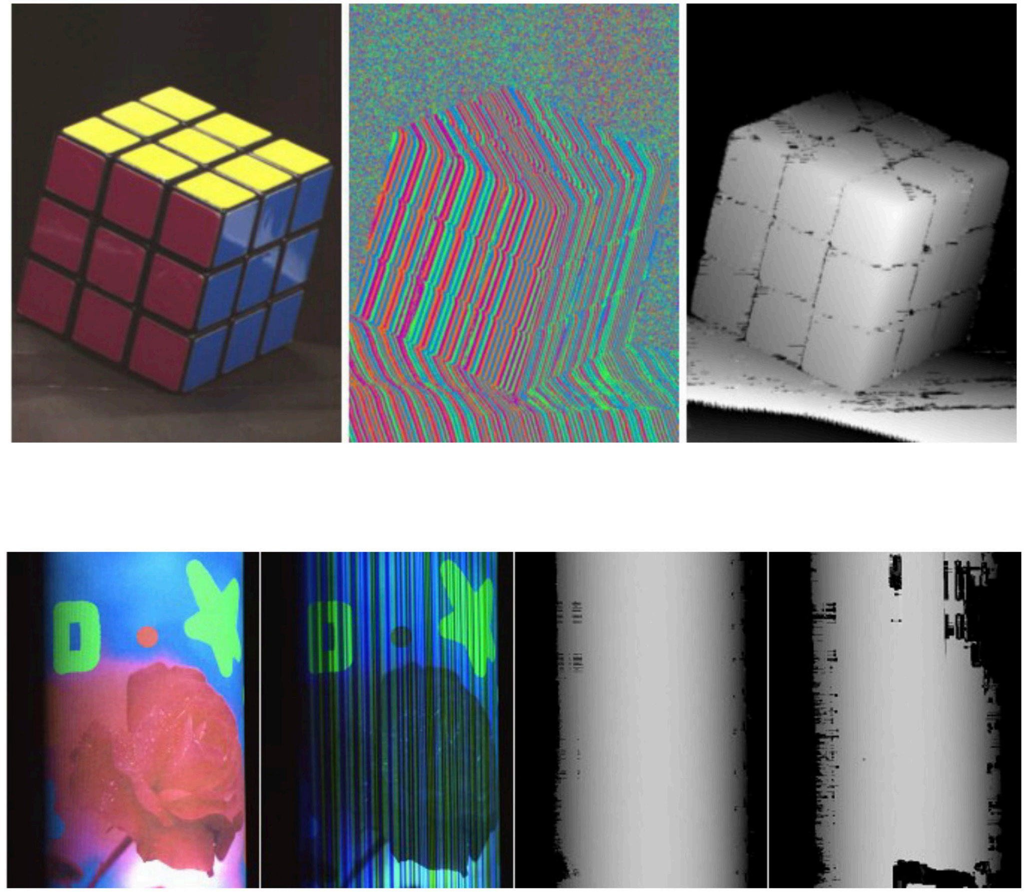
Recent progress in coded structured light as a technique to solve the correspondence Problem: a survey (J. BATLLE, E. MOUADDIB and J. SALVI, 1996)



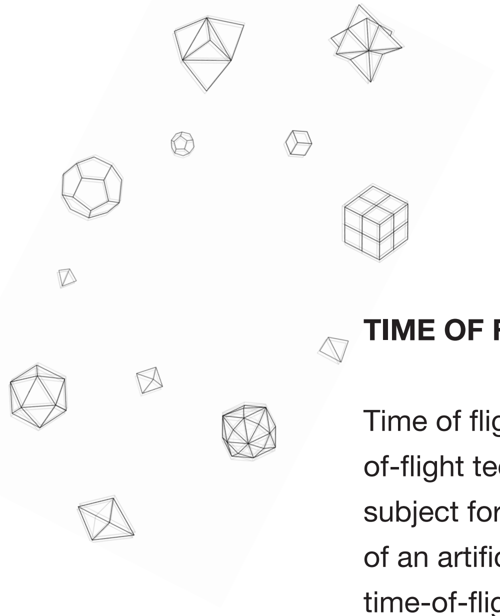
High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging (Changsoo Je, Sang Wook Lee and Rae-Hong Park, 2004)



High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging
(Changsoo Je, Sang Wook Lee and Rae-Hong Park, 2004)

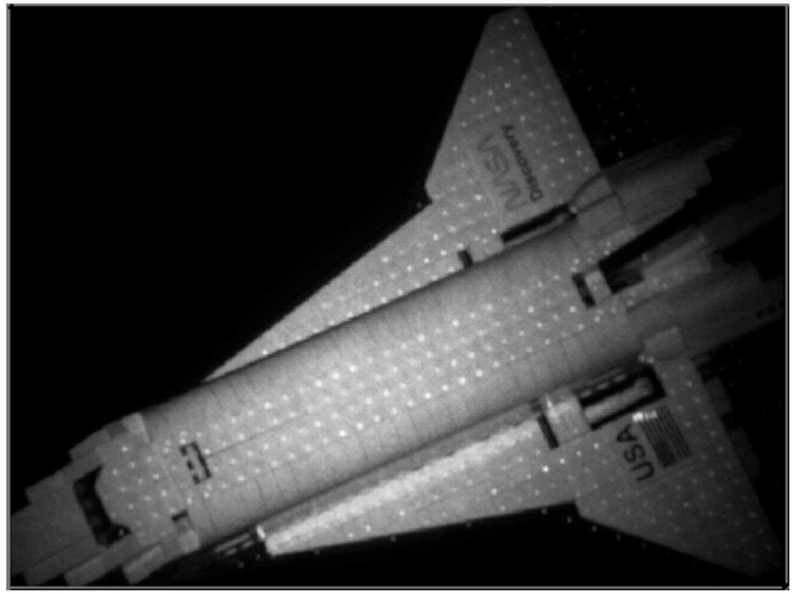


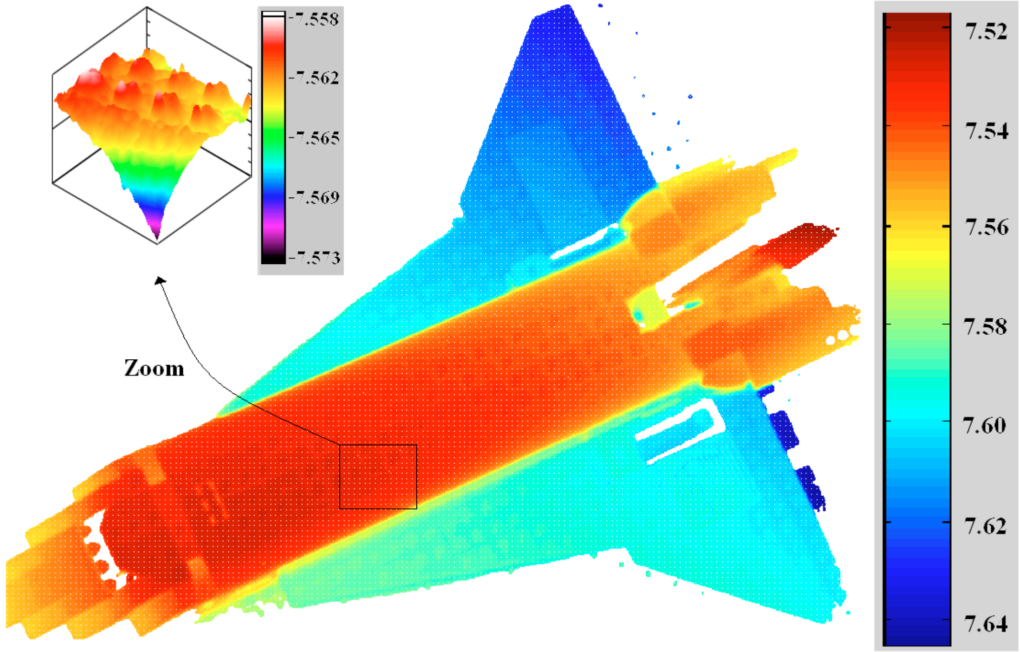
High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging
(Changsoo Je, Sang Wook Lee and Rae-Hong Park, 2004)

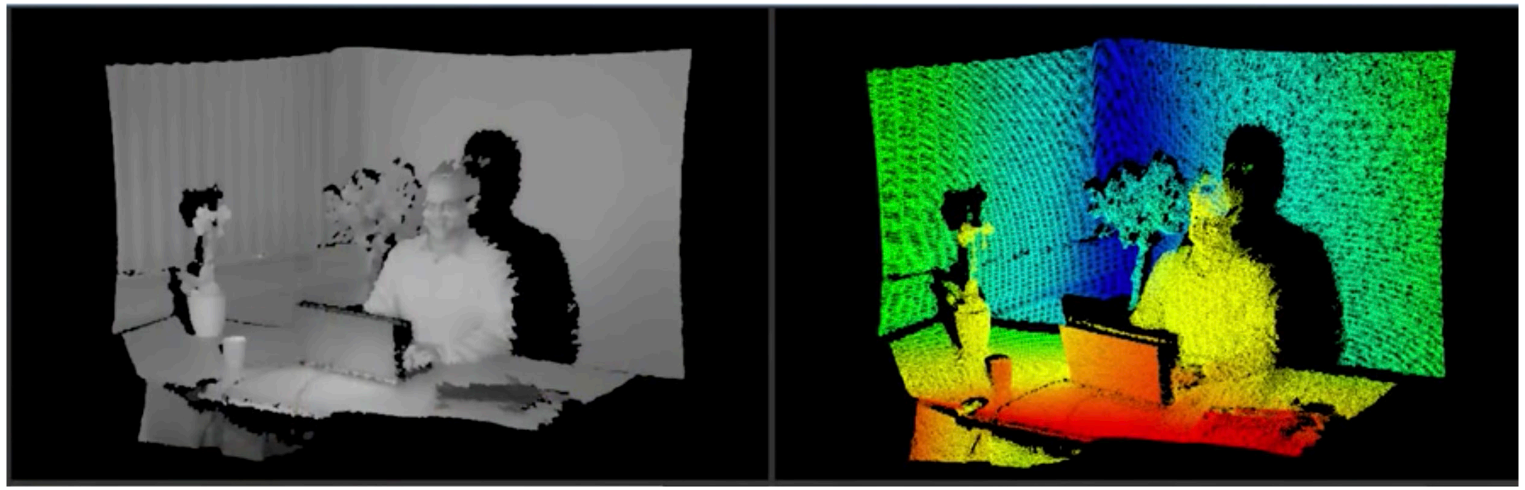


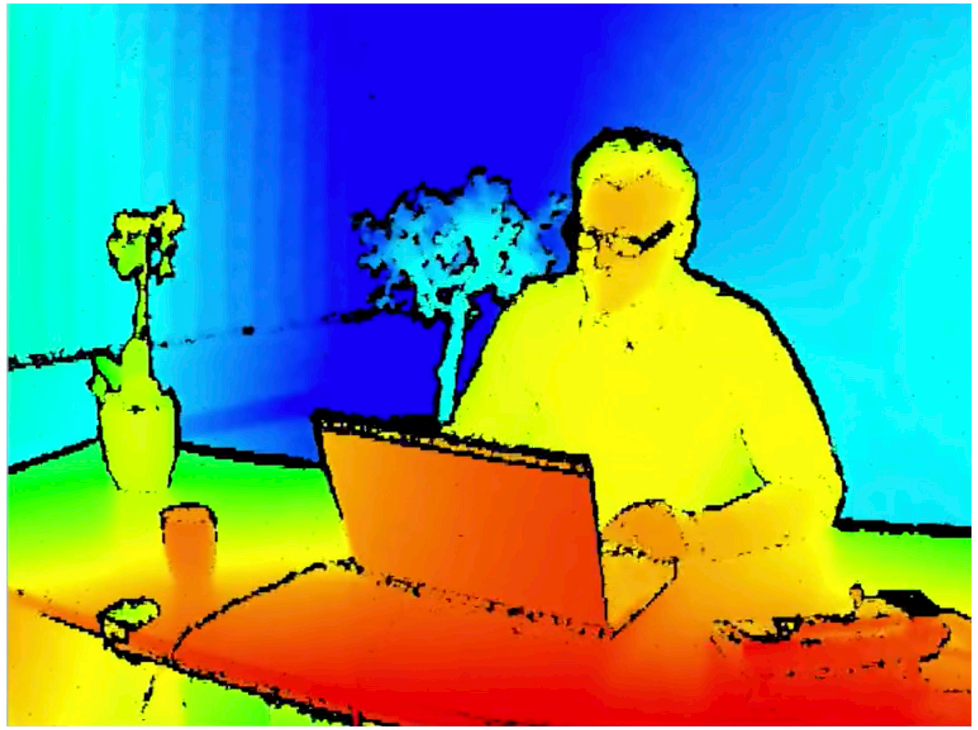
TIME OF FLIGHT (ToF)

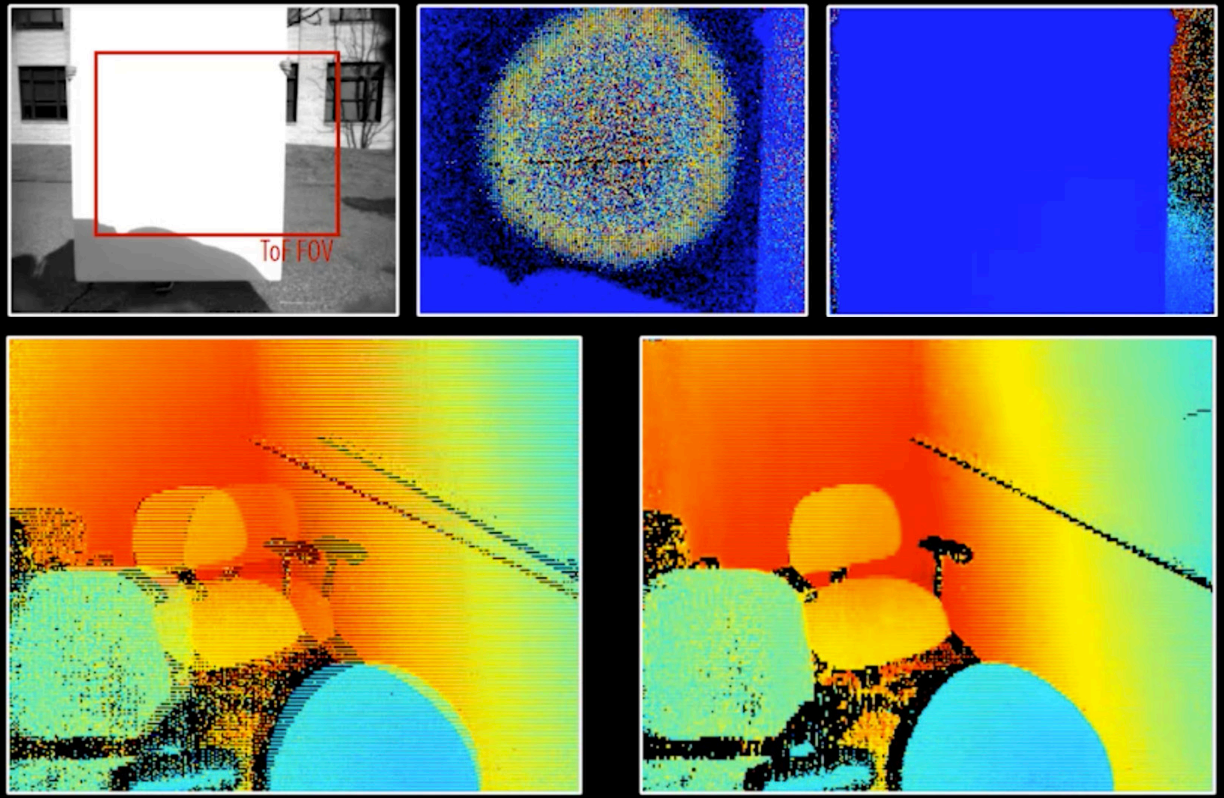
Time of flight — is a range imaging camera system that employs time-of-flight techniques to resolve distance between the camera and the subject for each point of the image, by measuring the round trip time of an artificial light signal provided by a laser or an LED. Laser-based time-of-flight cameras are part of a broader class of scannerless LIDAR, in which the entire scene is captured with each laser pulse, as opposed to point-by-point with a laser beam such as in scanning LIDAR systems. Time-of-flight camera products for civil applications began to emerge around 2000, as the semiconductor processes allowed the production of components fast enough for such devices. The systems cover ranges of a few centimeters up to several kilometers. The distance resolution is about 1 cm. The spatial resolution of time-of-flight cameras is generally low compared to standard 2D video cameras, with most commercially available devices at 320 × 240 pixels or less as of 2011. Compared to other 3D laser scanning methods for capturing 3D images, TOF cameras operate more quickly by providing up to 160 operations per second.

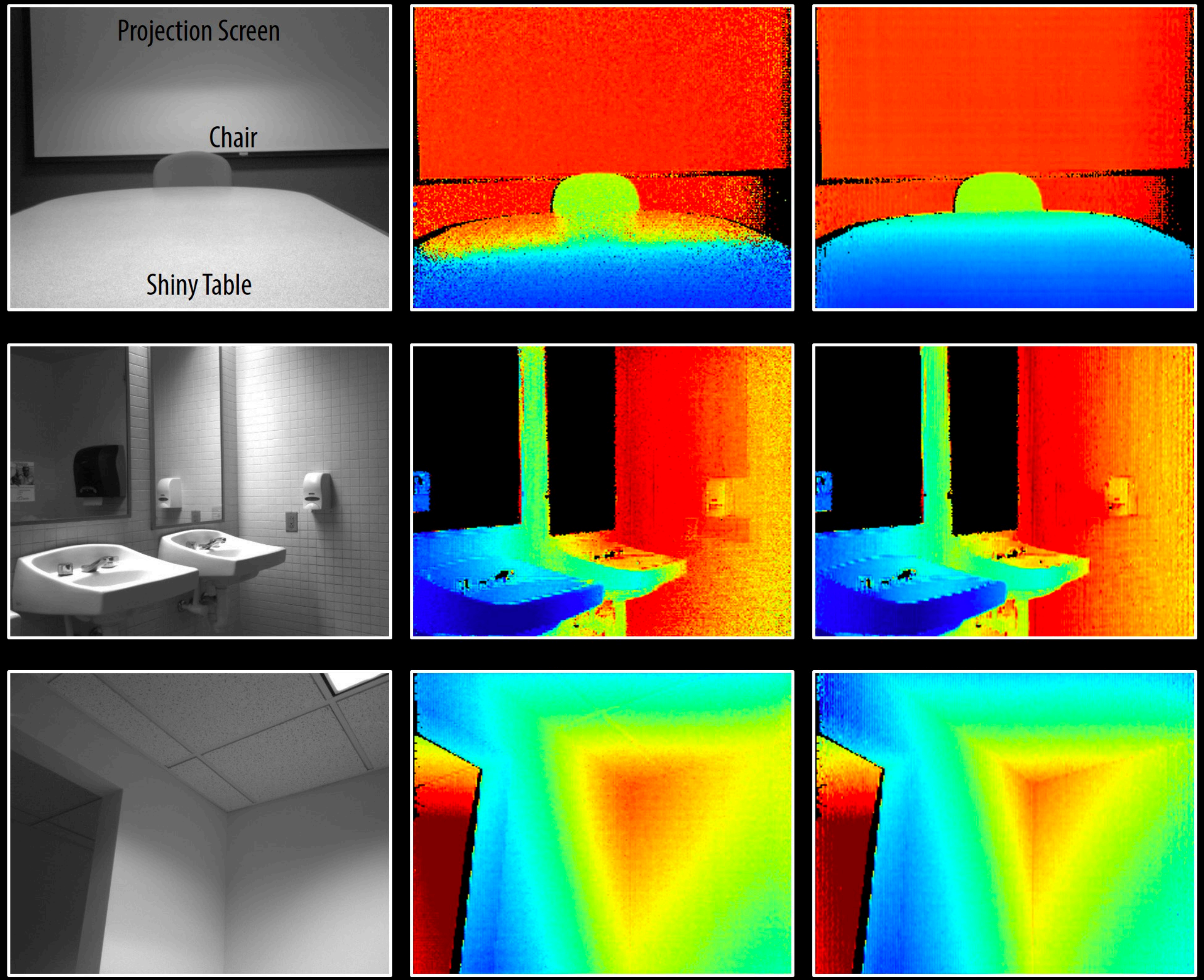



7.558
7.562
7.565
7.569
7.573
Zoom
7.52
7.54
7.56
7.58
7.60
7.62
7.64



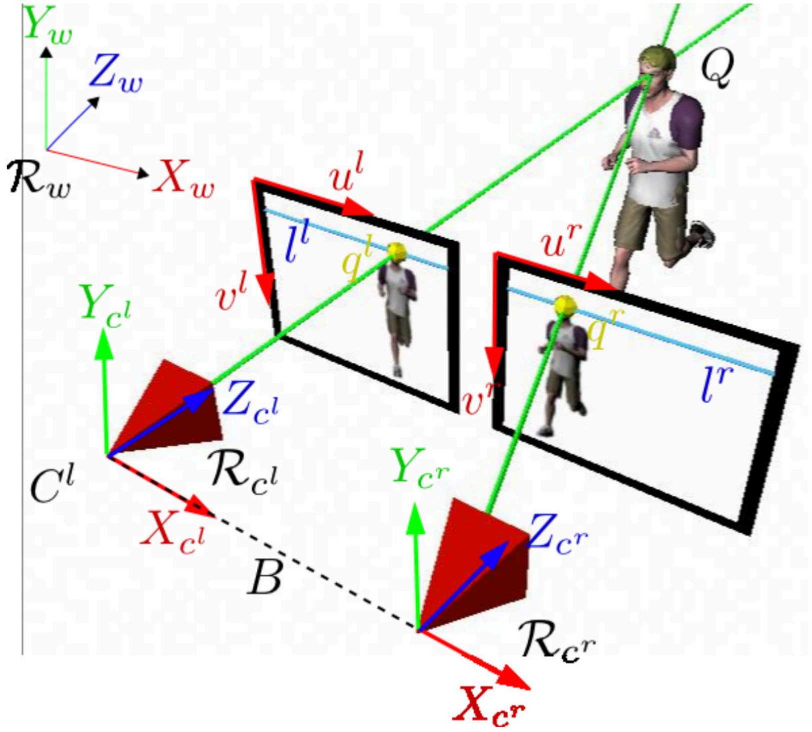


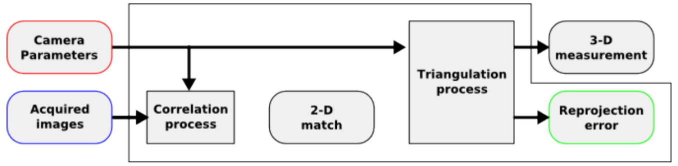

ToF FOV


Projection Screen
Chair
Shiny Table

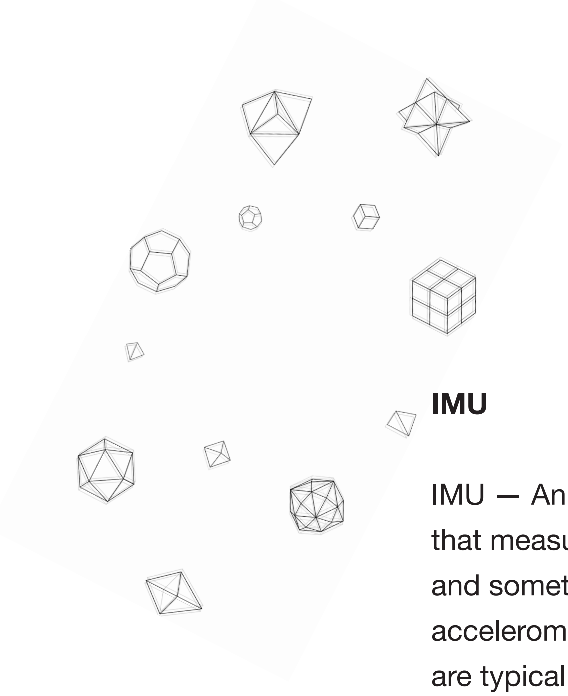
STEREO VISION

Stereo vision — is the extraction of 3D information from digital images, such as those obtained by a CCD camera. By comparing information about a scene from two vantage points, 3D information can be extracted by examining the relative positions of objects in the two panels. This is similar to the biological process Stereopsis. Stereoscopic images are often stored as MPO (multi picture object) files. Recently, researchers pushed to develop methods to reduce the storage needed for these files in order to maintain the high quality of the stereo image.
In traditional stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene, in a manner similar to human binocular vision. By comparing these two images, the relative depth information can be obtained in the form of a disparity map, which encodes the difference in horizontal coordinates of corresponding image points. The values in this disparity map are inversely proportional to the scene depth at the corresponding pixel location.
For a human to compare the two images, they must be superimposed in a stereoscopic device, with the image from the right camera being shown to the observer's right eye and from the left one to the left eye.





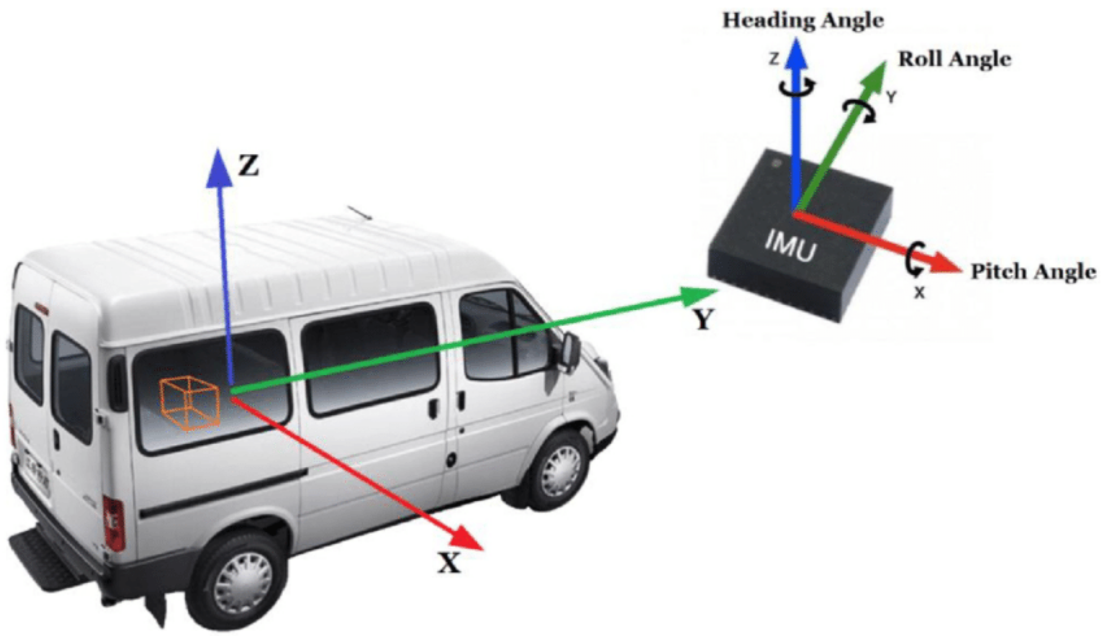
Stereo-vision sensor in standart configuration. Data flow diagram for the stereo-vision process.
(Evaluation Method for Automotive Stereo-Vision Systems, J. Morat, F. Devernay, J. Ibanez-Guzman, and S. Cornou, 2007)

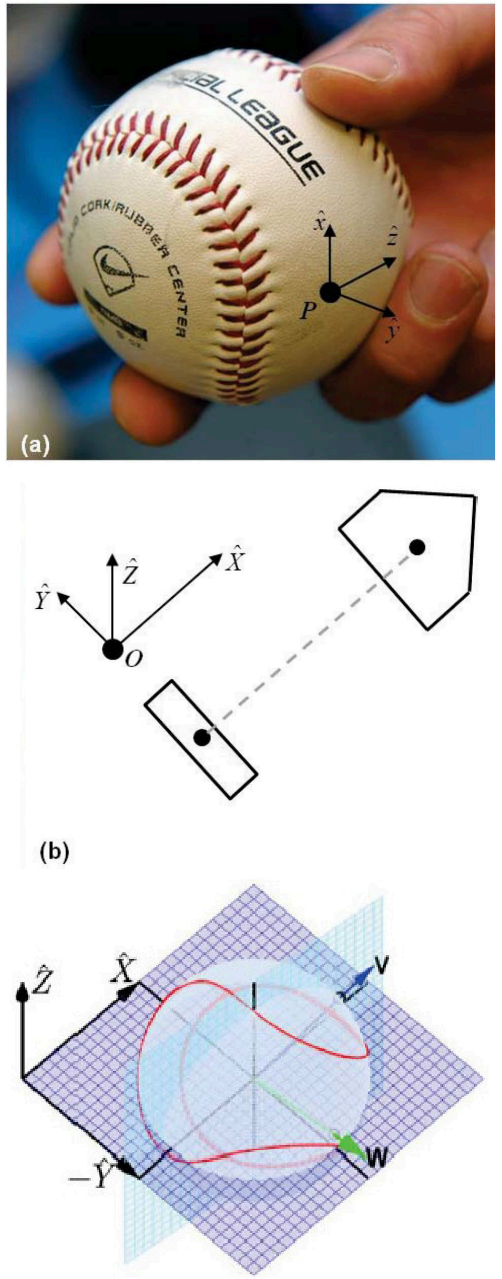


IMU

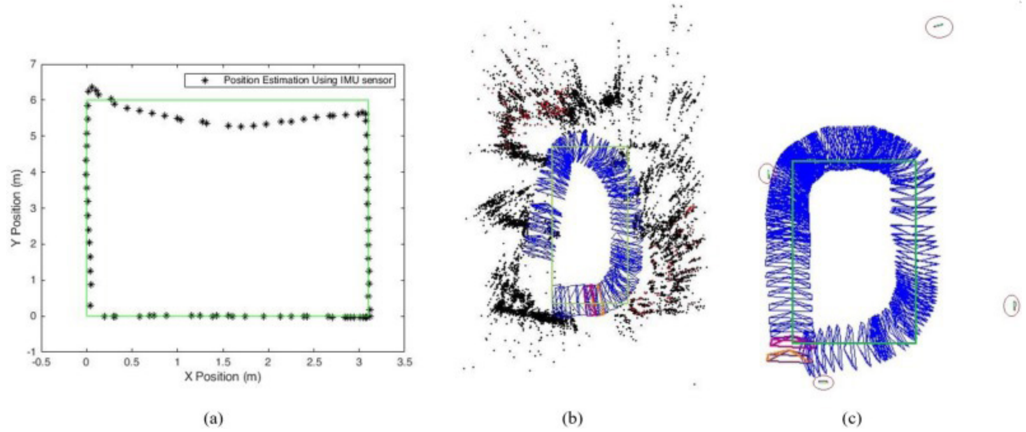
IMU — An inertial measurement unit (IMU) is an electronic device that measures and reports a body's specific force, angular rate, and sometimes the orientation of the body, using a combination of accelerometers, gyroscopes, and sometimes magnetometers. IMUs are typically used to maneuver aircraft (an attitude and heading reference system), including unmanned aerial vehicles (UAVs), among many others, and spacecraft, including satellites and landers. Recent developments allow for the production of IMU-enabled GPS devices. An IMU allows a GPS receiver to work when GPS-signals are unavailable, such as in tunnels, inside buildings, or when electronic interference is present. A wireless IMU is known as a WIMU.

An inertial measurement unit works by detecting linear acceleration using one or more accelerometers and rotational rate using one or more gyroscopes. Some also include a magnetometer which is commonly used as a heading reference. Typical configurations contain one accelerometer, gyro, and magnetometer per axis for each of the three vehicle axes: pitch, roll and yaw.

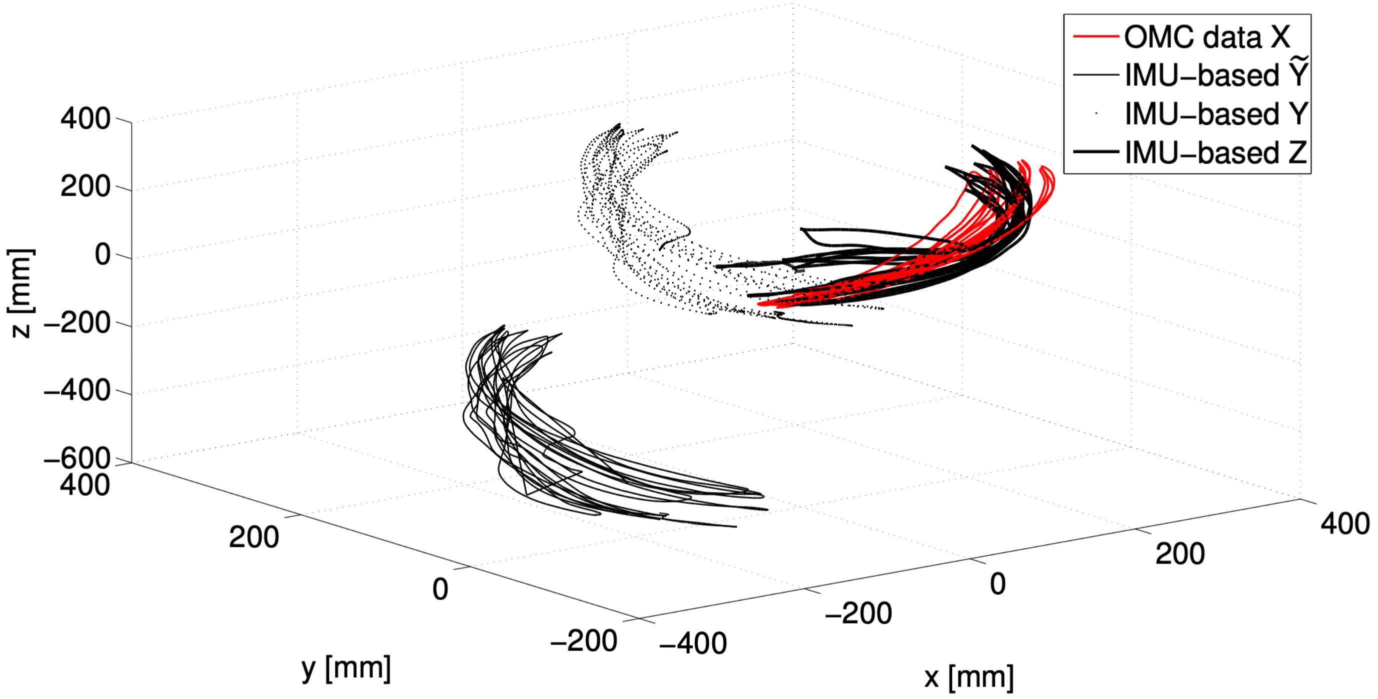




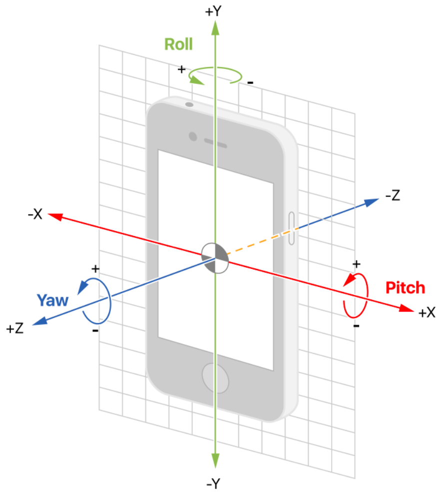
IMU for Vessel and Offshore Piping Survey
(Ahmed Islam, Syed Adeel Ahmed, Brandon Taravella)

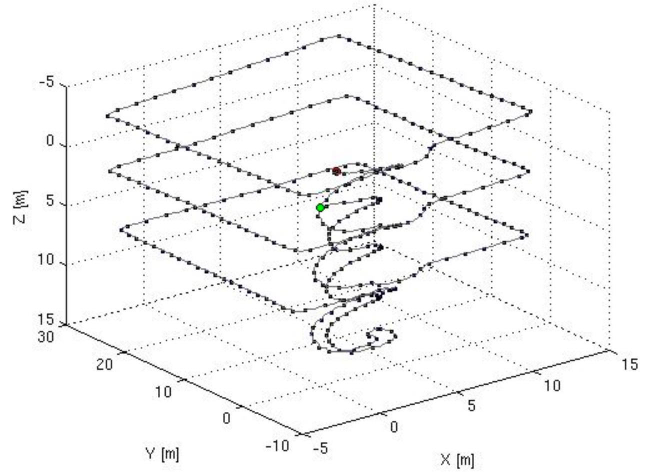


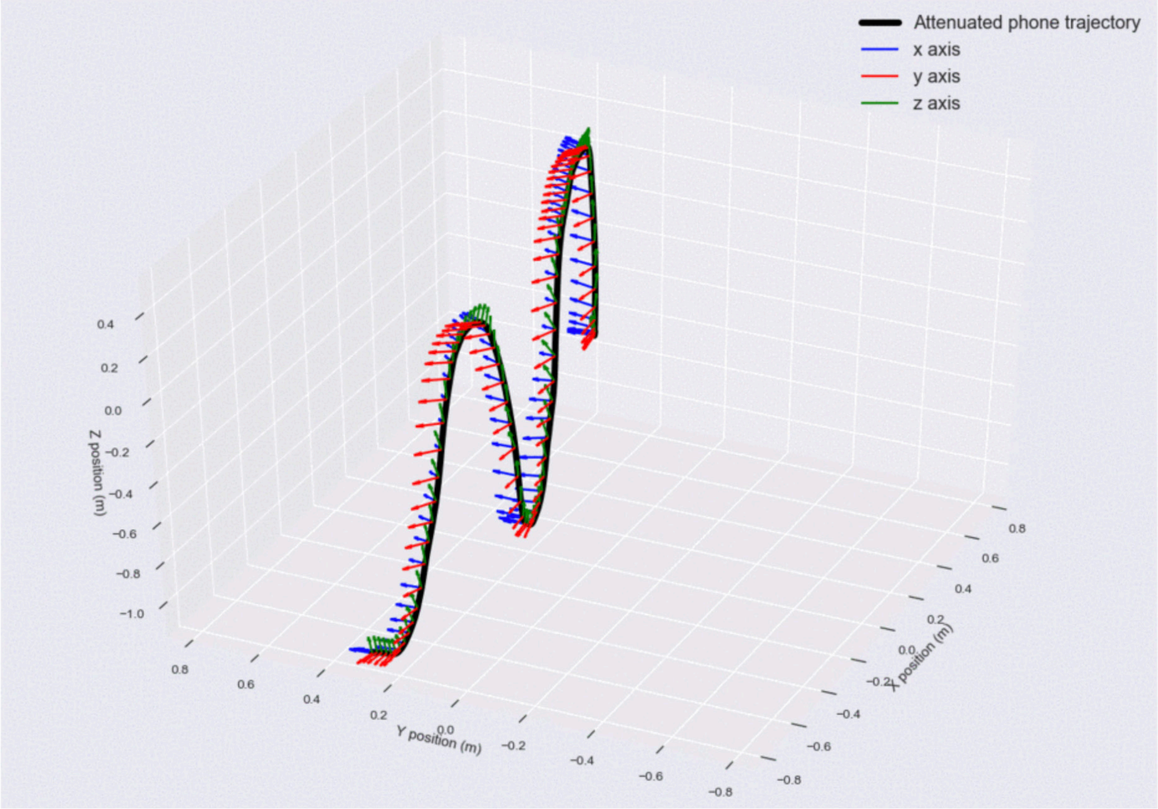
Hybrid Indoor Localization Using IMU Sensors and Smartphone Camera



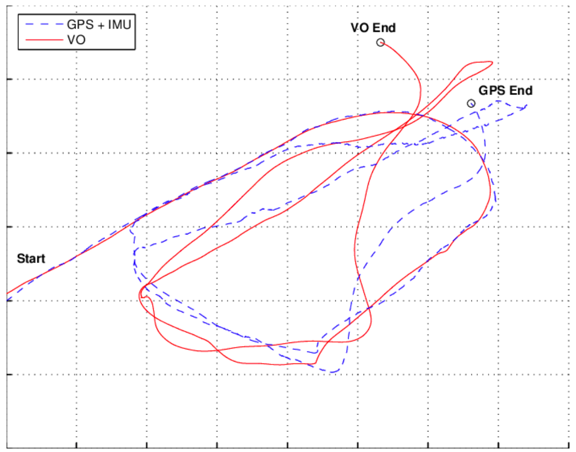
Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion
(Alessandro Filippeschi, Norbert Schmitz, Markus Miezal, Gabriele Bleser, Emanuele Ruffaldi and Didier Stricker)

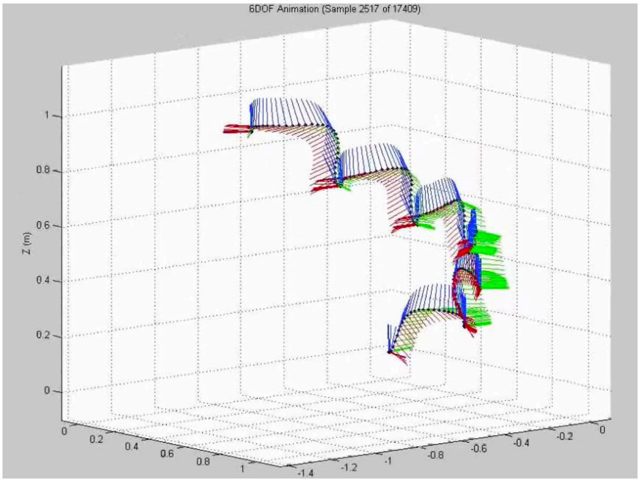


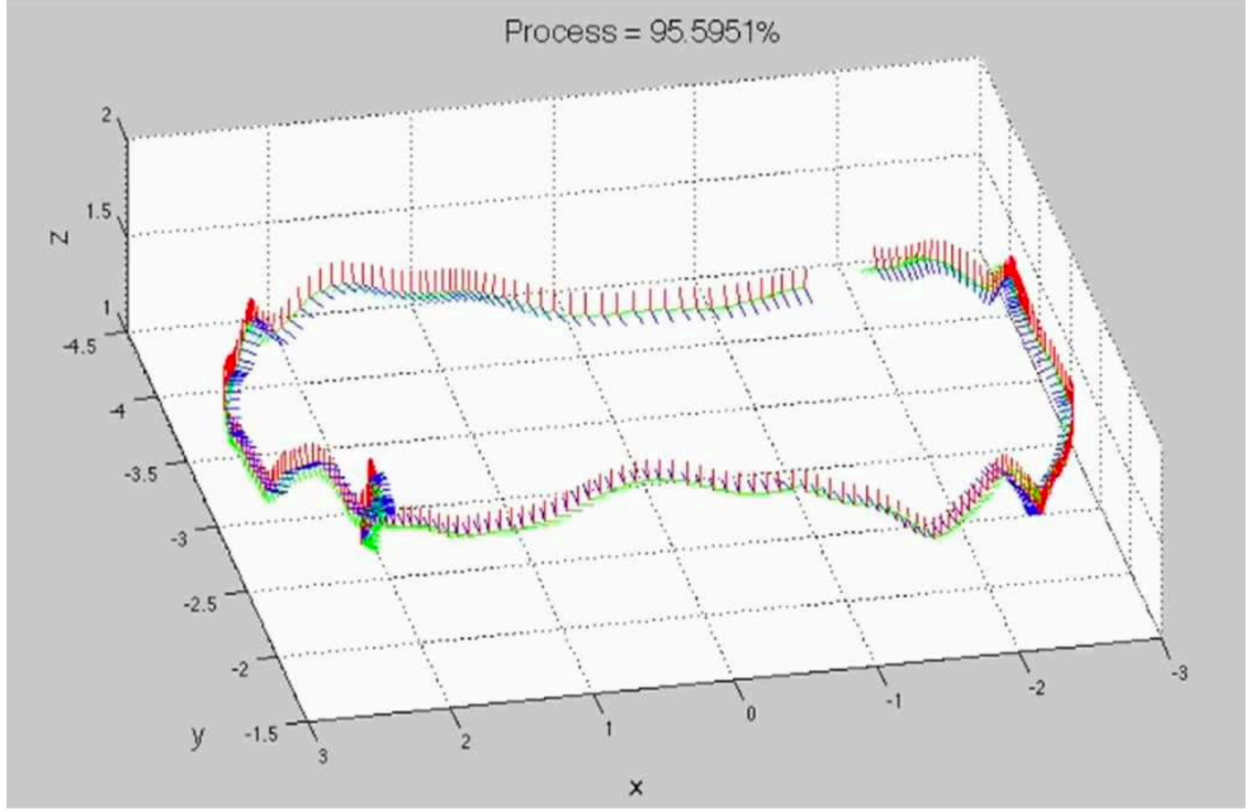




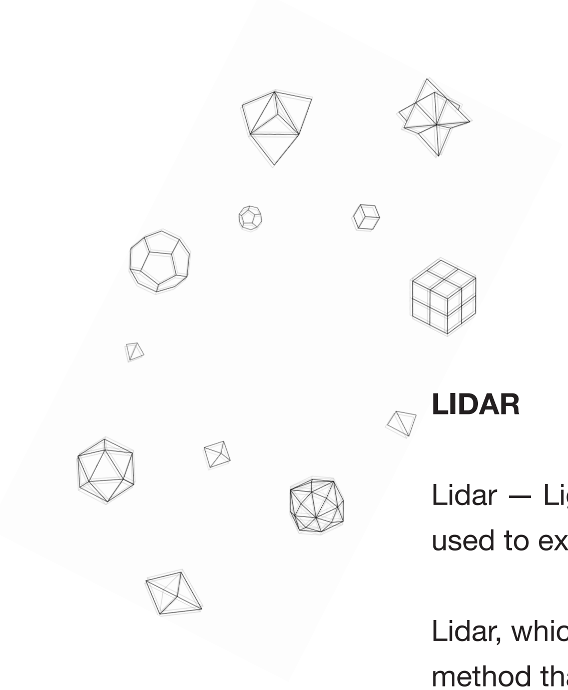
3D tracking with IMU







2D-3D tracking with IMU



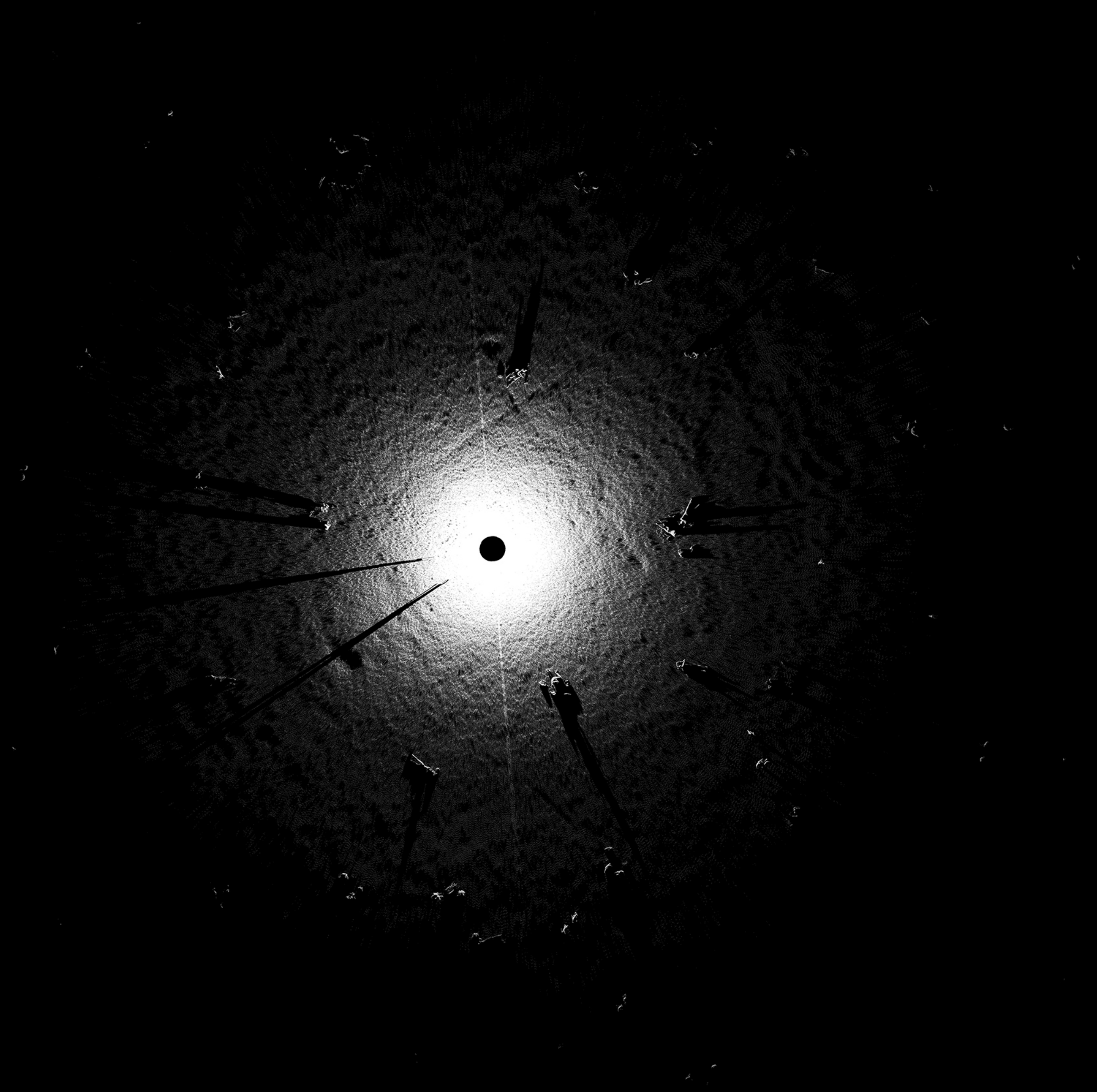
LIDAR

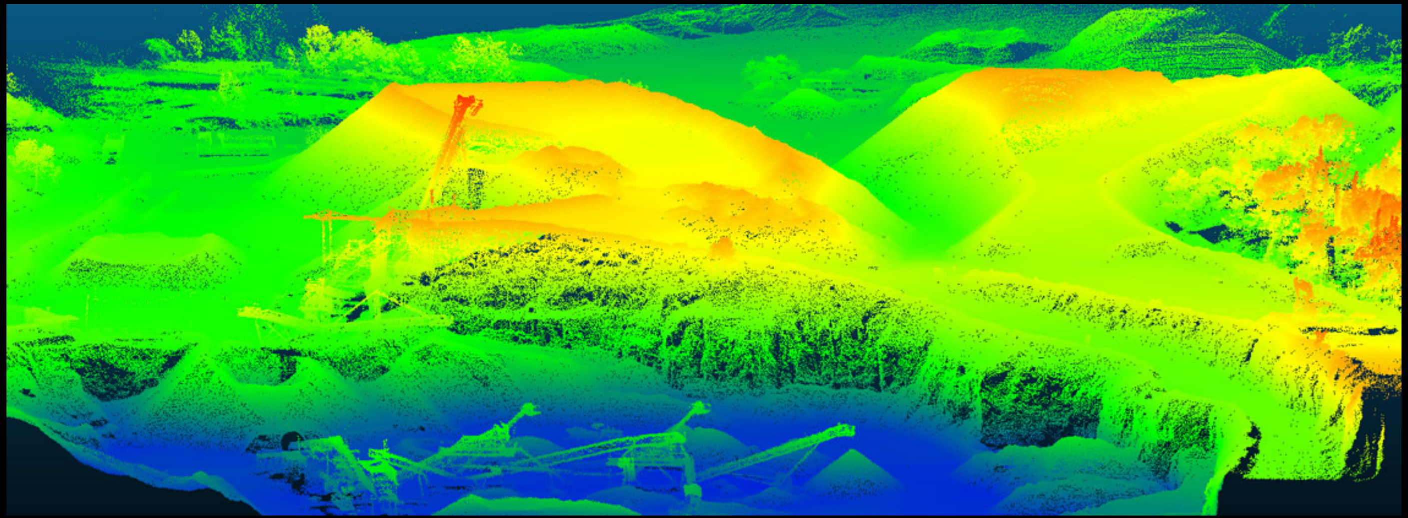
Lidar — Light Detection and Ranging — is a remote sensing method used to examine the surface of the Earth.

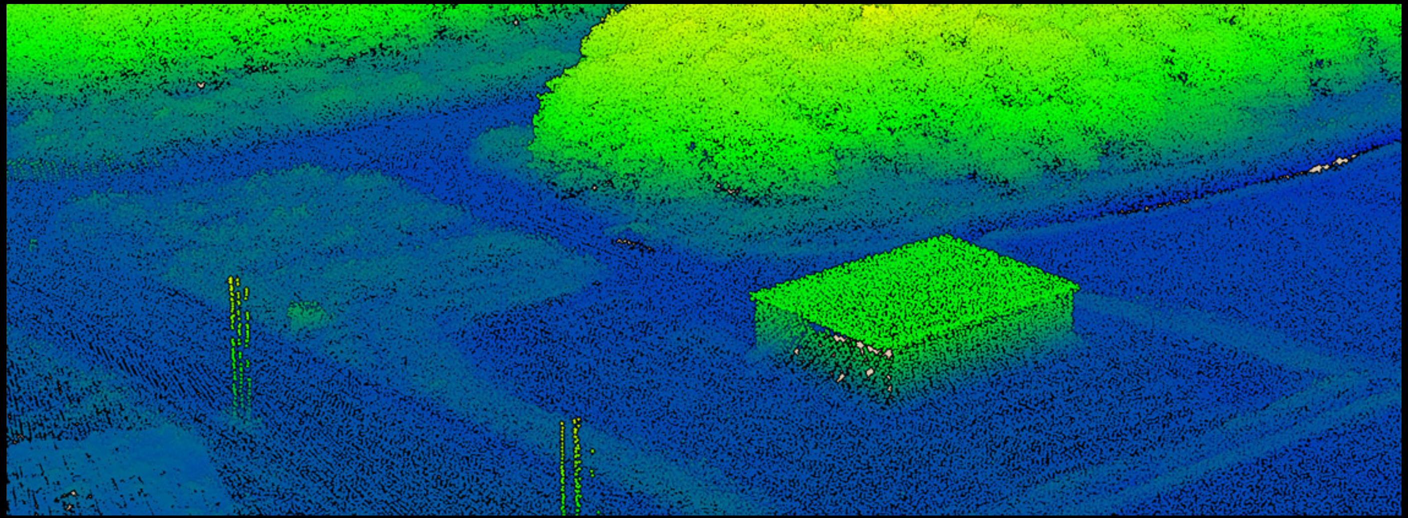
Lidar, which stands for Light Detection and Ranging, is a remote sensing method that uses light in the form of a pulsed laser to measure ranges (variable distances) to the Earth. These light pulses—combined with other data recorded by the airborne system— generate precise, three-dimensional information about the shape of the Earth and its surface characteristics.

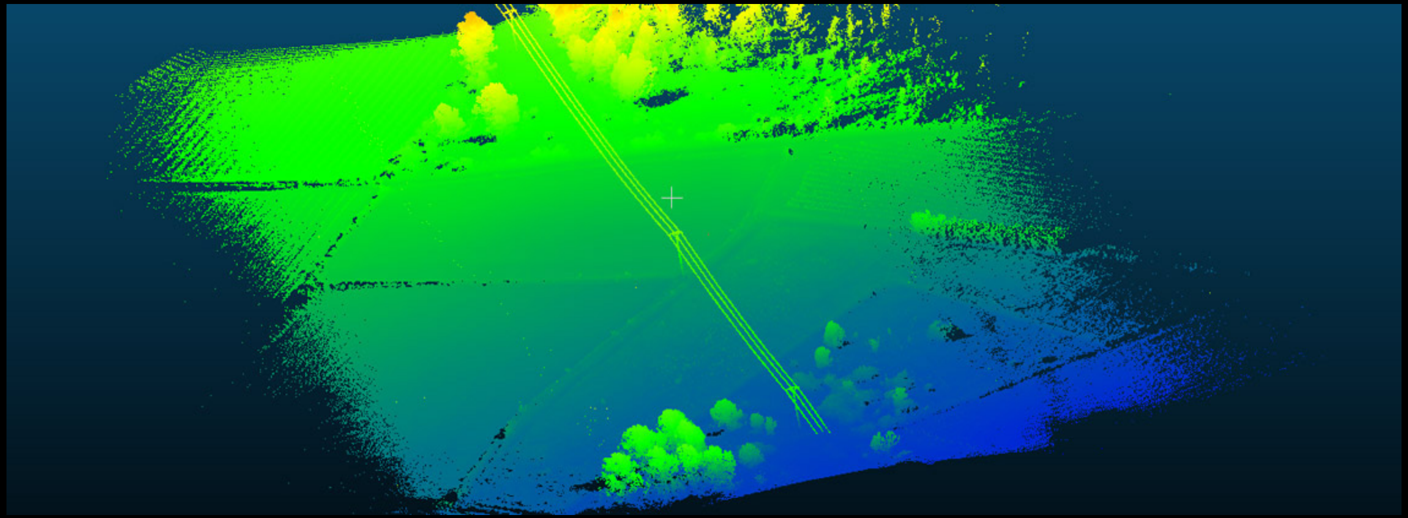
A lidar instrument principally consists of a laser, a scanner, and a specialized GPS receiver. Airplanes and helicopters are the most commonly used platforms for acquiring lidar data over broad areas. Two types of lidar are topographic and bathymetric. Topographic lidar typically uses a near-infrared laser to map the land, while bathymetric lidar uses water-penetrating green light to also measure seafloor and riverbed elevations.

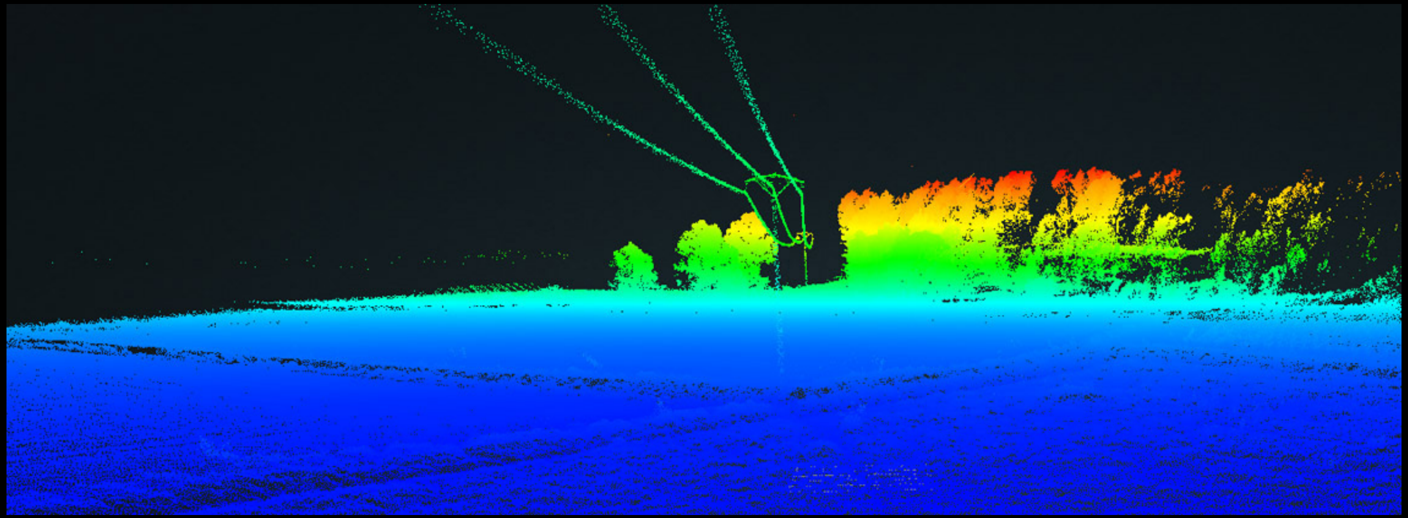
Lidar systems allow scientists and mapping professionals to examine both natural and manmade environments with accuracy, precision, and flexibility. Scientists are using lidar to produce more accurate shoreline maps, make digital elevation models for use in geographic information systems, to assist in emergency response operations, and in many other applications.

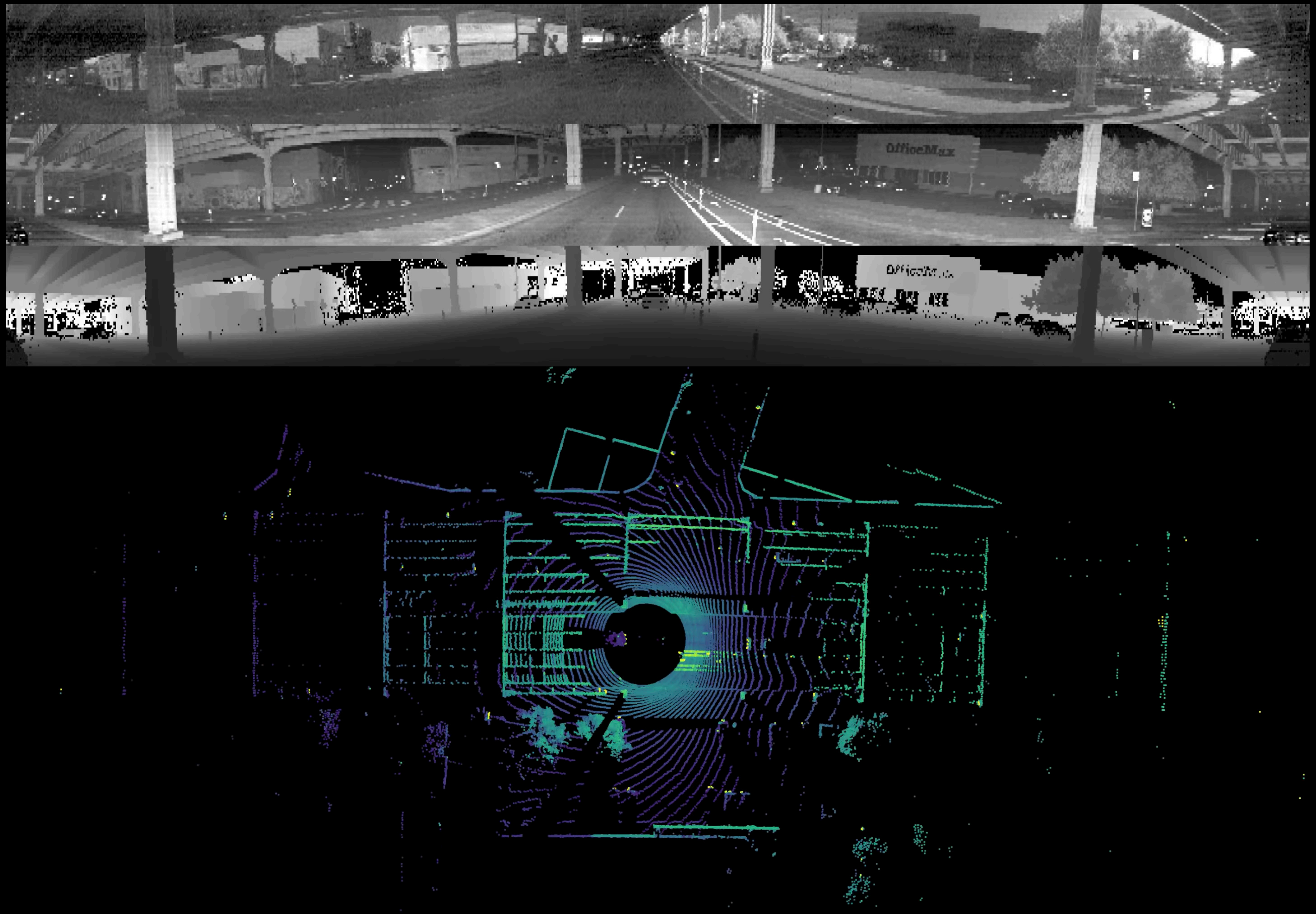


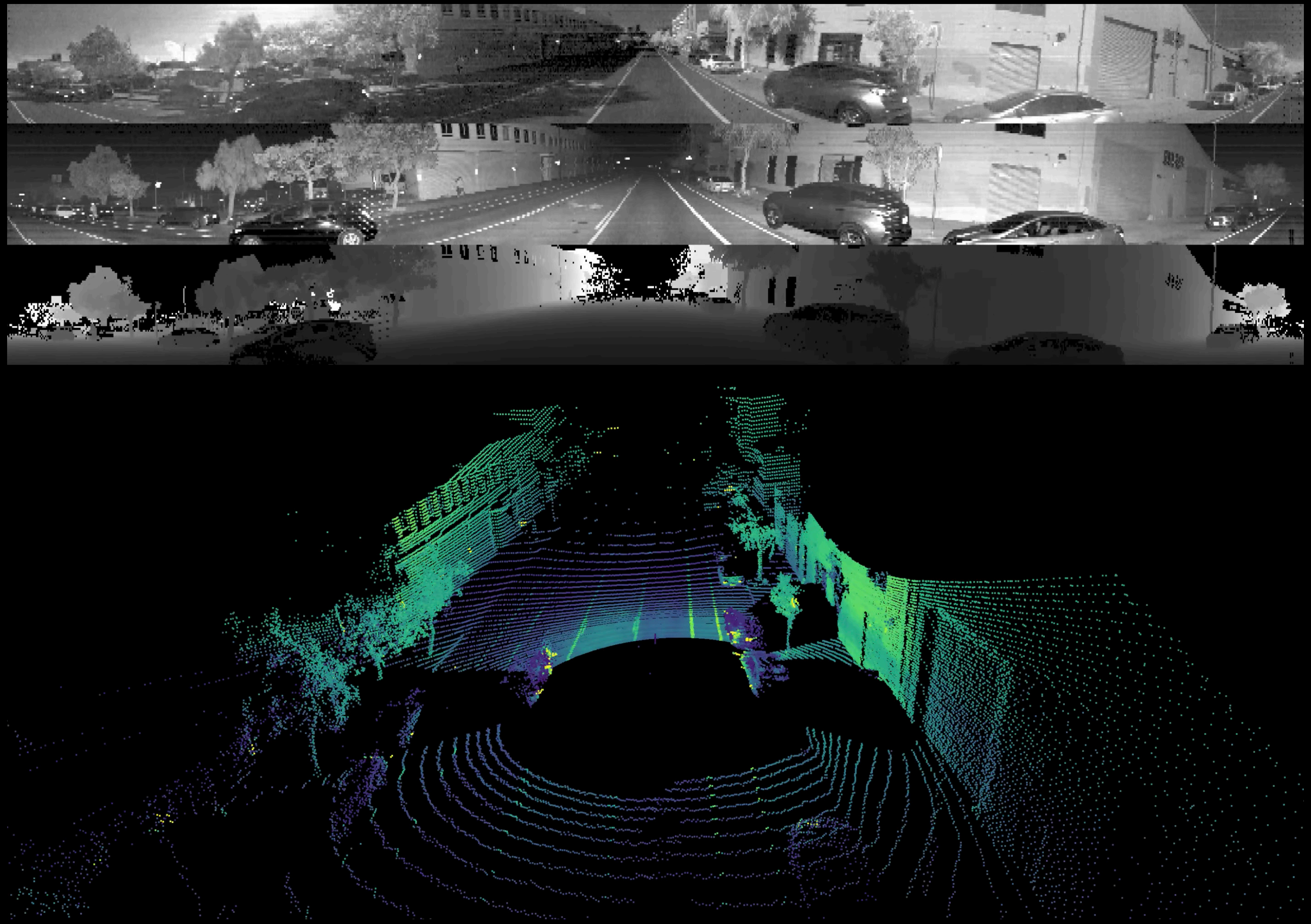




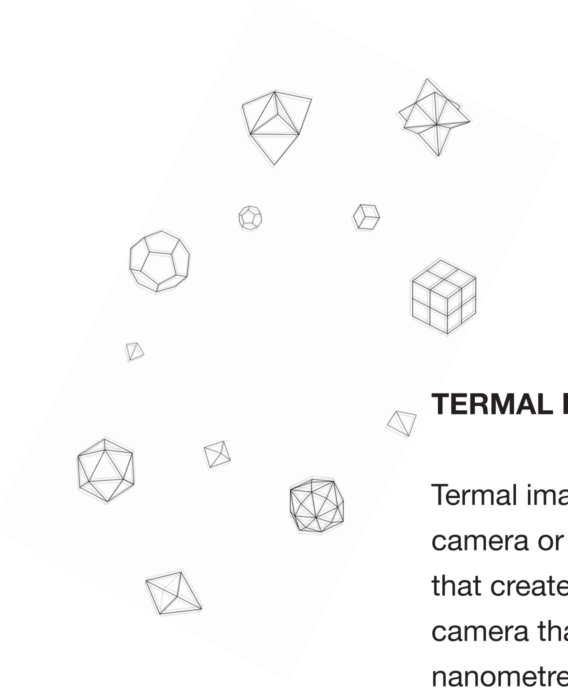






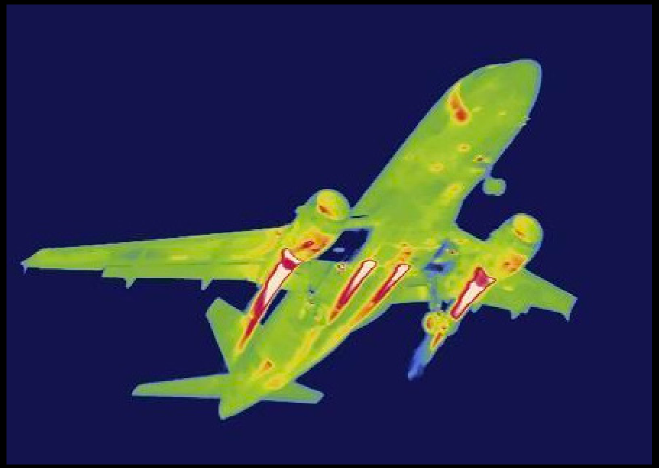
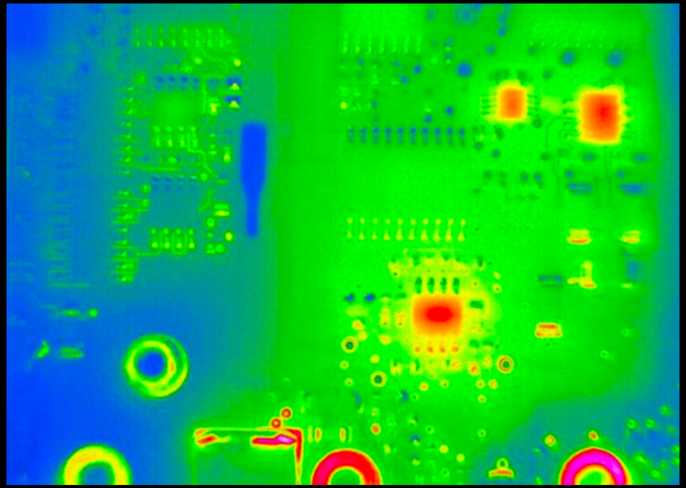


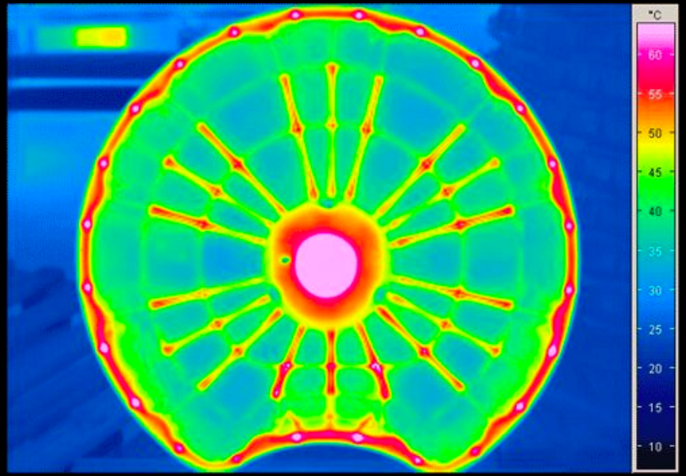
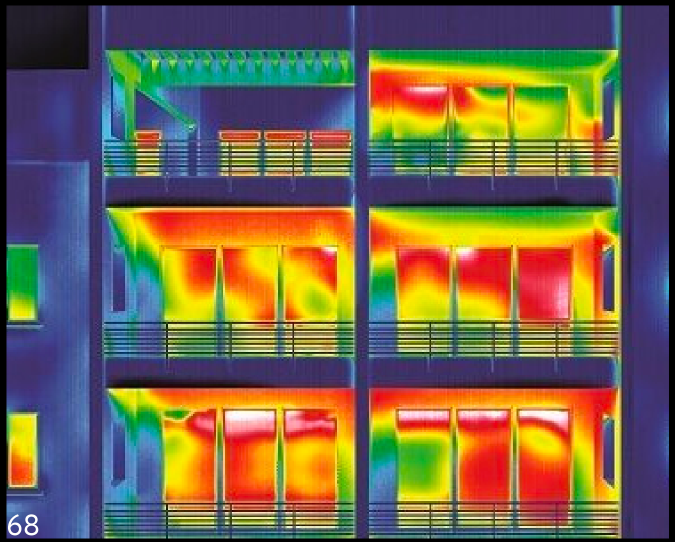
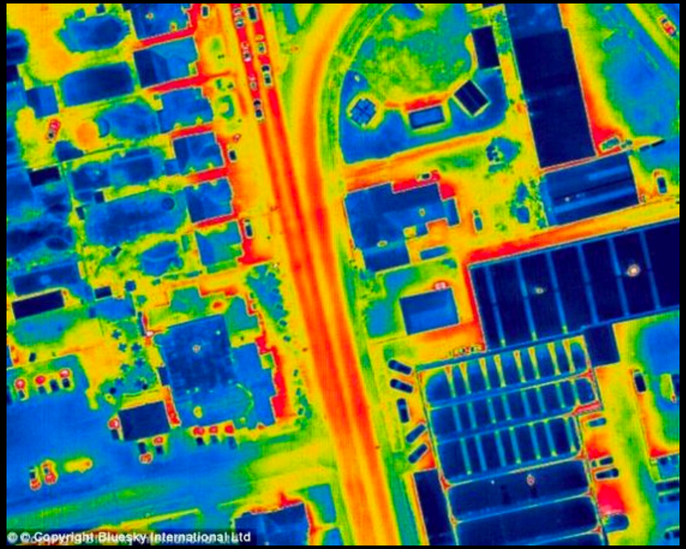
Ouster OS1-64 lidar sensor point cloud and 2D ambient, signal, and range images



TERMAL IMAGER

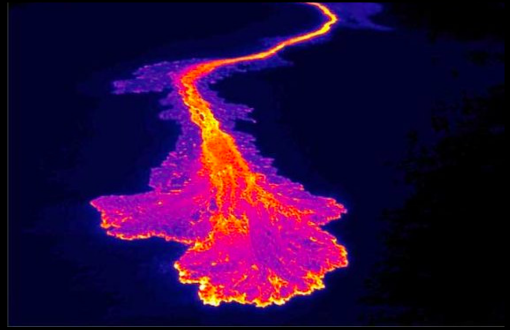
Termal imager — a thermographic camera (also called an infrared camera or thermal imaging camera or thermal imager) is a device that creates an image using infrared radiation, similar to a common camera that forms an image using visible light. Instead of the 400–700 nanometre range of the visible light camera, infrared cameras are sensitive to wavelengths from about 1,000 nm (1 μm) to about 14,000 nm (14 μm). The practice of capturing and analyzing the data they provide is called thermography.

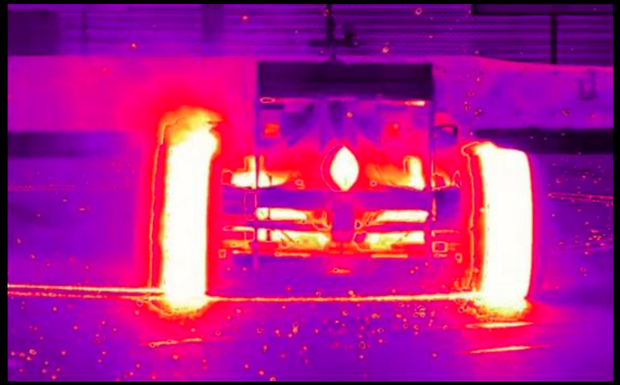
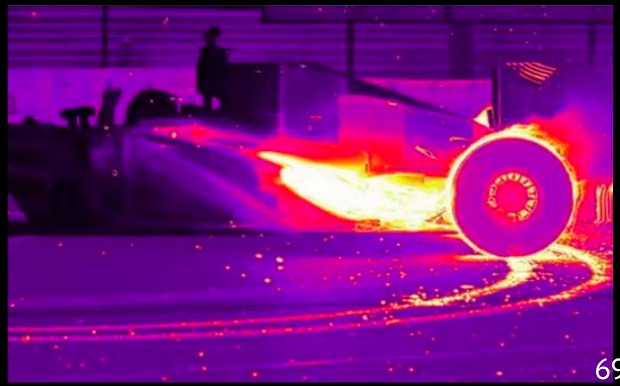



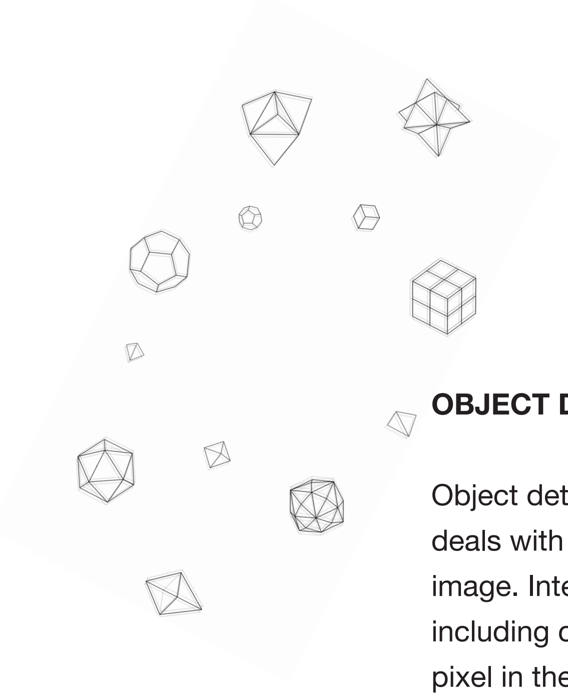

°C
60
55
50
45
40
35
30
25
20
15
10


© Copyright Bluesky International Ltd



36.3℃
36.5℃
36.5℃
36.8℃

FLIR
114
83.6
2012-07-16
3:43 PM
ε
0.95


WWW.TTCIRCUIT.COM



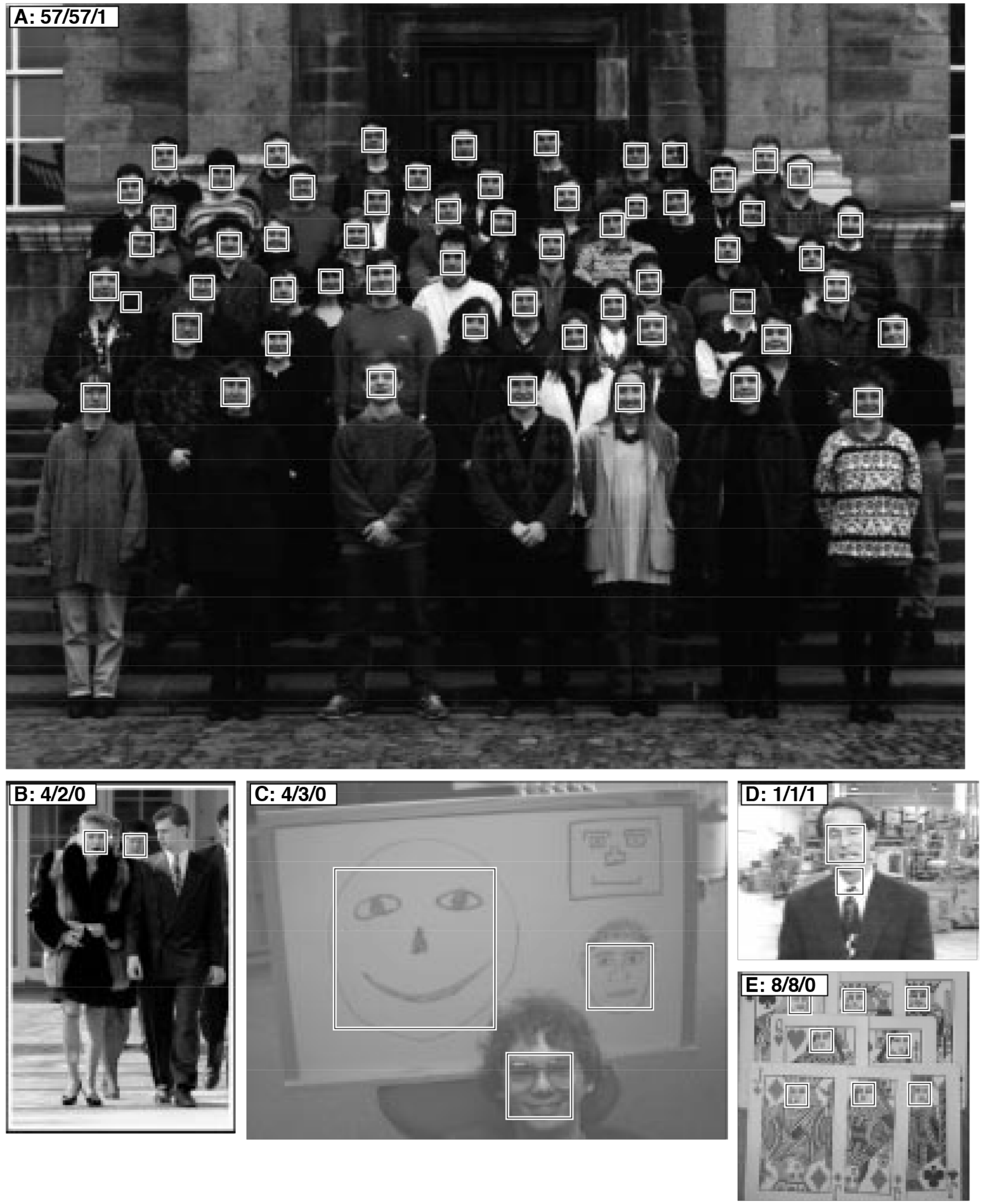


OBJECT DETECTION

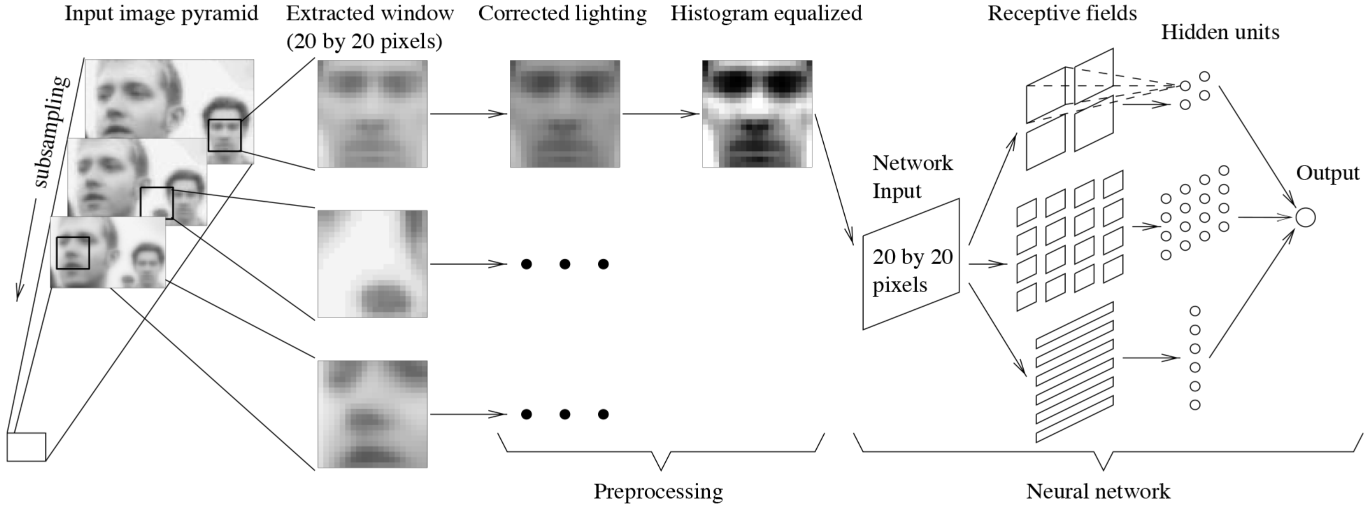
Object detection — is a common Computer Vision problem which deals with identifying and locating object of certain classes in the image. Interpreting the object localisation can be done in various ways, including creating a bounding box around the object or marking every pixel in the image which contains the object (called segmentation). Object detection was studied even before the breakout popularity of CNNs in Computer Vision. While CNNs are capable of automatically extracting more complex and better features, taking a glance at the conventional methods can at worst be a small detour and at best an inspiration.

Object detection before Deep Learning was a several step process, starting with edge detection and feature extraction using techniques like SIFT, HOG etc. These image were then compared with existing object templates, usually at multi scale levels, to detect and localize objects present in the image.

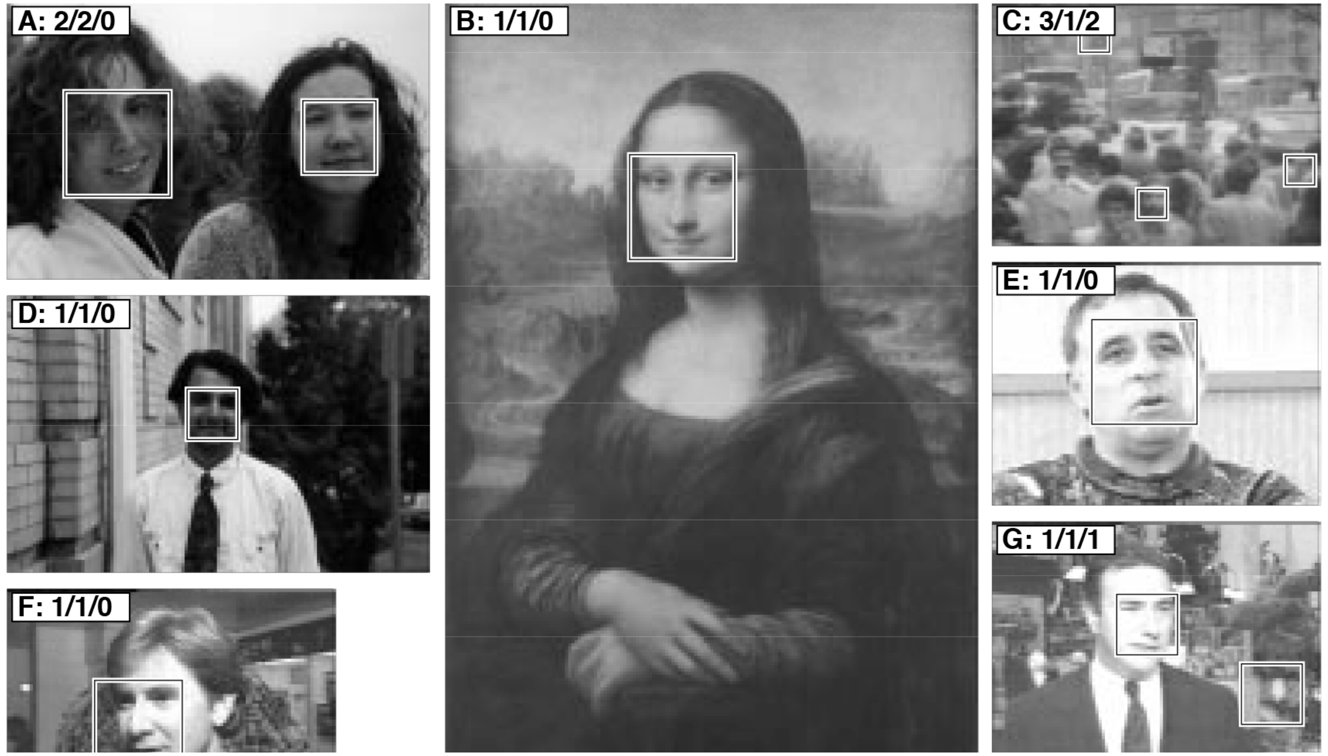
Part-based object detection results for people, bicycles, and horses (Felzenszwalb, McAllester, and Ramanan 2008) c 2008 IEEE. The first three columns show correct detections, while the rightmost column shows false positives.



Face detection results produced by Rowley, Baluja, and Kanade (1998a) c 1998 IEEE.



A neural network for face detection (Rowley, Baluja, and Kanade 1998a) c 1998 IEEE.



Face detection results produced by Rowley, Baluja, and Kanade (1998a) c 1998 IEEE.

Face detection results produced by Rowley, Baluja, and Kanade (1998a) c 1998 IEEE.

Instance (known object) recognition (Lowe 1999) c 1999 IEEE

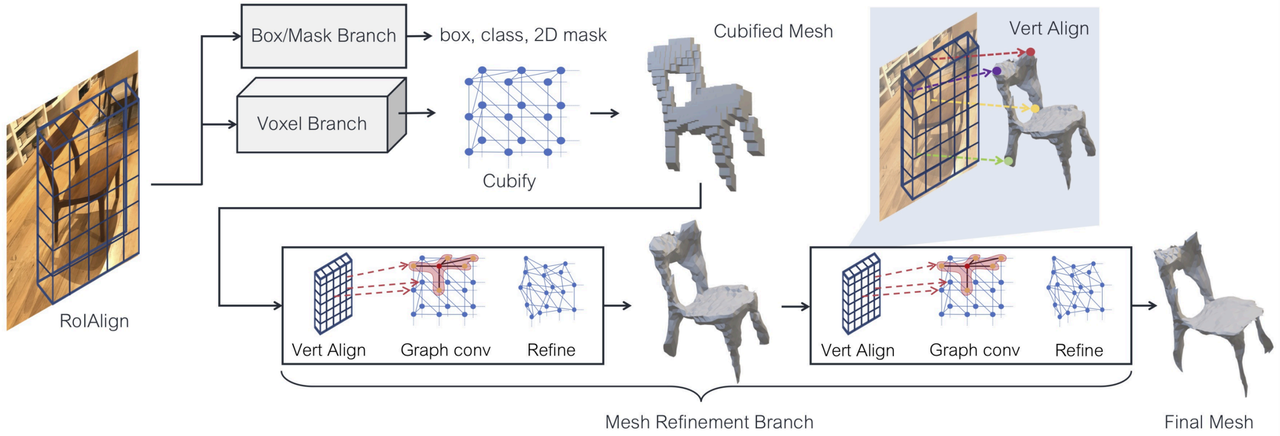
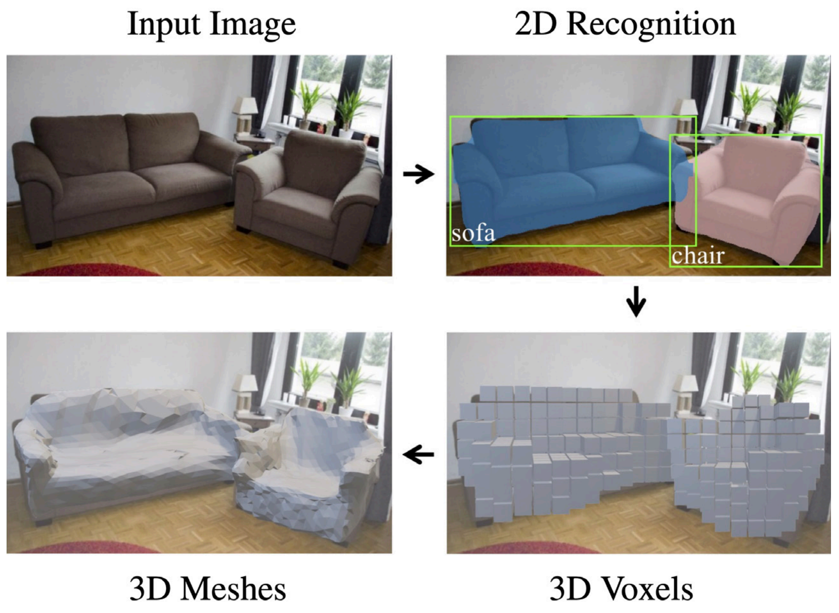
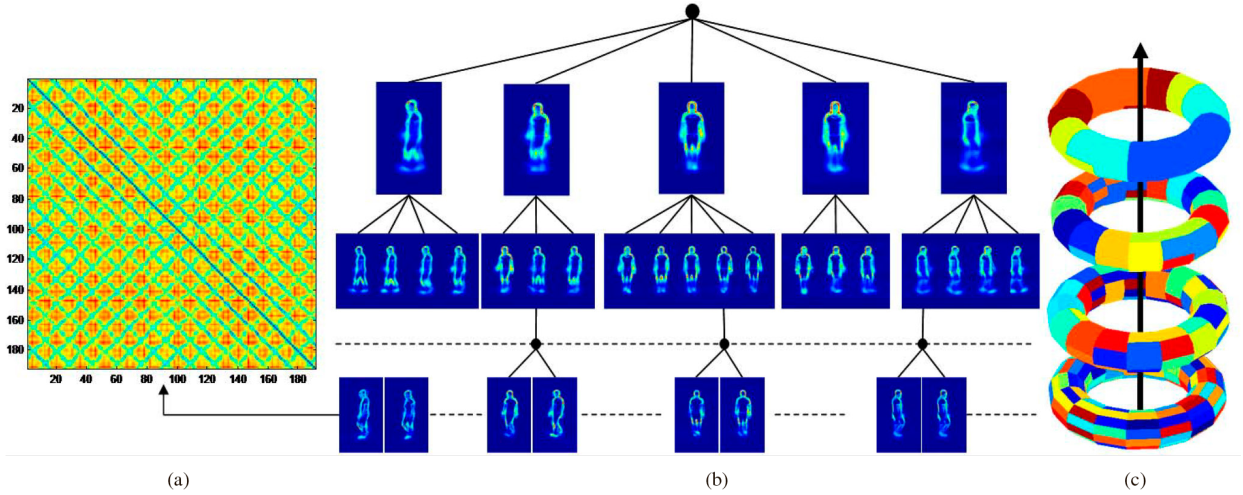


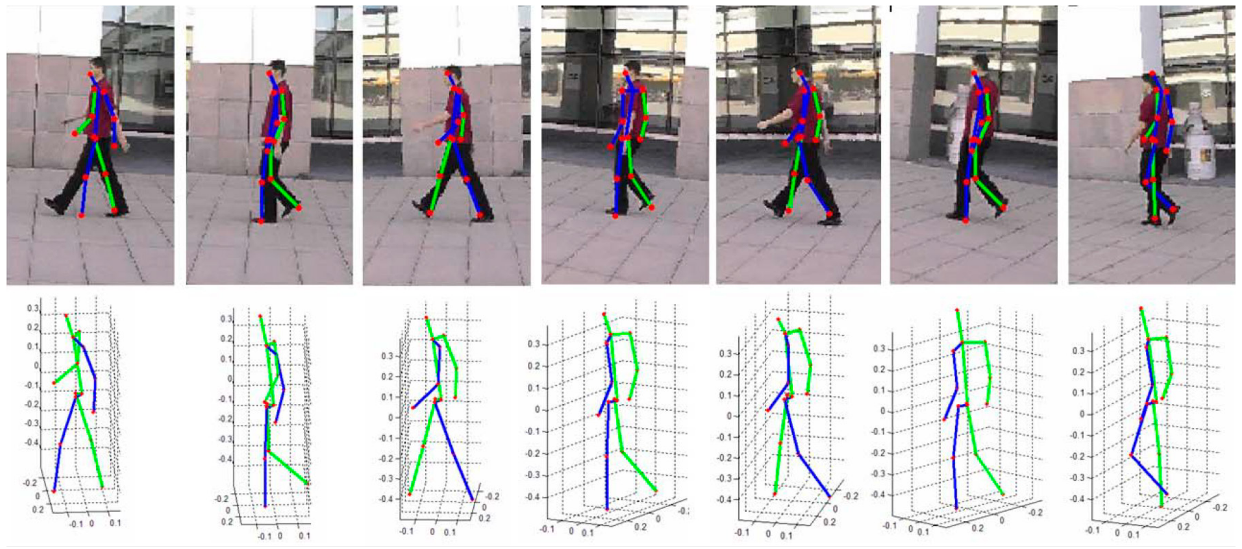
Figure 3. System overview of Mesh R-CNN. We augment Mask R-CNN with 3D shape inference. The *voxel branch* predicts a coarse shape for each detected object which is further deformed with a sequence of refinement stages in the *mesh refinement branch*.



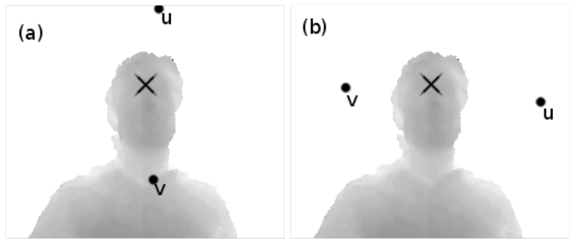
Mesh R-CNN (Georgia Gkioxari, Jitendra Malik, Justin Johnson, 2019)

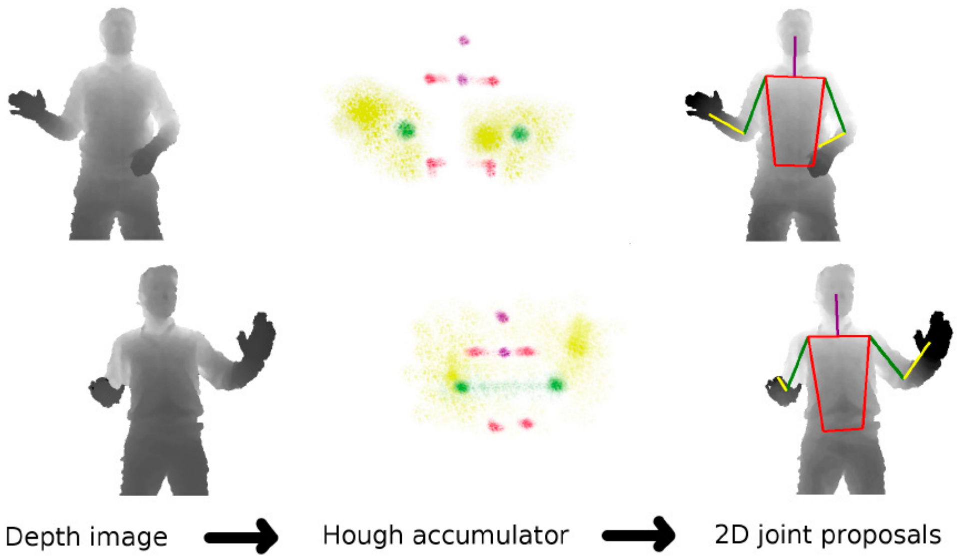
Mesh R-CNN (Georgia Gkioxari, Jitendra Malik, Justin Johnson, 2019)



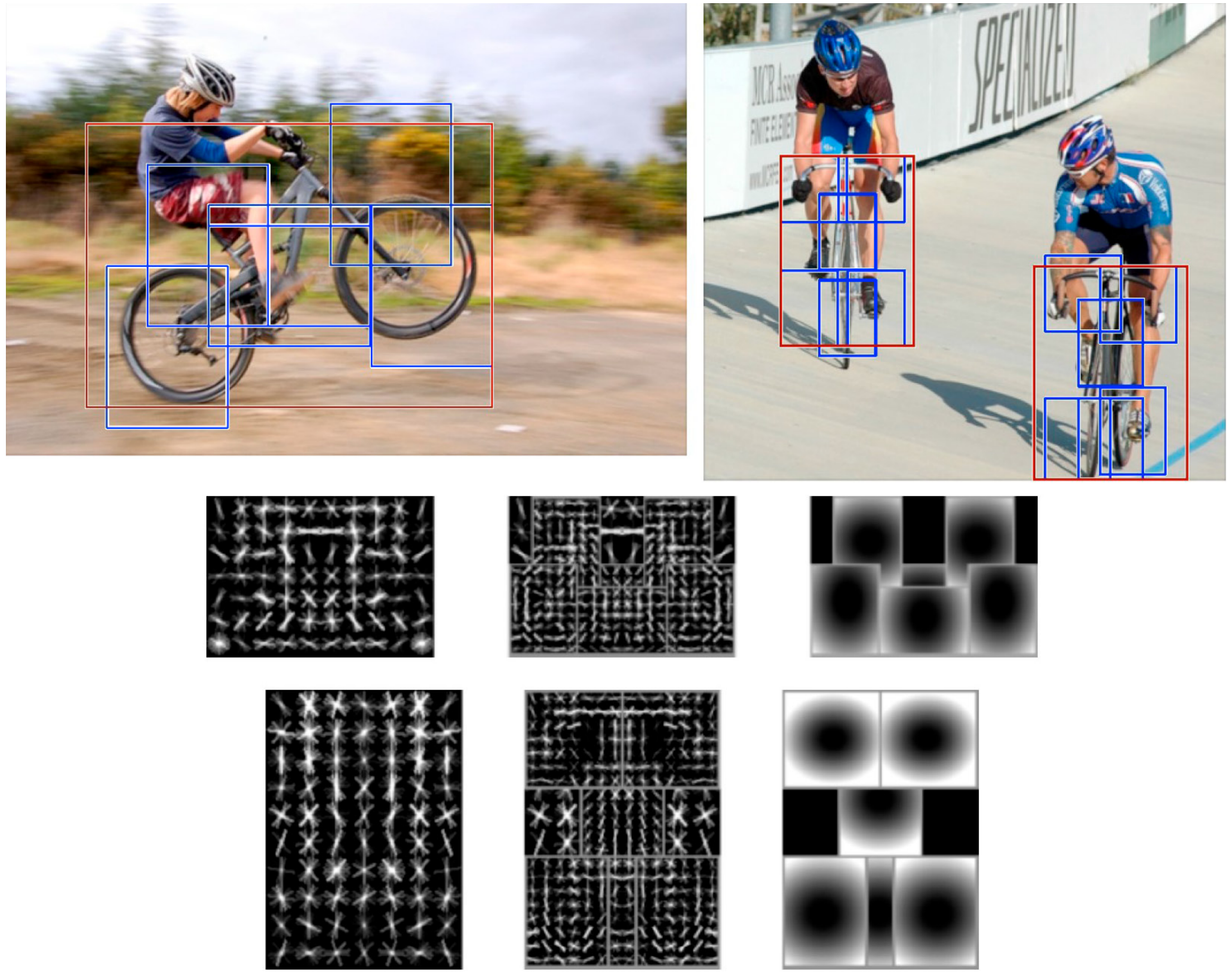


Pose detection using random forests (Rogez, Rihan, Ramalingam et al. 2008) c 2008 IEEE.





Static pose estimation from depth images using random regression forests and Hough voting (Brian Holt and Richard Bowden 2012)

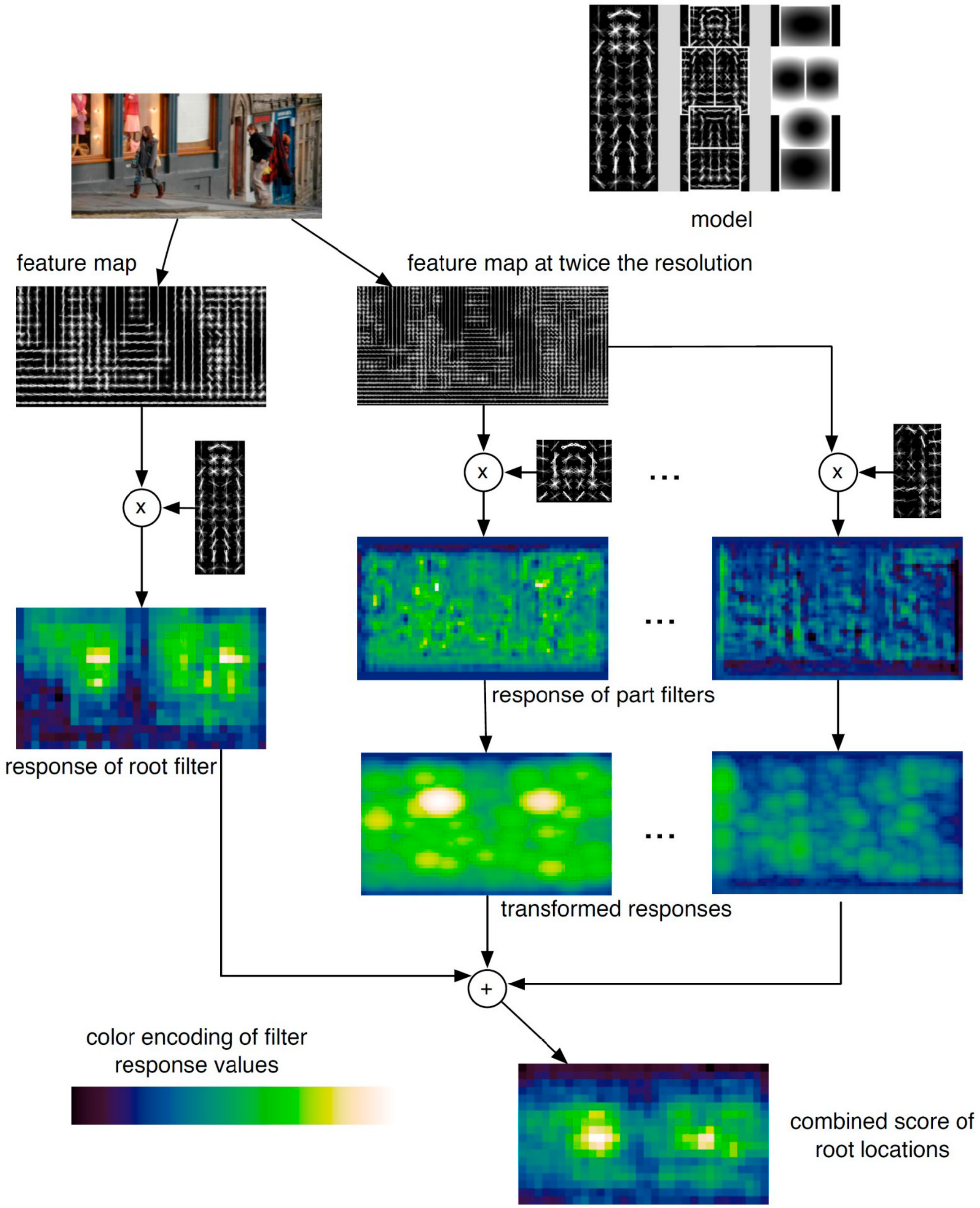


root filters
coarse resolution

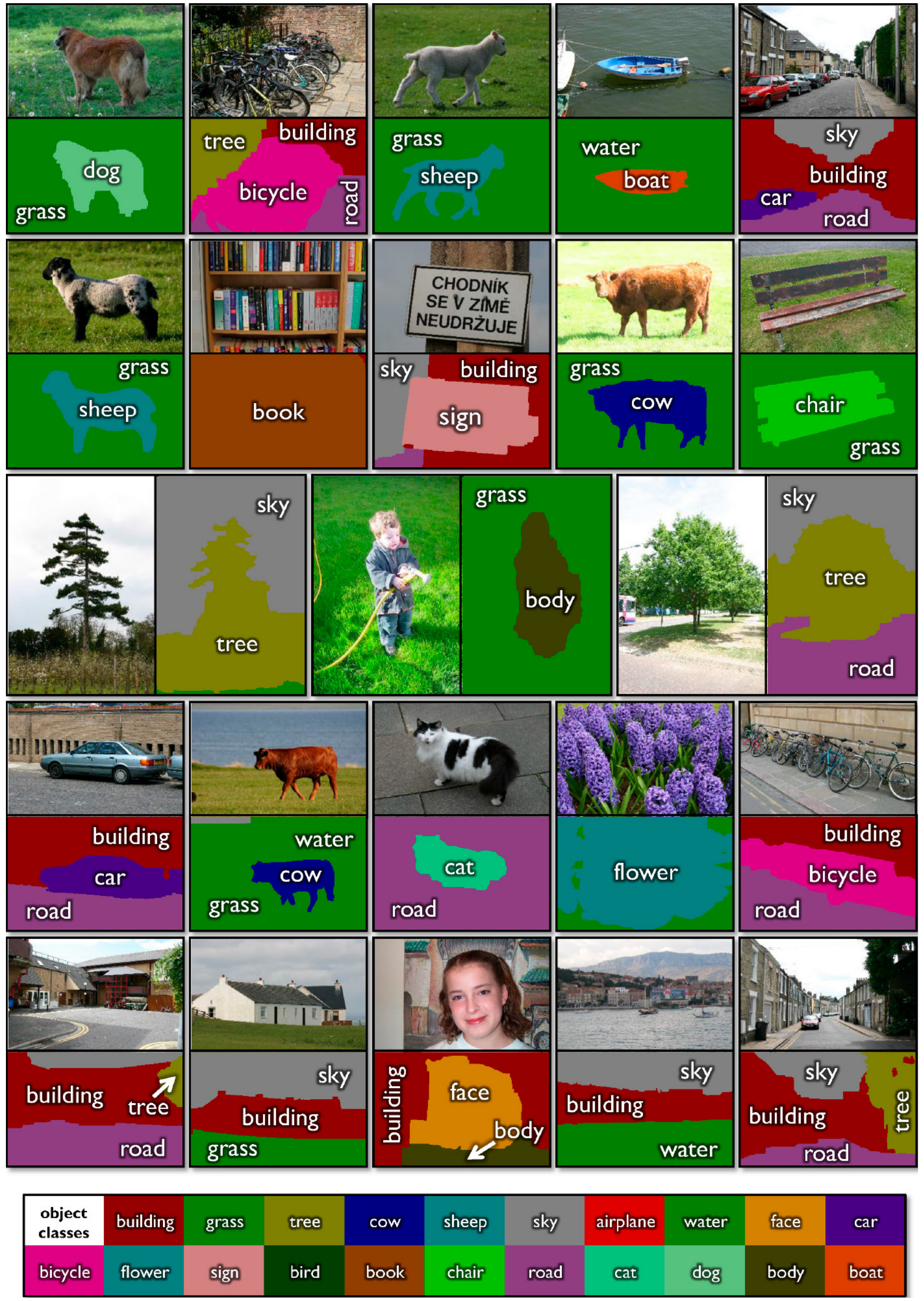
part filters
finer resolution

deformation
models

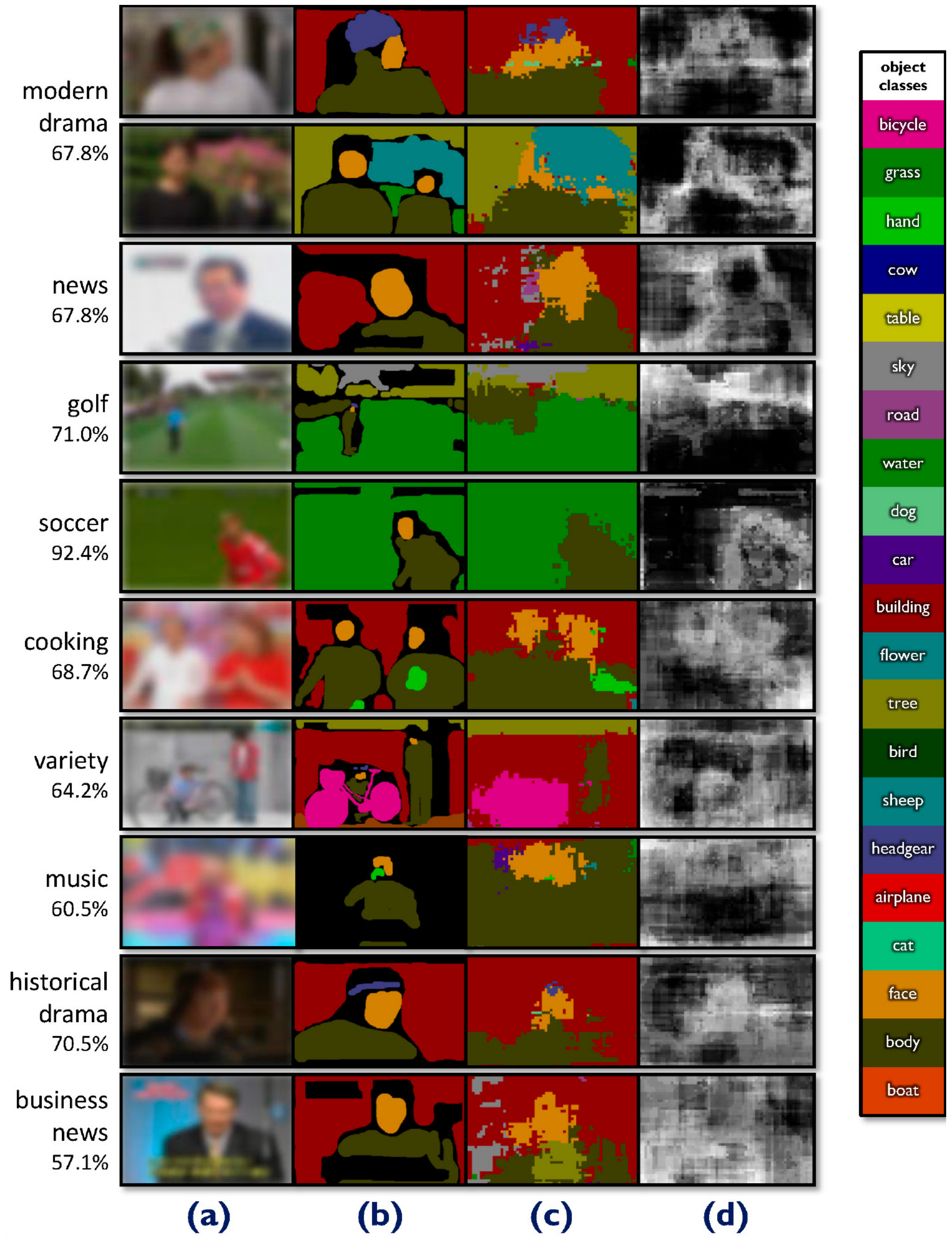
Rowley, Baluja, and Kanade: Neural Network-Based Face Detection (PAMI, January 1998)



Rowley, Baluja, and Kanade: Neural Network-Based Face Detection (PAMI, January 1998)

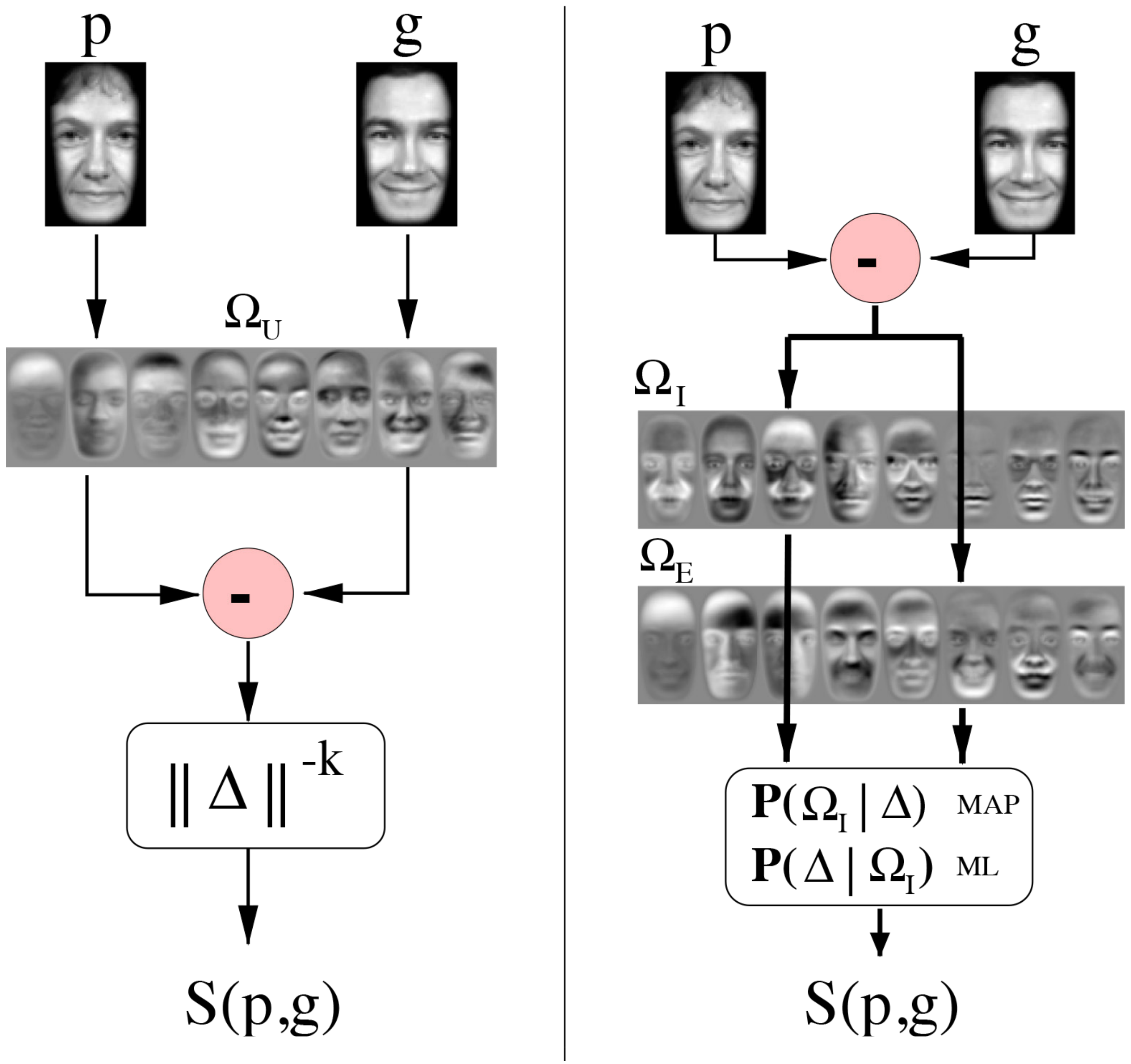


Example results on the MSRC 21-class datavabe.
Simultaneous recognition and segmentation (Shotton, Winn, Rother et al. 2009) c 2009 Springer.

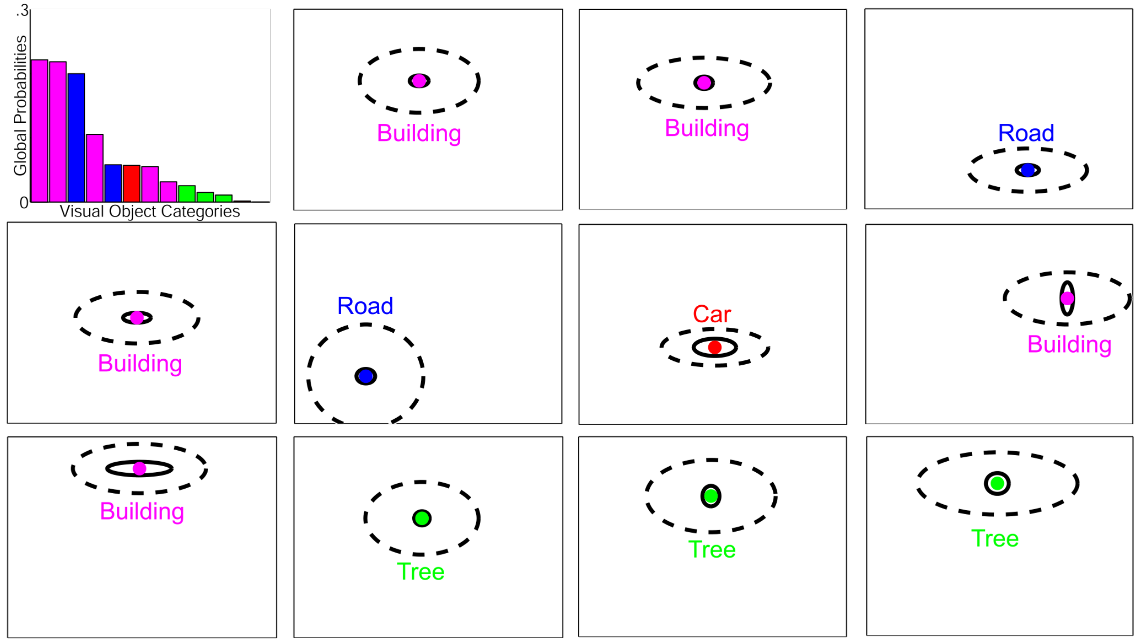


Example results on the television sequences. (a) Test images (blurred out for copyright reasons). (b) The hand-labeled ground truth. (c) The most likely labels inferred by the texture-layout potentials. (d) The entropy of the inferred class label distributions: white is high entropy, black low entropy. Class color key is given right, and pixel-wise segmentation accuracies for each dataset are shown.
Simultaneous recognition and segmentation (Shotton, Winn, Rother et al. 2009) c 2009 Springer.

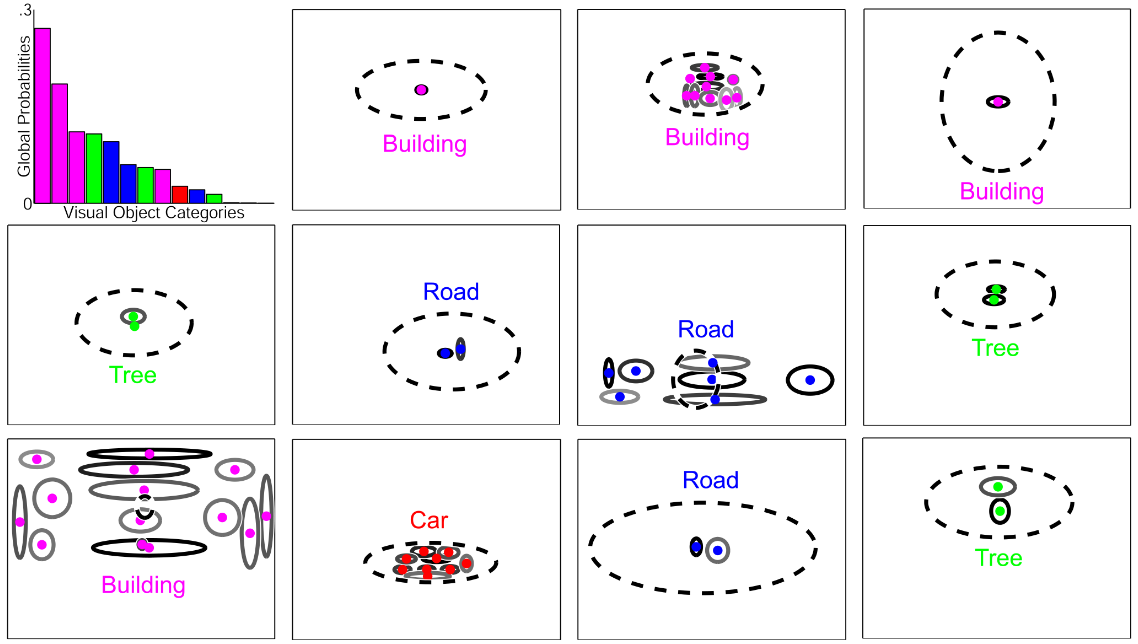
Standard Eigenfaces, “Dual” Eigenfaces: Intrapersonal, Extrapersonal.
Bayesian Face Recognition (Baback Moghaddam Tony Jebara Alex Pentland February 2002)



Bayesian Face Recognition (Baback Moghaddam Tony Jebara Alex Pentland February 2002)

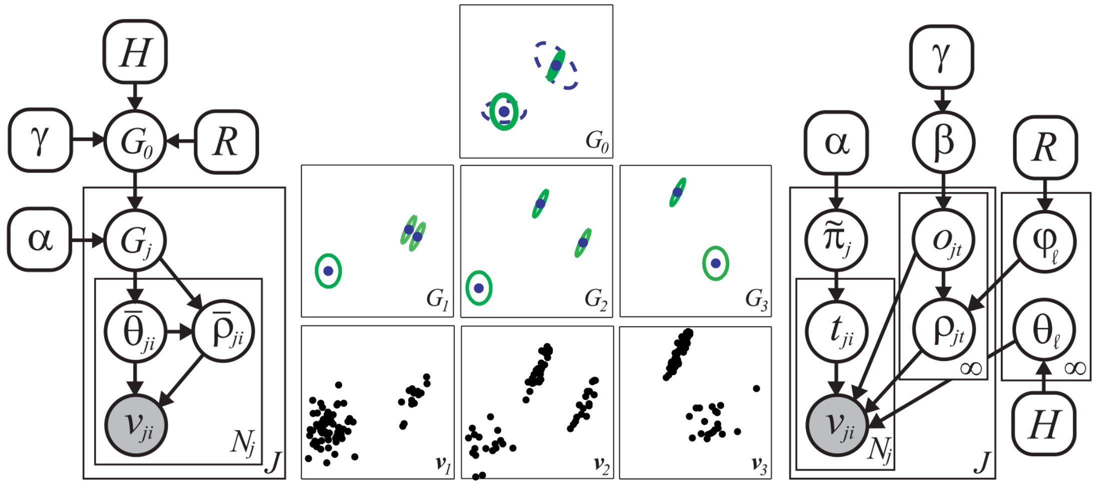


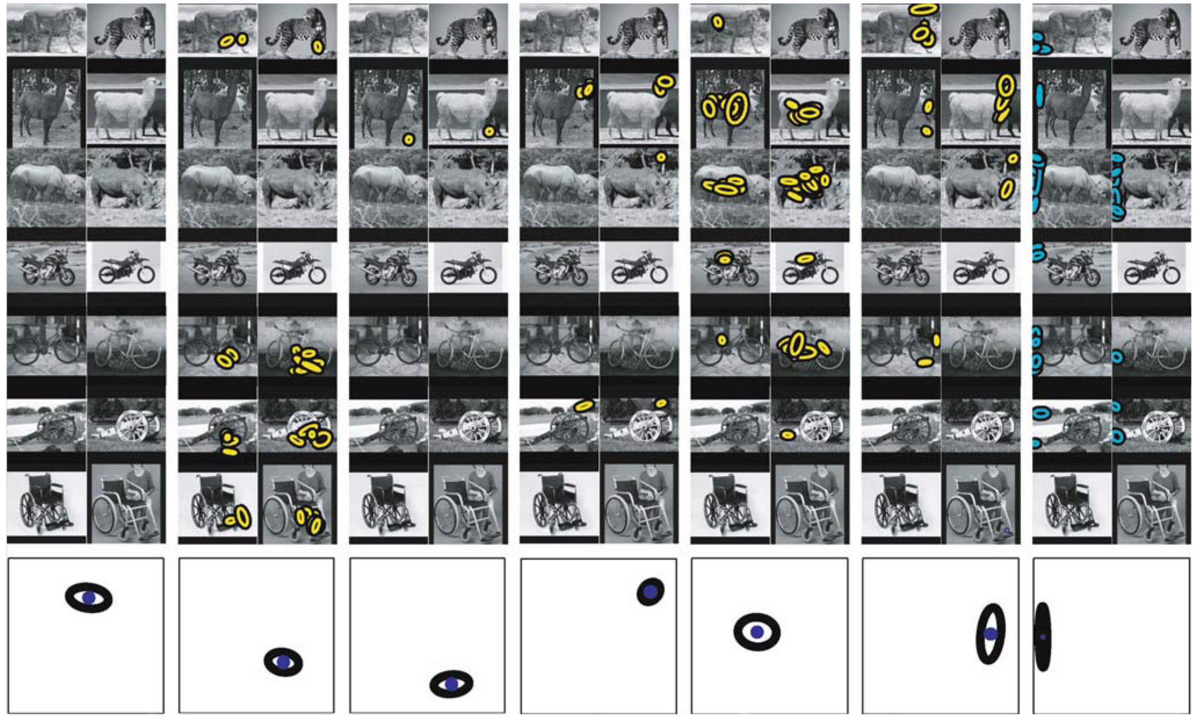
Single Part TDP



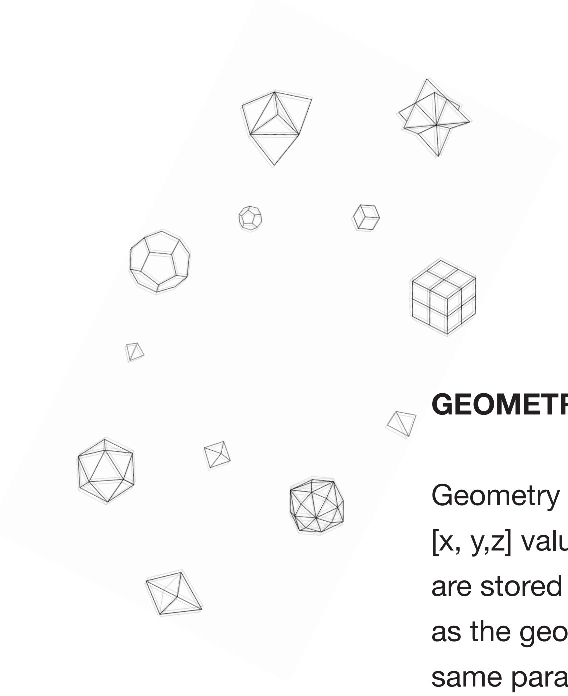
Multiple Part TDP

Contextual scene models for object recognition
Describing Visual Scenes Using Transformed Objects and Parts
(Erik B. Sudderth, Antonio Torralba, William T. Freeman, Alan S. Willsky 2007)



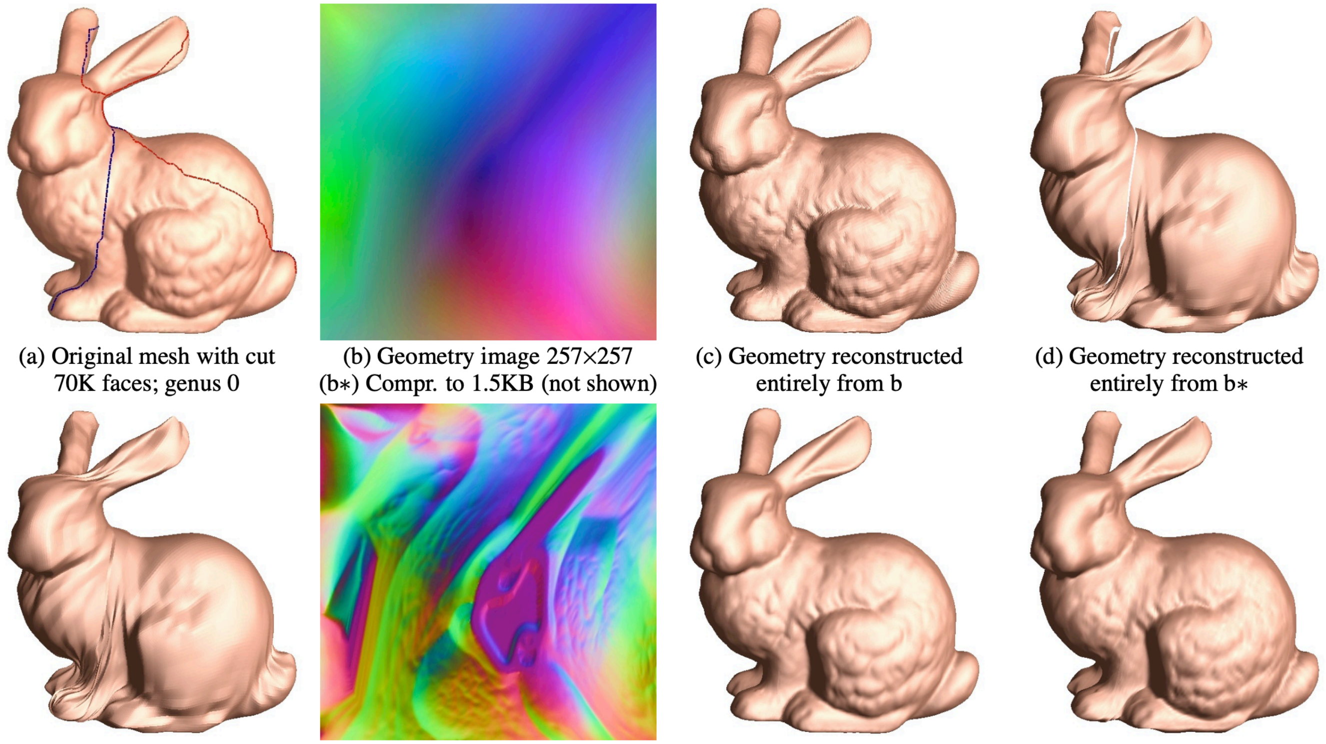


Contextual scene models for object recognition
Describing Visual Scenes Using Transformed Objects and Parts
(Erik B. Sudderth, Antonio Torralba, William T. Freeman, Alan S. Willsky 2007)

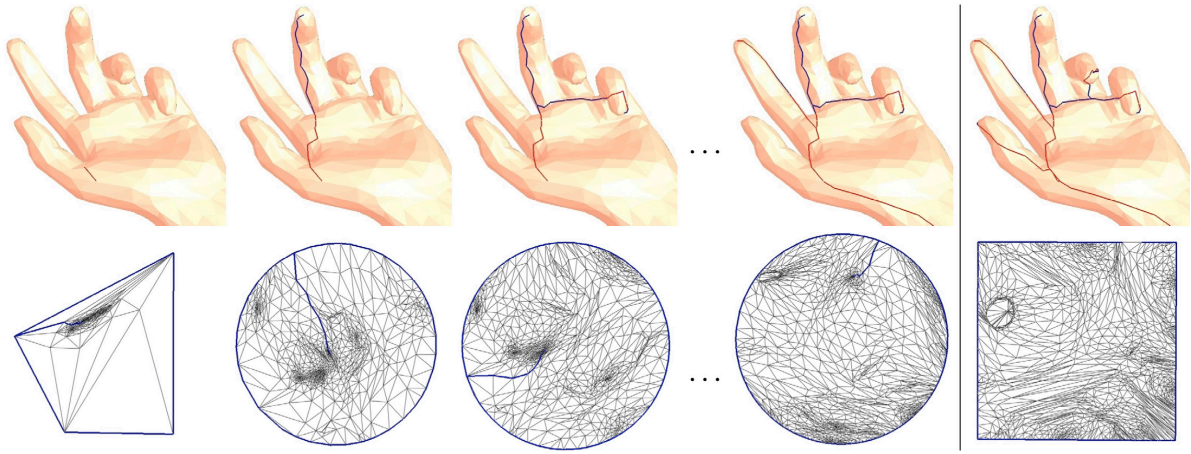


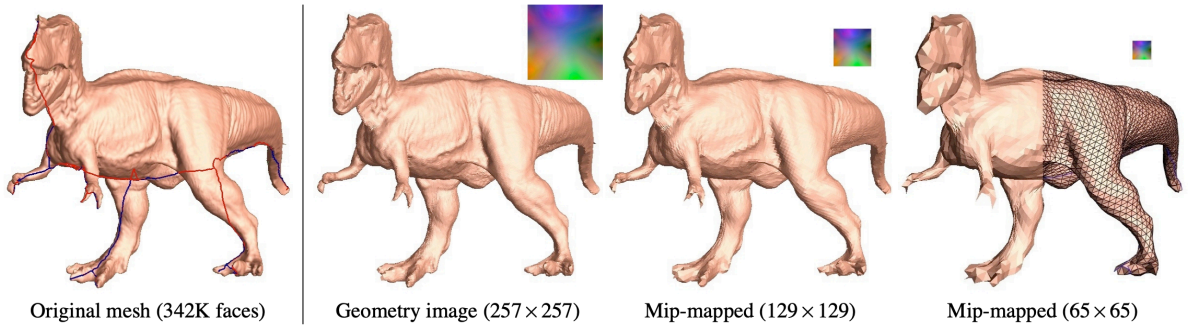
GEOMETRY IMAGES

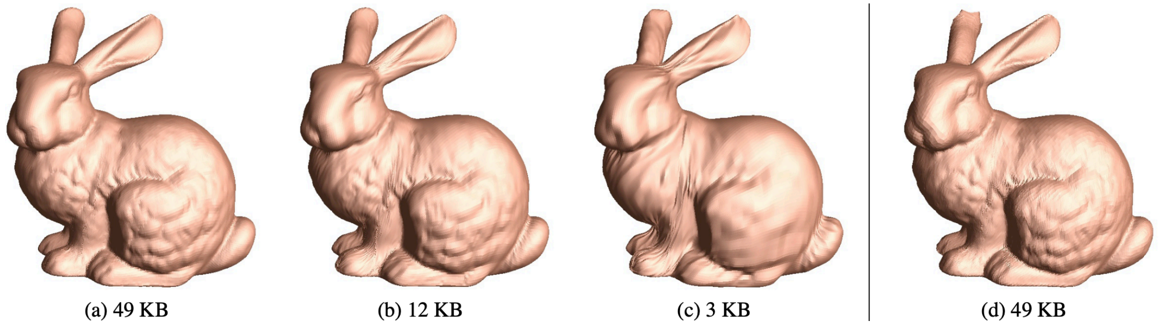
Geometry images — captures geometry as a simple n × n array of [x, y,z] values. Other surface attributes, such as normals and colors, are stored as additional square images, sharing the same domain as the geometry. Because the geometry and attributes share the same parametrization, the parametrization itself is implicit — “texture coordinates” are absent. Moreover, this parametrization fully utilizes the texture domain (with no wasted space). Geometry images can be encoded using traditional image compression algorithm, such as wavelet based coders. Also, geometry images are ideally suited for hardware rendering. They may be transmitted to the graphics pipeline in a compressed form just like texture images. And, they eliminate expensive pointer-based structures such as indexed vertex lists.



(a) Original mesh with cut 70K faces; genus 0
(b) Geometry image 257×257 (b∗) Compr. to 1.5KB (not shown)
(c) Geometry reconstructed entirely from b
(d) Geometry reconstructed entirely from b∗
(e) Geometry of d topology-fused using sideband data
(f) Normal-map image 512×512 (f∗) Compr. to 24KB (not shown)
(g) Geometry of c normal-mapped using f
(h) Geometry of e normal-mapped using f∗

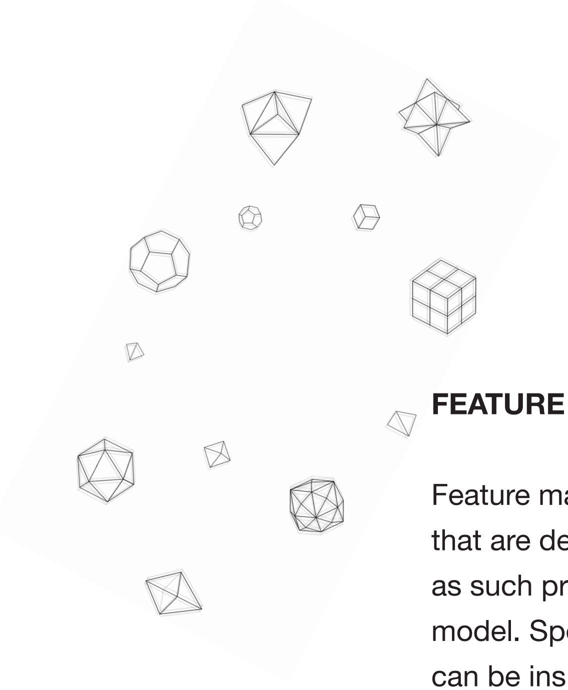






Geometry Images
(Xianfeng Gu, Steven J. Gortle, Hugues Hoppe, 2002)

Geometry Images
(Xianfeng Gu, Steven J. Gortle, Hugues Hoppe, 2002)

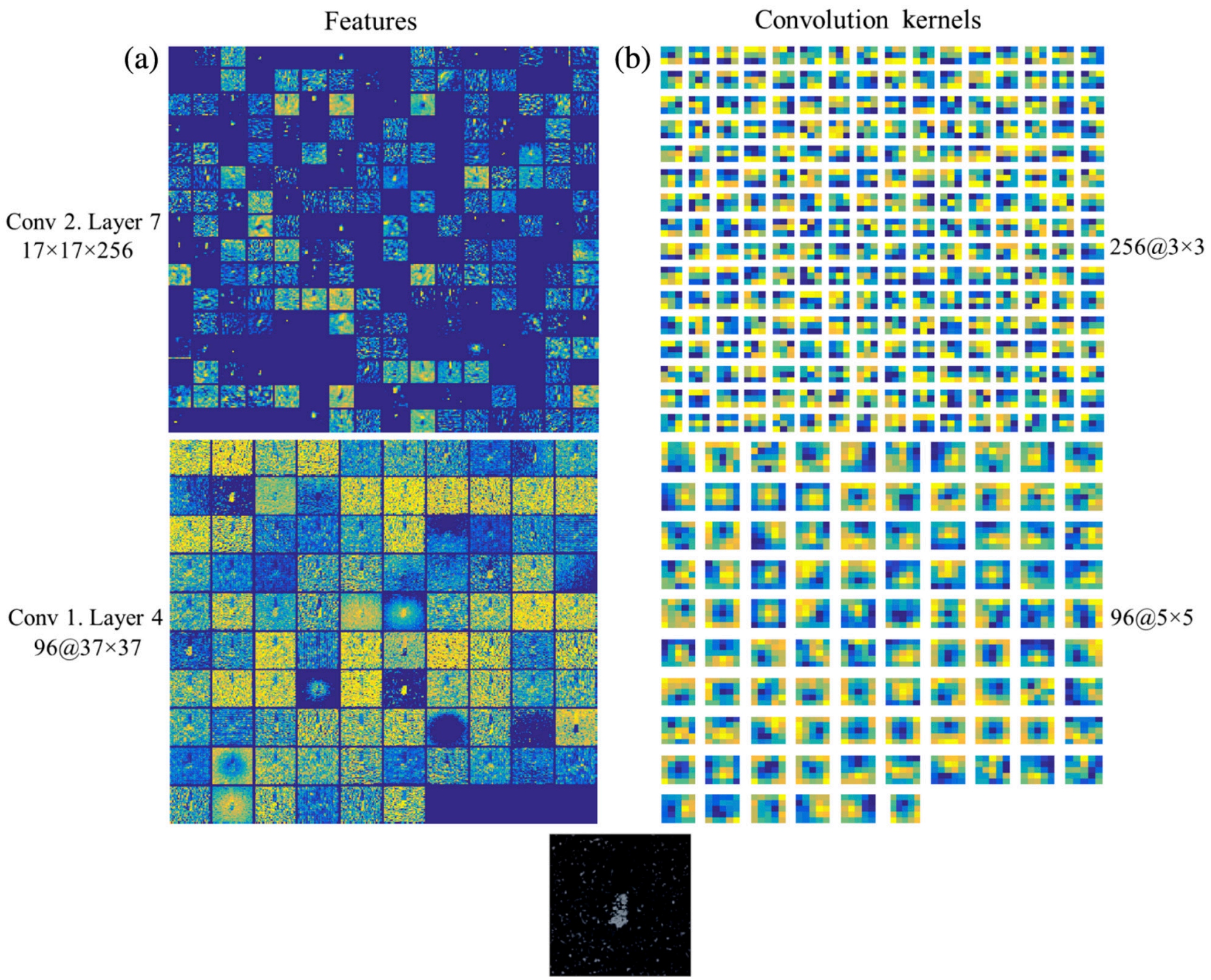


FEATURE MAPS

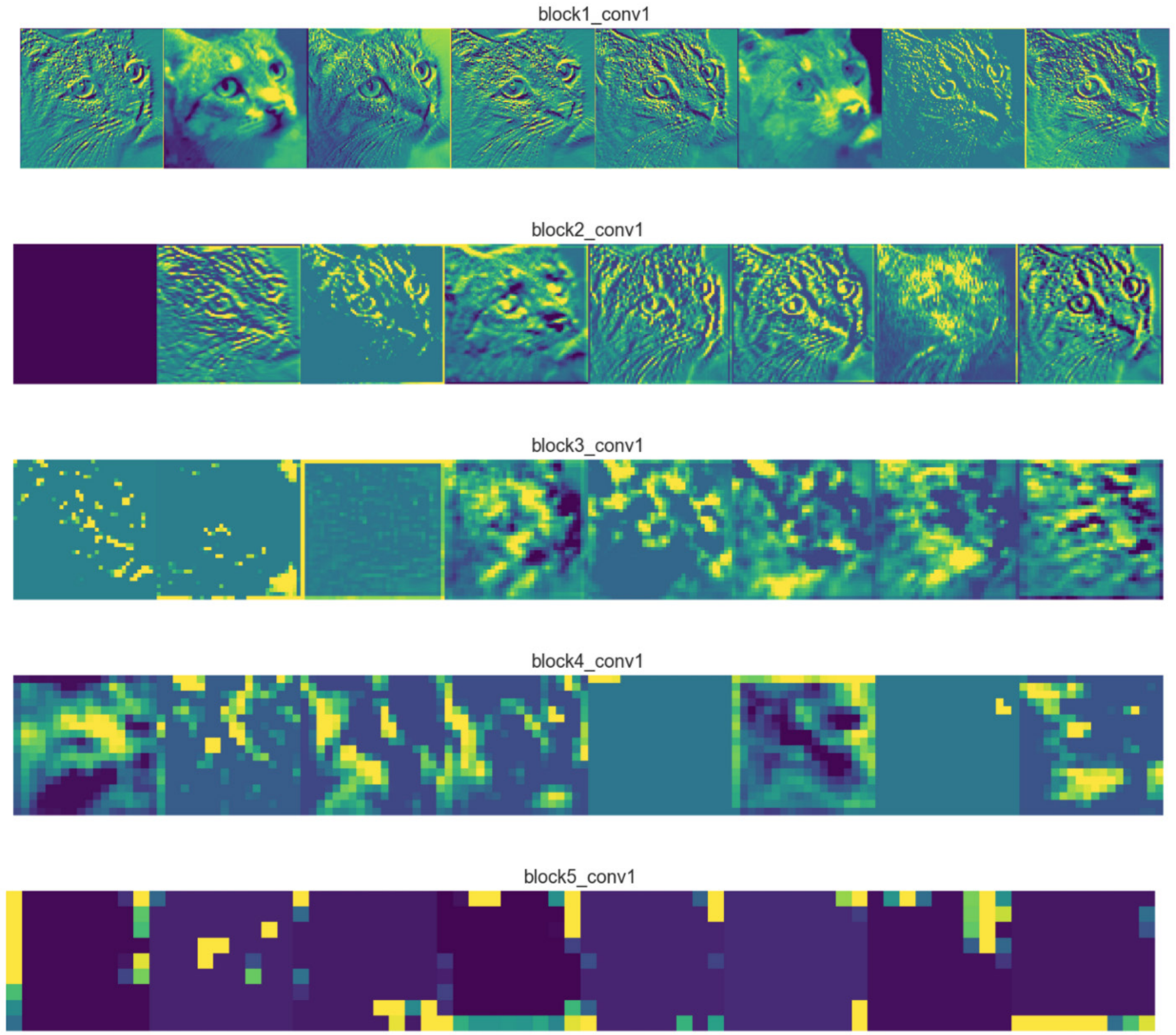
Feature maps — convolutional neural networks, have internal structures that are designed to operate upon two-dimensional image data, and as such preserve the spatial relationships for what was learned by the model. Specifically, the two-dimensional filters learned by the model can be inspected and visualized to discover the types of features that the model will detect, and the activation maps output by convolutional layers can be inspected to understand exactly what features were detected for a given input image.

In a convolutional neural network units within a hidden layer are segmented into «feature maps» where the units within a feature map share the weight matrix, or in simple terms look for the same feature. The hidden units within a feature map are unique in that they are connected to different units in the lower layer. So for the first hidden layer, units within a feature map will be connected to different regions of the input image.

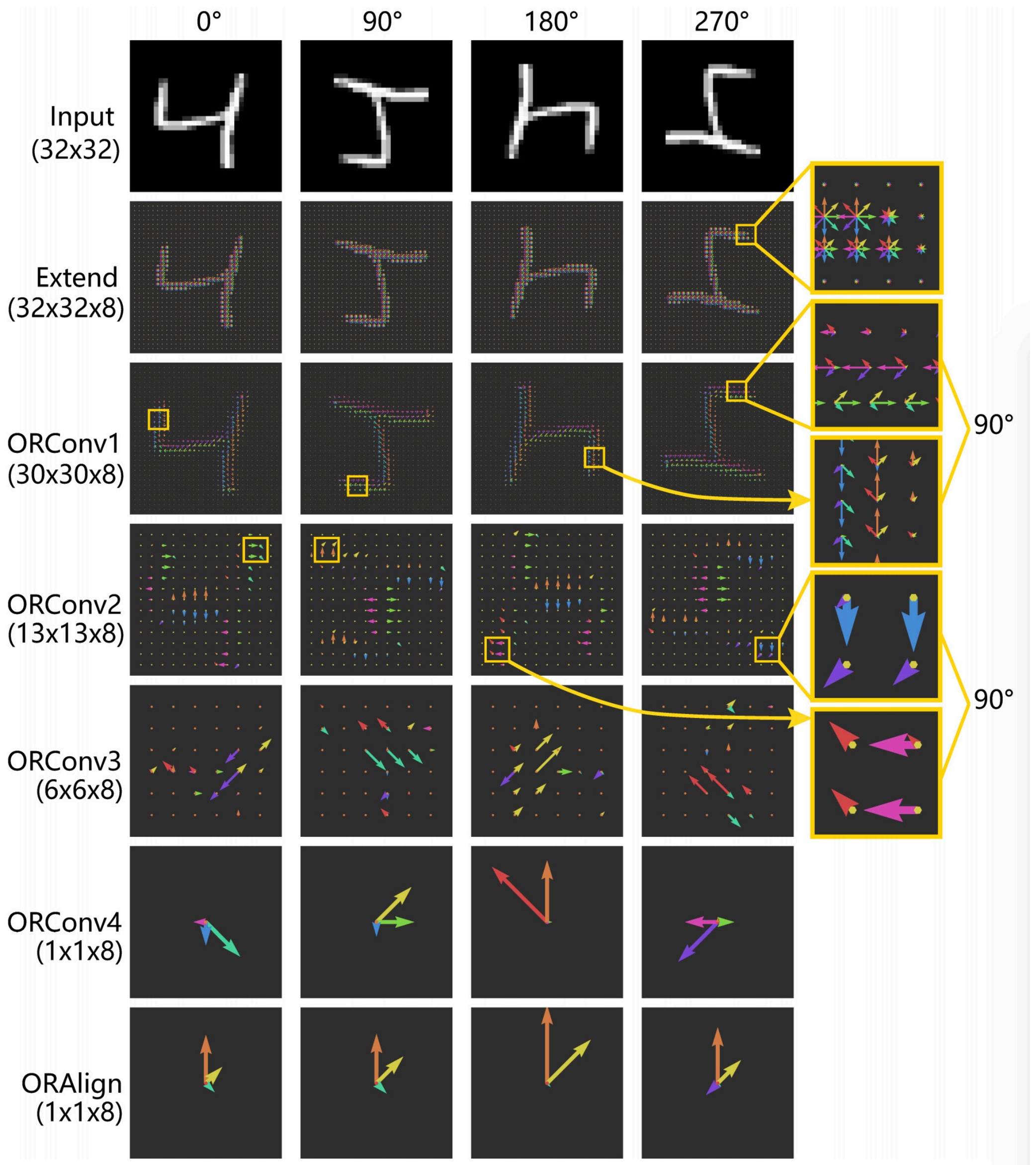
So in summary, a hidden layer is segmented into feature maps where each unit in a feature map looks for the same feature but at different positions of the input image.



Deep features visualization: (a) output feature maps and (b) convolution kernels of the considered. Deep feature extraction and combination for synthetic aperture radar target classification (Moussa Amrani, Feng Jiang, 2017)



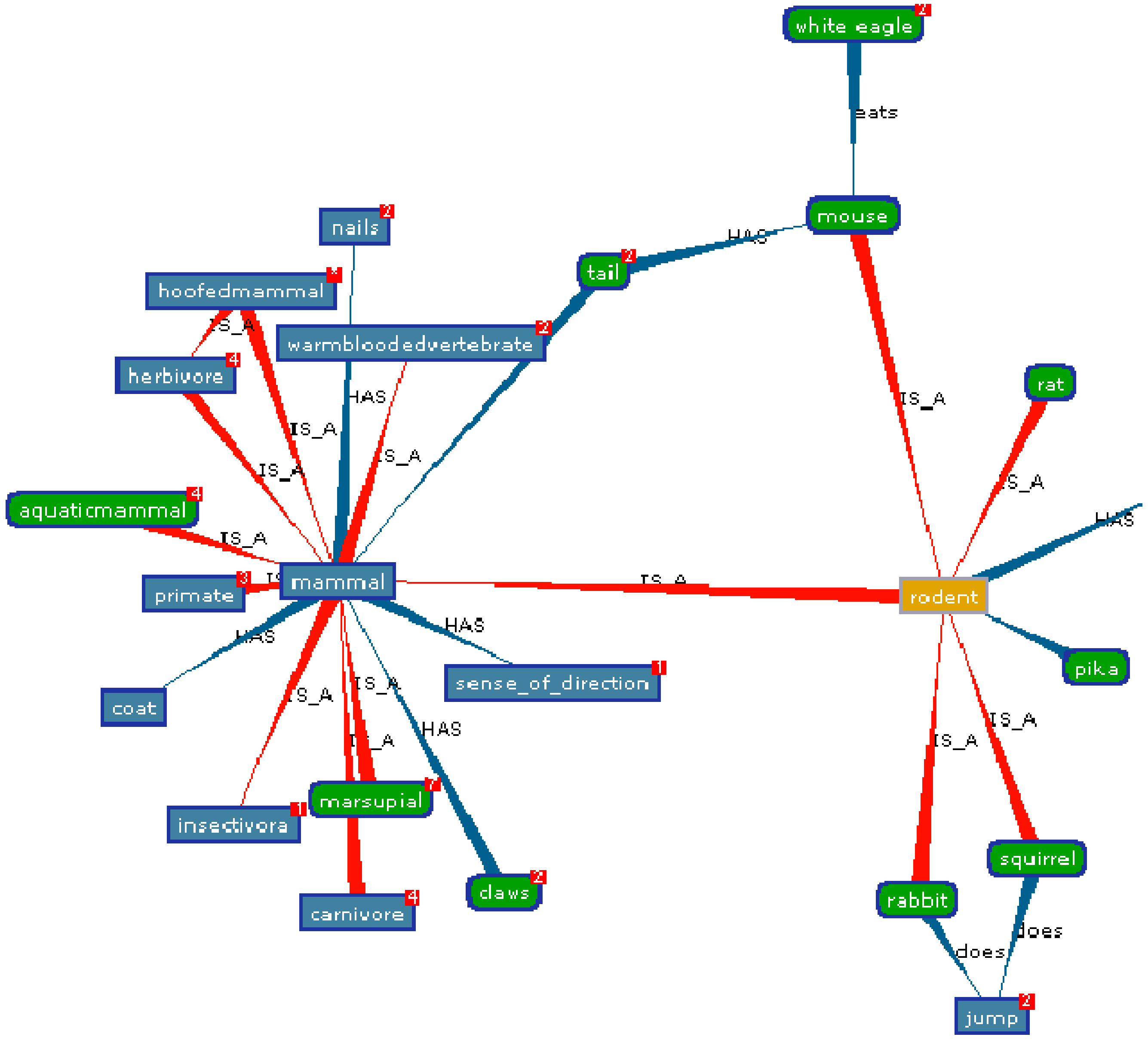
8 feature maps per layer. Block1_conv1 actually contains 64 feature maps, since we have 64 filters in that layer. But we are only visualizing the first 8 per layer in this figure.



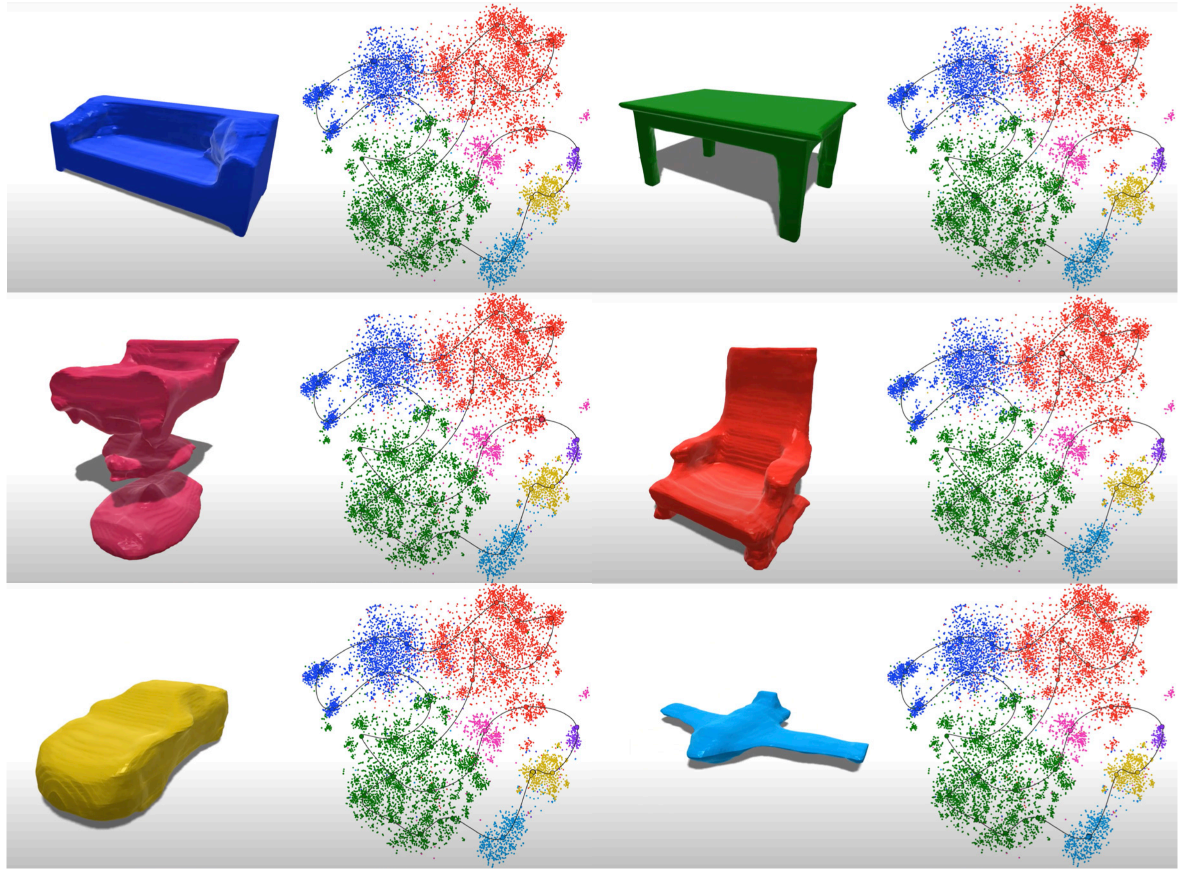
Feature map
Oriented Response Networks (Yanzhao Zhou, Qixiang Ye, Qiang Qiu and and Jianbin Jiao, 2017)

SEMANTIC SPACE VISUALIZATION

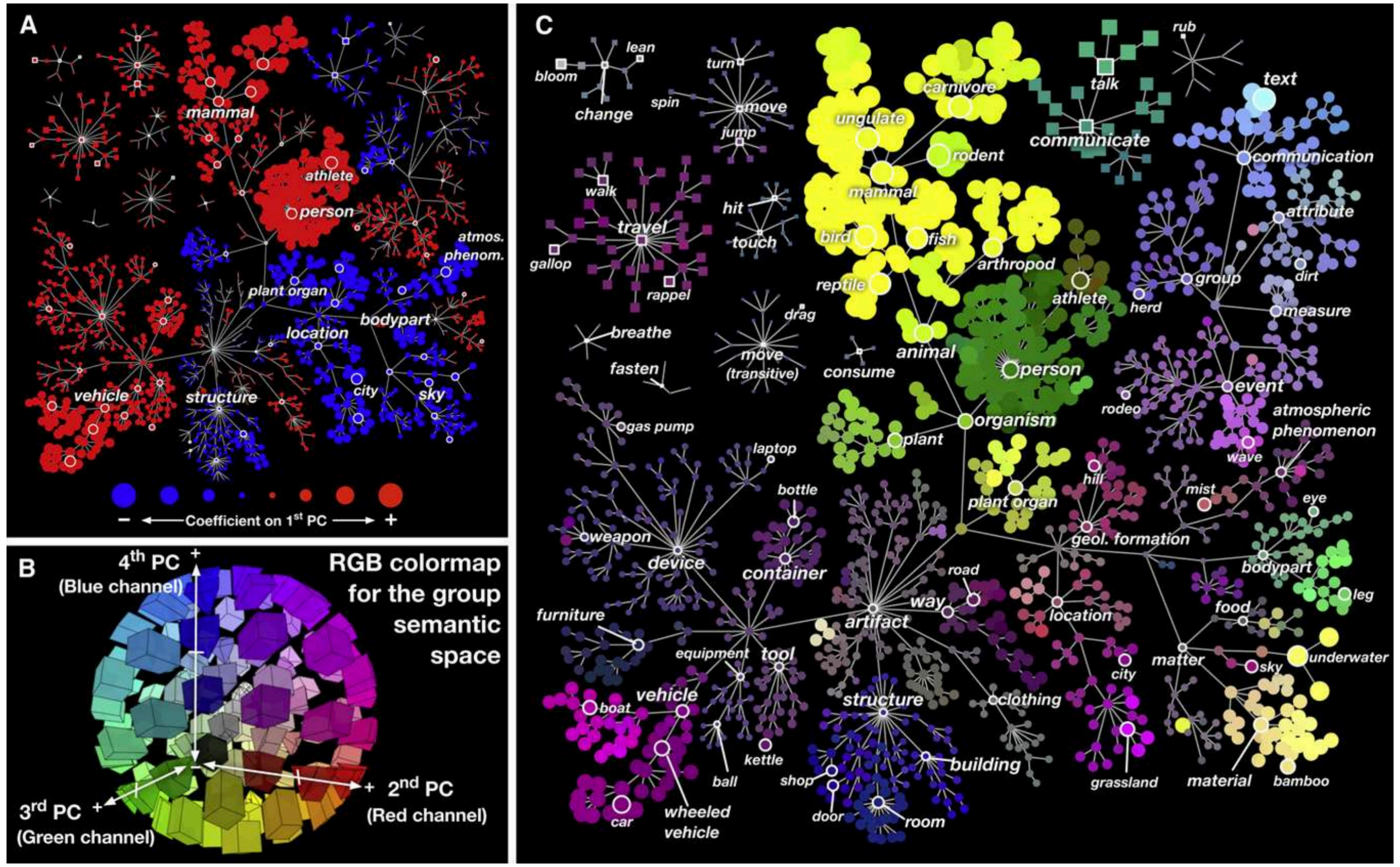
Semantic space visualization — the visualization of the simulation results must be done in conformity with beneficiaries perception and professional domain understanding. It means that right data must be identified before. Semantic technologies provide new ways for accessing data and acquiring knowledge.



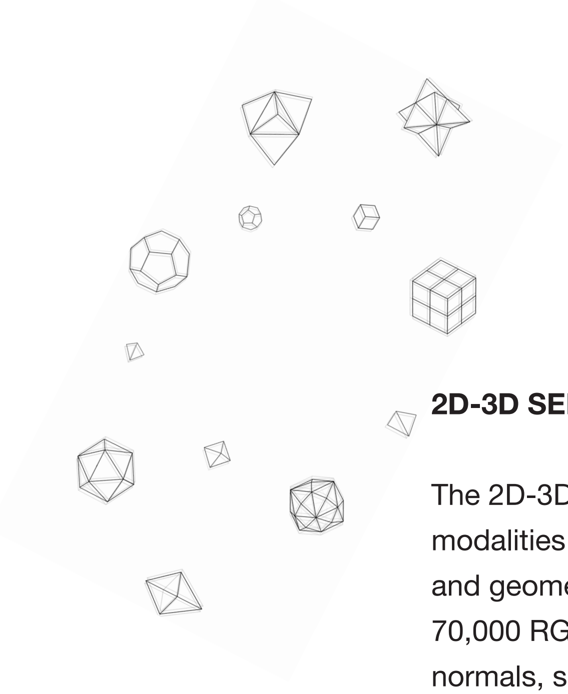
Visualization of Semantic Space for animal kingdom.
Dynamic Semantic Visual Information Management
(Julian Szymanski, Duch Wlodzislaw, 2010)



Visualization of the DeepSDF latent space using t-SNE

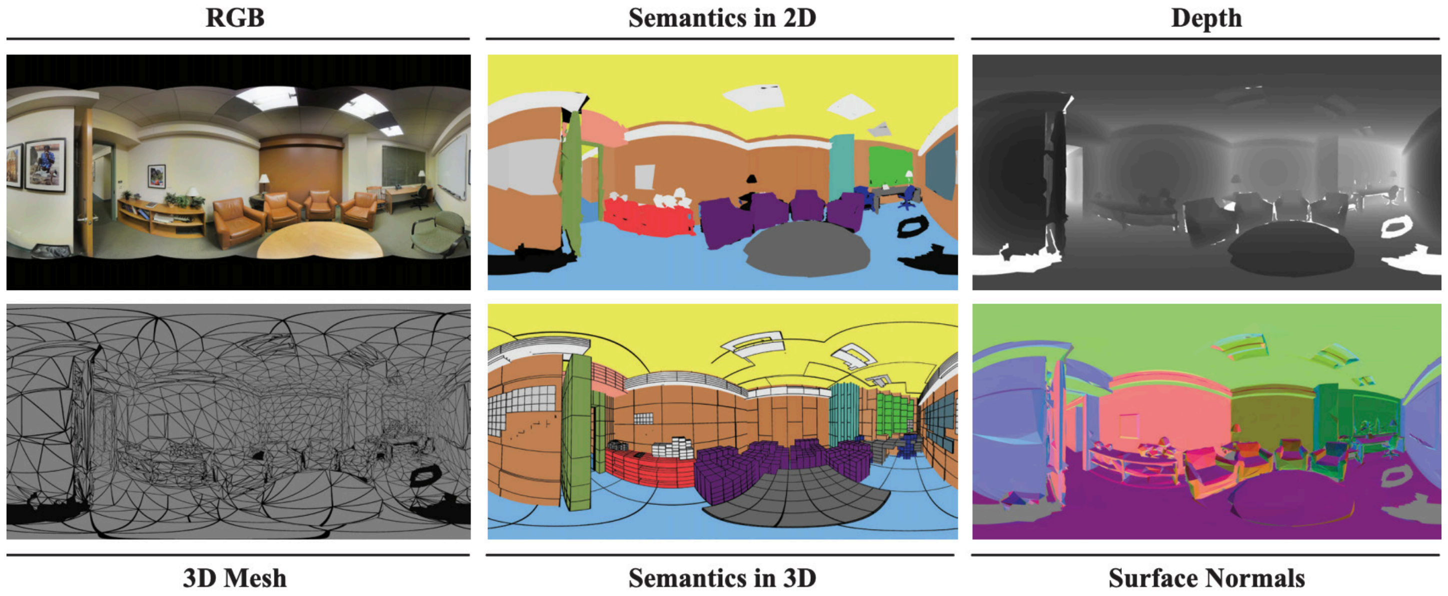


A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain (Alexander G. Huth, Shinji Nishimoto, An T. Vu and Jack L. Gallant, 2012)



2D-3D SEMANTIC DATASETS

The 2D-3D-S dataset — provides a variety of mutually registered modalities from 2D, 2.5D and 3D domains, with instance-level semantic and geometric annotations. It covers over 6,000 m2 and contains over 70,000 RGB images, along with the corresponding depths, surface normals, semantic annotations, global XYZ images (all in forms of both regular and 360° equirectangular images) as well as camera information. It also includes registered raw and semantically annotated 3D meshes and point clouds. In addition, the dataset contains the raw RGB and Depth imagery along with the corresponding camera information per scan location. The dataset enables development of joint and cross-modal learning models and potentially unsupervised approaches utilizing the regularities present in large-scale indoor spaces.



Joint 2D-3D-Semantic Data for Indoor Scene Understanding
(Iro Armen, SAlexander Sax, Amir R. Zamir, Silvio Savarese, 2017)

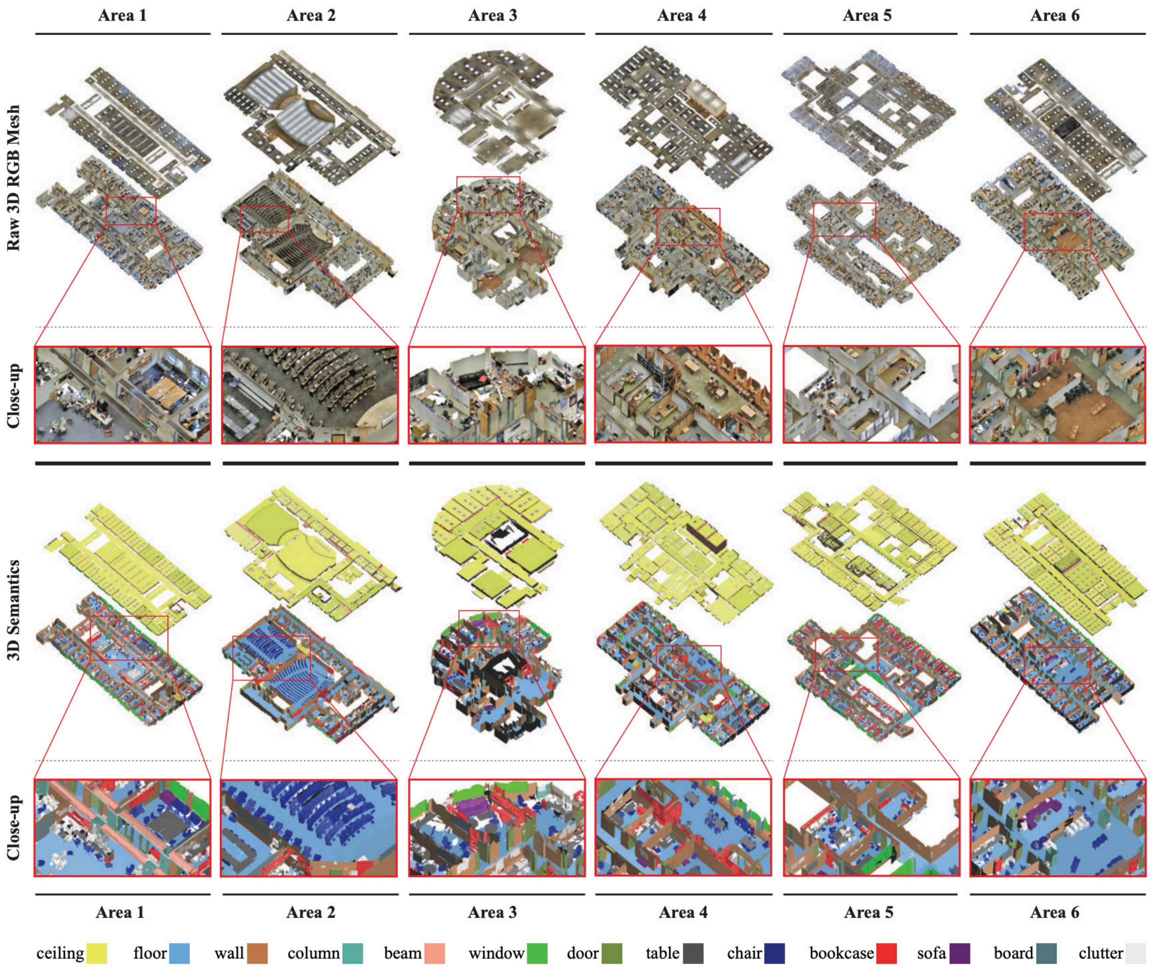
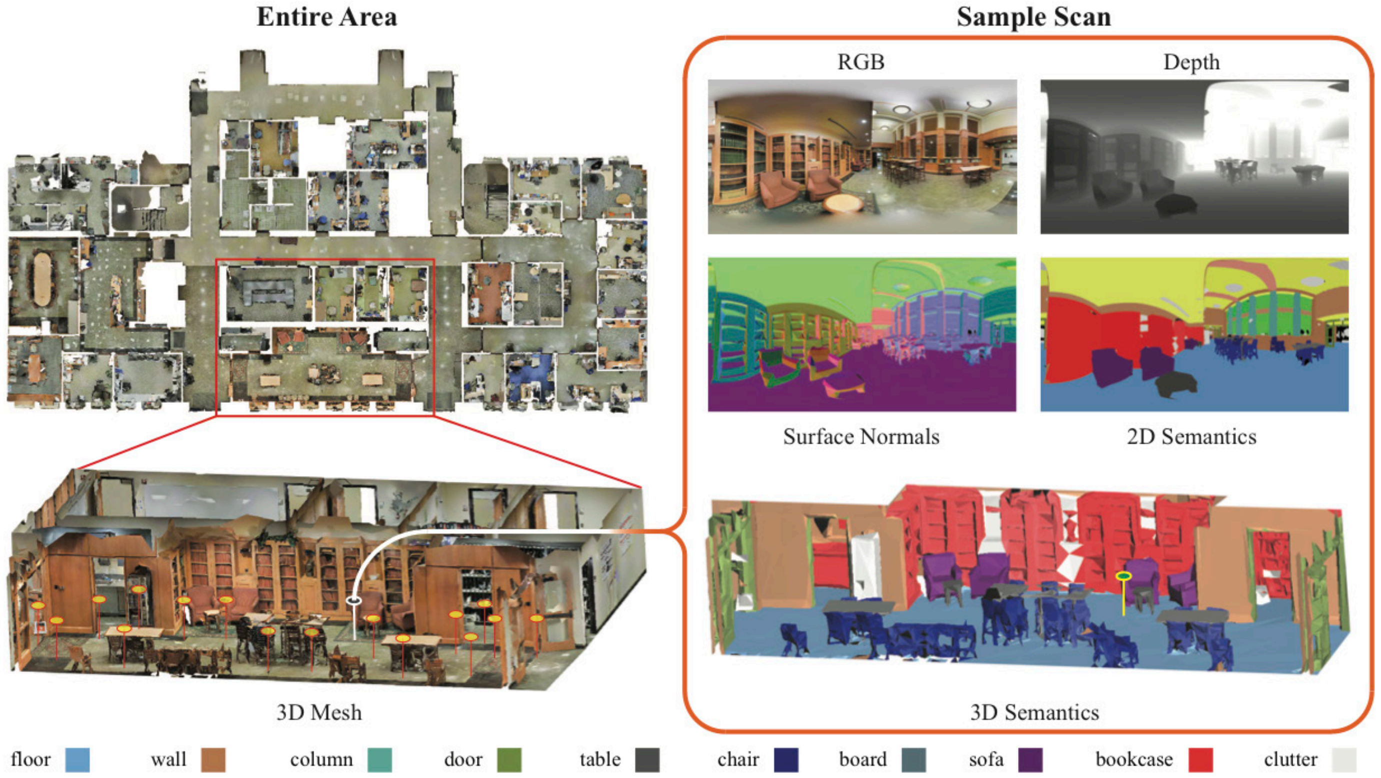


Figure 2: **3D Modalities.** The dataset includes both the textured and semantic 3D mesh models of all areas as well as their point clouds.



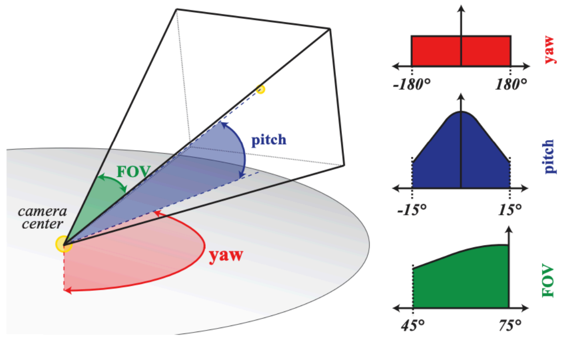


Figure 5: **Sampling distributions.** We sample camera parameters from the above distributions of yaw, pitch and Field of View (FOV).

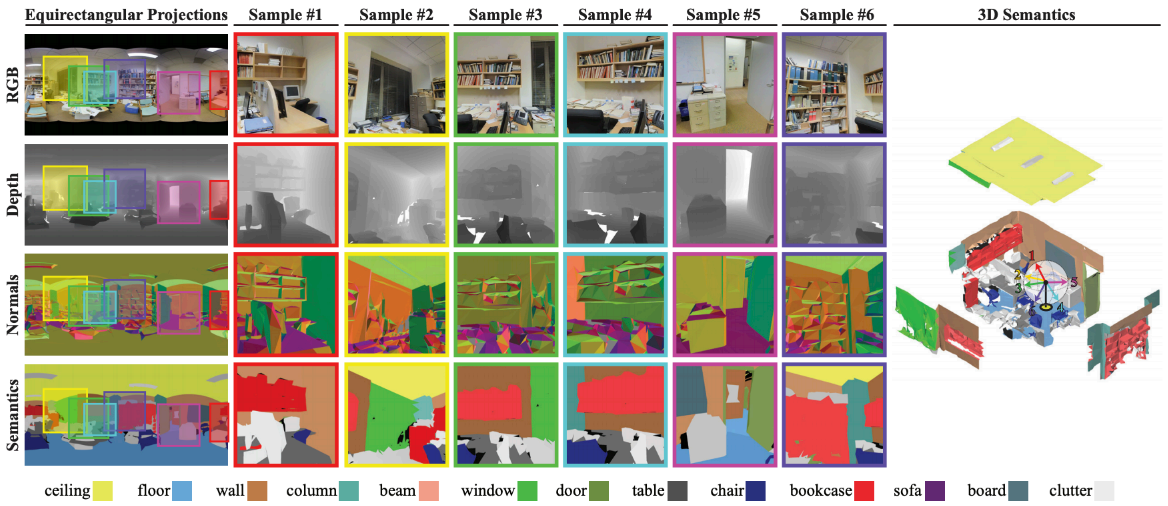


Figure 4: **Sampling images from the equirectangular projection.** We use the equirectangular projections to sample 72 images per scan location, all with consistent depth, surface normal, and semantic information. The sampling distributions are provided in Figure 5.

Joint 2D-3D-Semantic Data for Indoor Scene Understanding
(Iro Armen, SAlexander Sax, Amir R. Zamir, Silvio Savarese, 2017)

Joint 2D-3D-Semantic Data for Indoor Scene Understanding
(Iro Armen, SAlexander Sax, Amir R. Zamir, Silvio Savarese, 2017)

PHYGITALISM

Physical + Digital = Phygital.

We believe that society is already moving to a new type of communication – the symbiosis of the digital and physical world.

We develop phygital-products for business and explore new ways of human-computer interaction. We are trying to contribute to the creation of a new communication environment with the help of amazing technologies, business models, science and art.

We work with a large stack of technologies:
– **xR** (virtual, augmented, mixed realities)
– **CG** (computer graphics)
– **ML** (machine learning)
– **Web** (web services)
– **IoT** (internet of things and industrial internet)

The name of the company PHYGITALISM sounds like an art movement. We believe that phygital will inspire many artists to bring their exciting ideas to life through the existing physical world and technologies. Our goal is to create an exceptional world precedent when the name of the company was the beginning of an art movement.

Learn more www.phygitalism.com

ACRONYMS

2D Two-dimension
3D Three-dimension
FOV Field of View
GAPD Geiger-mode Avalanche Photo Diode
GPU Graphics Processing Unit
GUM Guide to the Expression of Uncertainty in Measurement
ICP Iterative Closest Point
IMU Inertial Measurement Unit
IR Infra-Red
LiDAR Light Detection And Ranging
NIR Near Infra-Red
PCL Point Cloud Library
RADAR Radio Detection and Ranging
RANSAC RANdom SAmple Consensus
SDK Software Development Kit
SfM Structure-from-Motion
SI International System
SNR Signal-to-Noise Ratio
SONAR Sound Detection and Ranging
TOF Time-of-Flight

REFERENCES

Best Thermal Cameras For Sale Today: The Ultimate Buying Guide. (2020). Retrieved 9 July 2020, from http://www.thermalcameras.guide/best-thermal-cameras-on-market/

Какая теплопроводность газобетона - определяем толщину стены. (2020). Retrieved 9 July 2020, from http://stroy-gazobeton.ru/61-kakaya-teploprovodnost-gazobetona

Fluke TiX1000 Infrared Camera. (2020). Retrieved 9 July 2020, from http://en-us.fluke.com/products/infrared-cameras/tix1000-infrared-camera.html#slideshow-272610531

M7 LWIR X-Long Range PTZ FLIR Thermal Imaging camera. (2020). Retrieved 9 July 2020, from https://www.x20.org/product/m7-ptz-long-range-thermal-imager/

Thermal imaging service developed to protect against the spread of coronavirus. (2020). Retrieved 9 July 2020, from https://www.smartcitiesworld.net/news/news/thermal-imaging-service-developed-to-protect-against-the-spread-of-coronavirus-5215

Watson, L. (2020). Council spends £24,000 on spy plane with thermal imaging cameras to catch immigrants living illegally in ‘sheds with beds’. Retrieved 9 July 2020, from https://www.dailymail.co.uk/news/article-2346759/Council-uses-spy-plane-thermal-imaging-cameras-catch-immigrants-living-illegally-garden-sheds.html

(2020). Retrieved 9 July 2020, from https://i.pinimg.com/564x/df/21/58/df21584d59bfd6775e4a7699d80e8cb0.jpg

(2020). Retrieved 9 July 2020, from https://i.pinimg.com/564x/73/75/2a/73752a38e6282aba8d1dd71ac25bb927.jpg

(2020). Retrieved 9 July 2020, from https://i.pinimg.com/564x/64/76/43/647643fbd173fd6e0ea90a9c632fa382.jpg

(2020). Retrieved 9 July 2020, from https://i.pinimg.com/564x/e6/b8/8b/e6b88bacbd5e683f81ac8385466e9d99.jpg

(2020). Retrieved 9 July 2020, from https://i.pinimg.com/564x/a2/20/e4/a220e4e9d174d8591e487469539ba467.jpg

(2020). Retrieved 9 July 2020, from https://www.youtube.com/watch?v=X7TljH5x2kE&feature=emb_logo

Redirect Notice. (2020). Retrieved 9 July 2020, from https://www.google.com/url?sa=i&source=imgres&cd=&cad=rja&uact=8&ved=2ahUKEwiV6r-U6KvqAhXR4KYKHU8jA7AQjRx6BAgBEAQ&url=http%3A%2F%2Fenddl22.net%2Fwordpress%2F2012%2F03&psig=AOvVaw0w4iQnyv7y0Sb_omc51caT&ust=1593684559650736

Redirect Notice. (2020). Retrieved 9 July 2020, from https://www.google.com/url?sa=i&source=imgres&cd=&cad=rja&uact=8&ved=2ahUKEwiEnKLj56vqAhVSwcQBHd0sAEwQjRx6BAgBEAQ&url=http%3A%2F%2Fmrl.engin.umich.edu%2FPE_PersOdometry.html&psig=AOvVaw0bvu8SacPGeWbE3eEnpnMC&ust=1593684562547525

(2020). Retrieved 9 July 2020, from https://www.researchgate.net/figure/Filtered-GPS-IMU-position-estimate-versus-VO-position-estimate-The-VO-trajectory_fig2_29642623