Extended Reality in Industry4.0 (ERI)

Lecture 16: 3D reconstruction

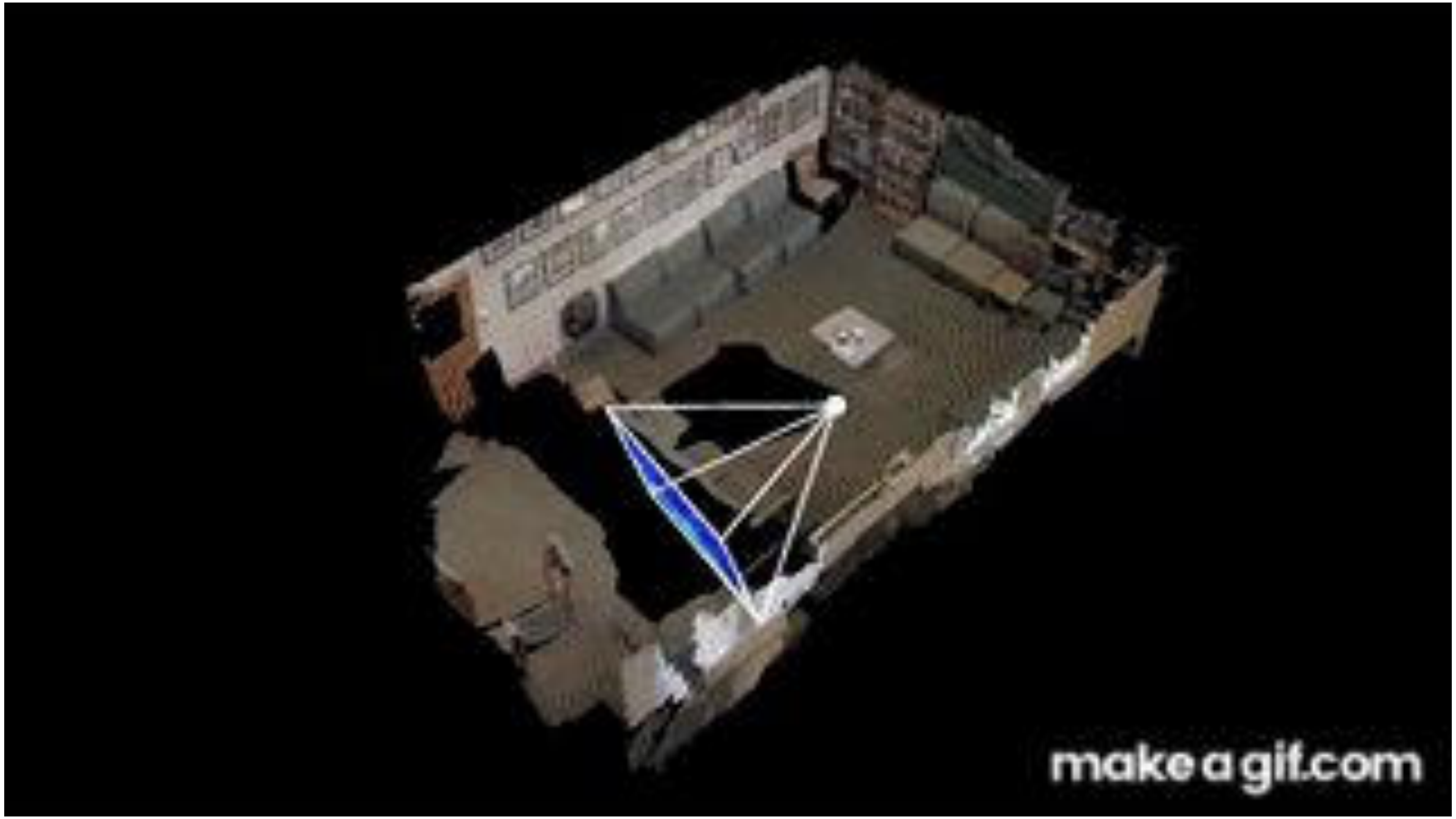
Dr. Kalpana Shankhwar, kalpana@iiitd.ac.in

Assistant Professor

Department of Human Centered Design,
IIIT Delhi

3D reconstruction

3D reconstruction is the process of creating a 3D model or representation of an object or scene from a collection of 2D images or other data sources (i.e. video) taken from known camera viewpoints.



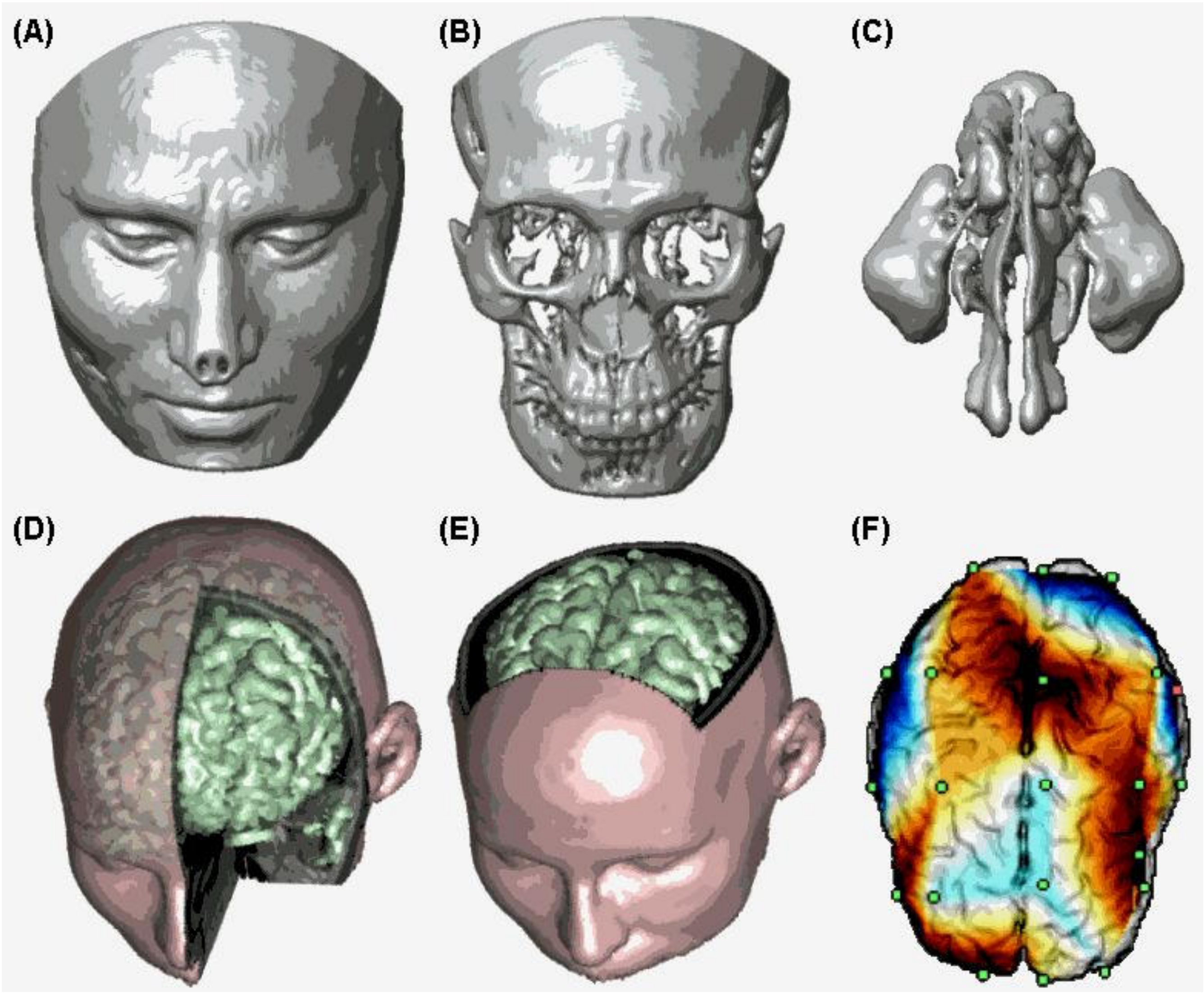
3D reconstruction

- This process relies on various types of algorithms and mathematical models to estimate the geometry and appearance of the objects depicted in the input data.

3D reconstruction applications

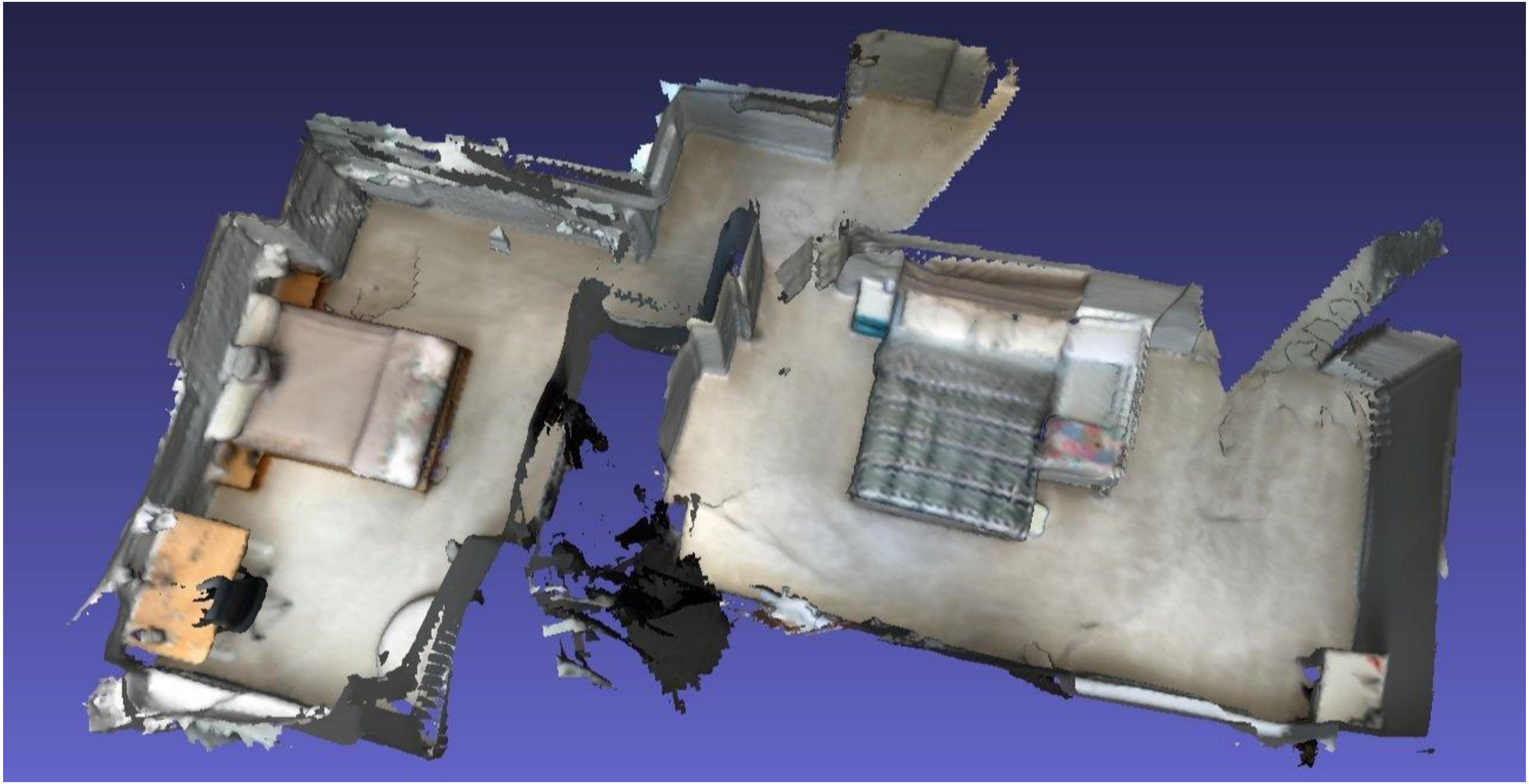
- Medical imaging
- Virtual reality
- Autonomous navigation
- 3D printing
- Metaverse

Medical imaging



Virtual reality

Navigation



Navigation

(a) Start to scan the environment

(c) Continue to scan the environment

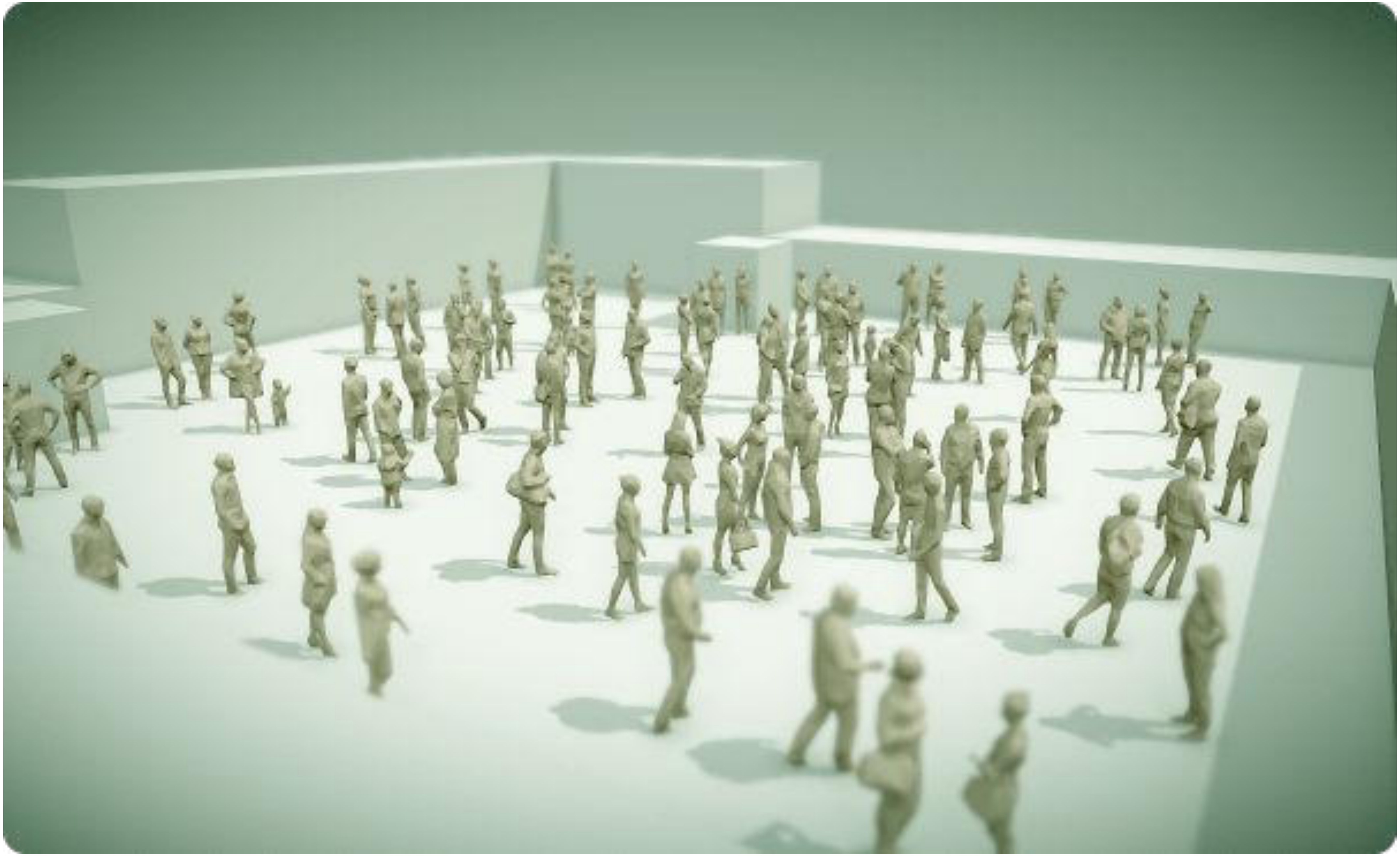
(b) Find and grasp a cube

(d) Detect and avoid obstascles

(e) Find target and place the cube at the target area

3D printing

Metaverse



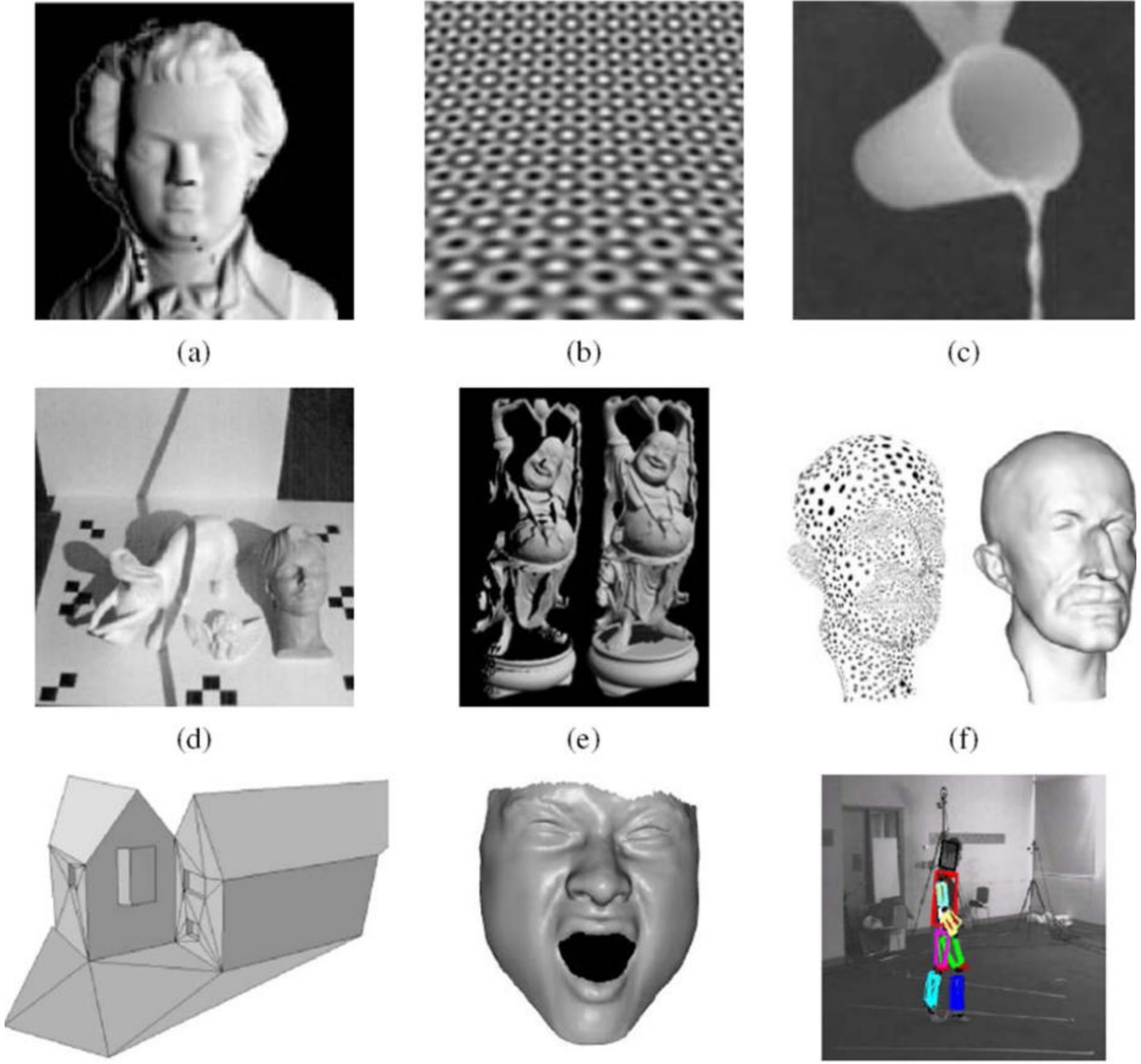


Figure 12.1 3D shape acquisition and modeling techniques: (a) shaded image (Zhang, Tsai, Cryer et al. 1999) c 1999 IEEE; (b) texture gradient (Garding 1992) c 1992 Springer; (c) real-time depth from focus (Nayar, Watanabe, and Noguchi 1996) c 1996 IEEE; (d) scanning a scene with a stick shadow (Bouguet and Perona 1999) c 1999 Springer; (e) merging range maps into a 3D model (Curless and Levoy 1996) c 1996 ACM; (f) point-based surface modeling (Pauly, Keiser, Kobbelt et al. 2003) c 2003 ACM; (g) automated modeling of a 3D building using lines and planes (Werner and Zisserman 2002) c 2002 Springer; (h) 3D face model from spacetime stereo (Zhang, Snavely, Curless et al. 2004) c 2004 ACM; (i)person tracking (Sigal, Bhatia, Roth et al. 2004) c 2004 IEEE.

Shape from shading

- Among the various cues that can be used to infer shape, the shading on a surface can provide a lot of information about **local surface orientations** and hence overall **surface shape**
- This approach becomes even more powerful when lights shining from different directions can be turned on and off separately (photometric stereo).

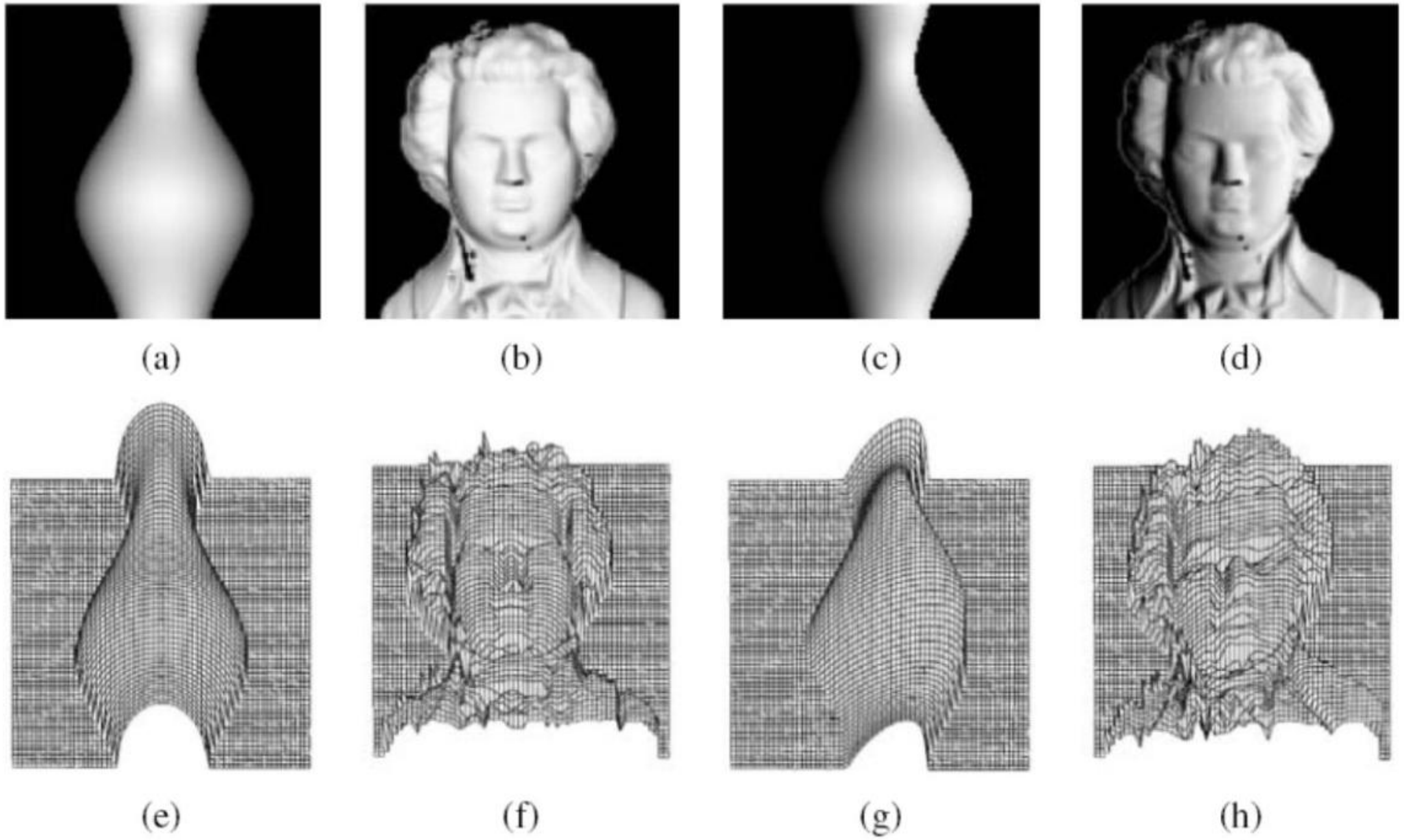
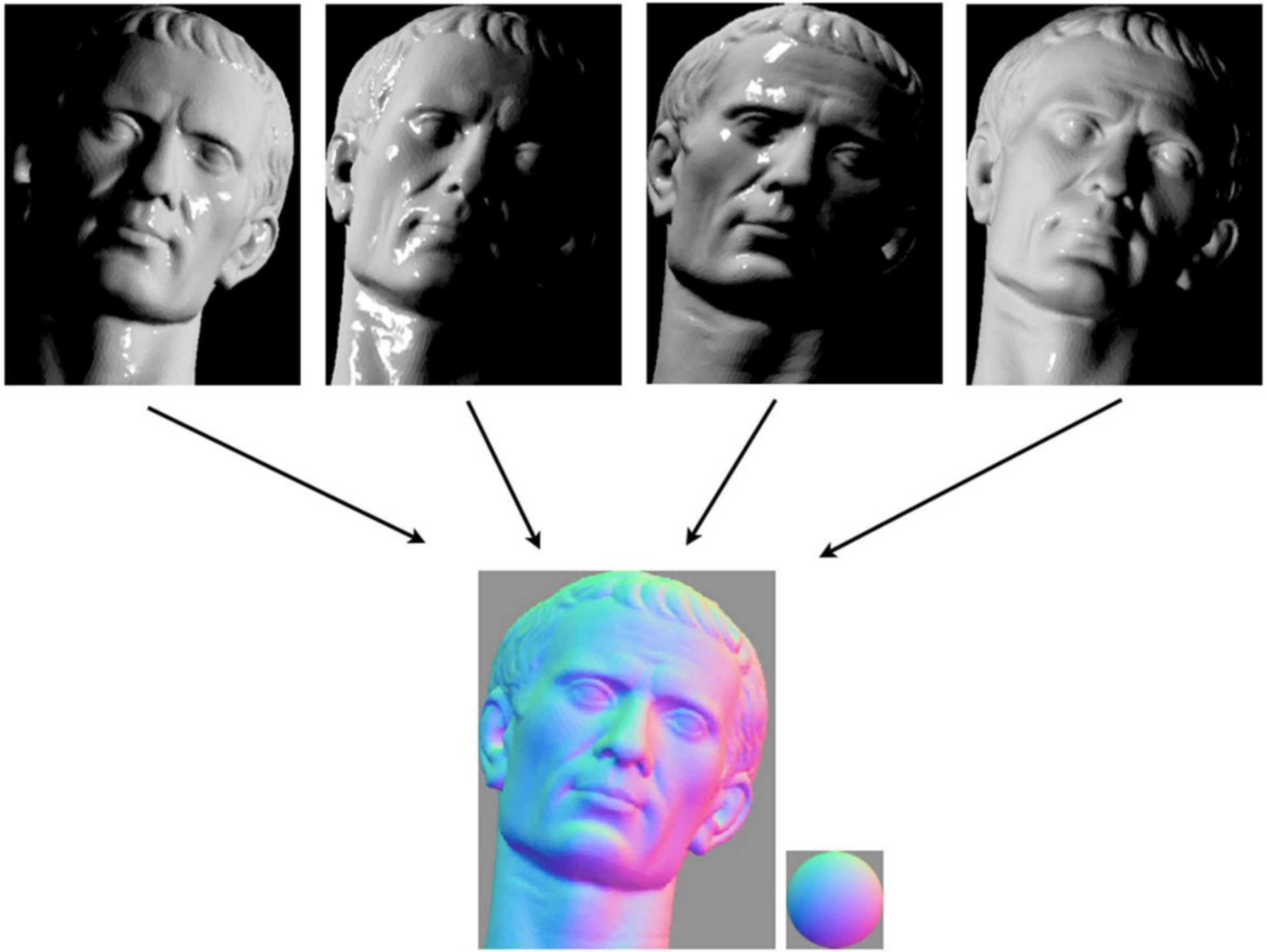


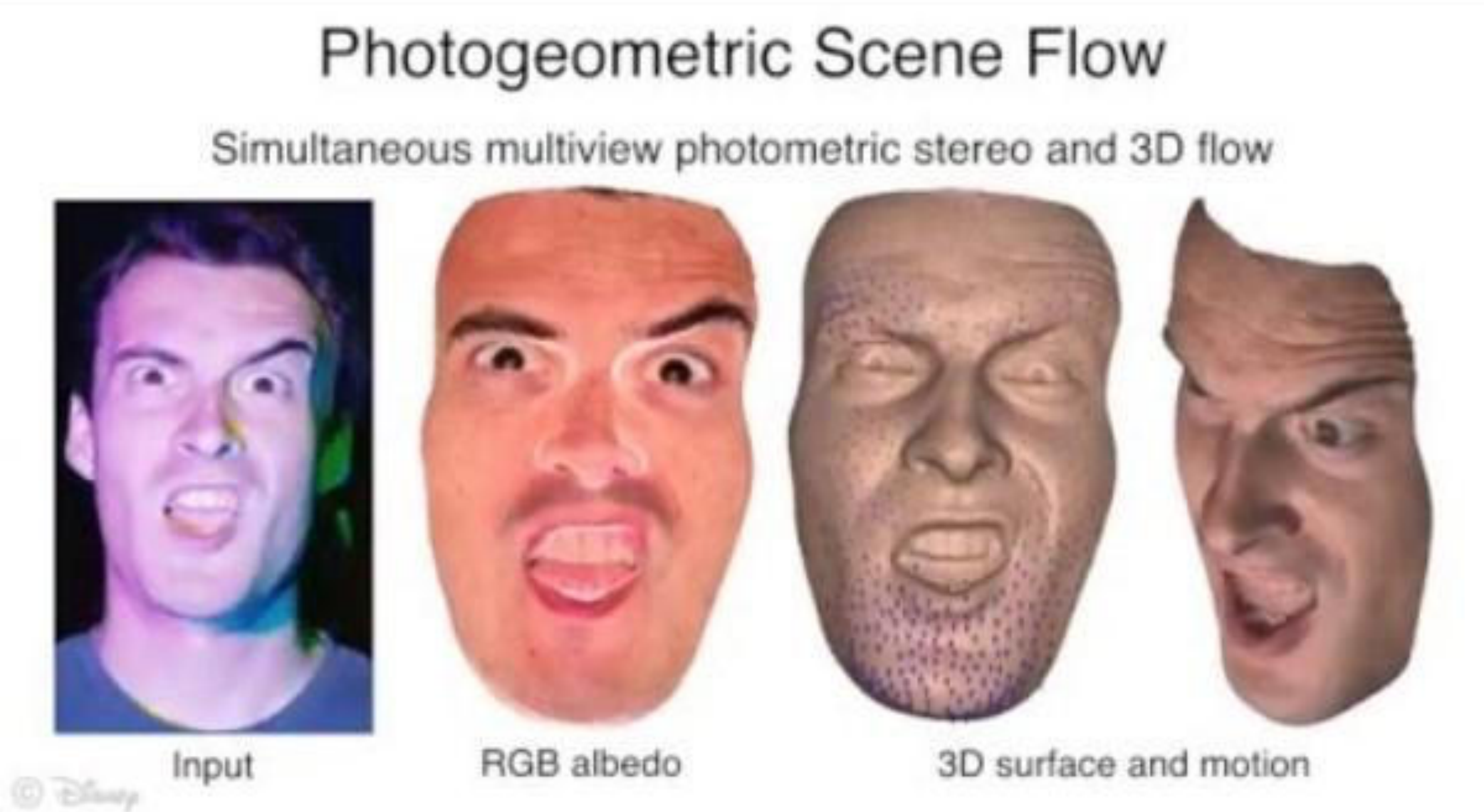
Figure 12.2 Synthetic shape from shading (Zhang, Tsai, Cryer *et al.* 1999) © 1999 IEEE: shaded images, (a–b) with light from in front (0, 0, 1) and (c–d) with light the front right (1, 0, 1); (e–f) corresponding shape from shading reconstructions using the technique of Tsai and Shah (1994).

Photometric stereo

Another way to make shape from shading more reliable is to use **multiple light sources** that can be selectively turned on and off.



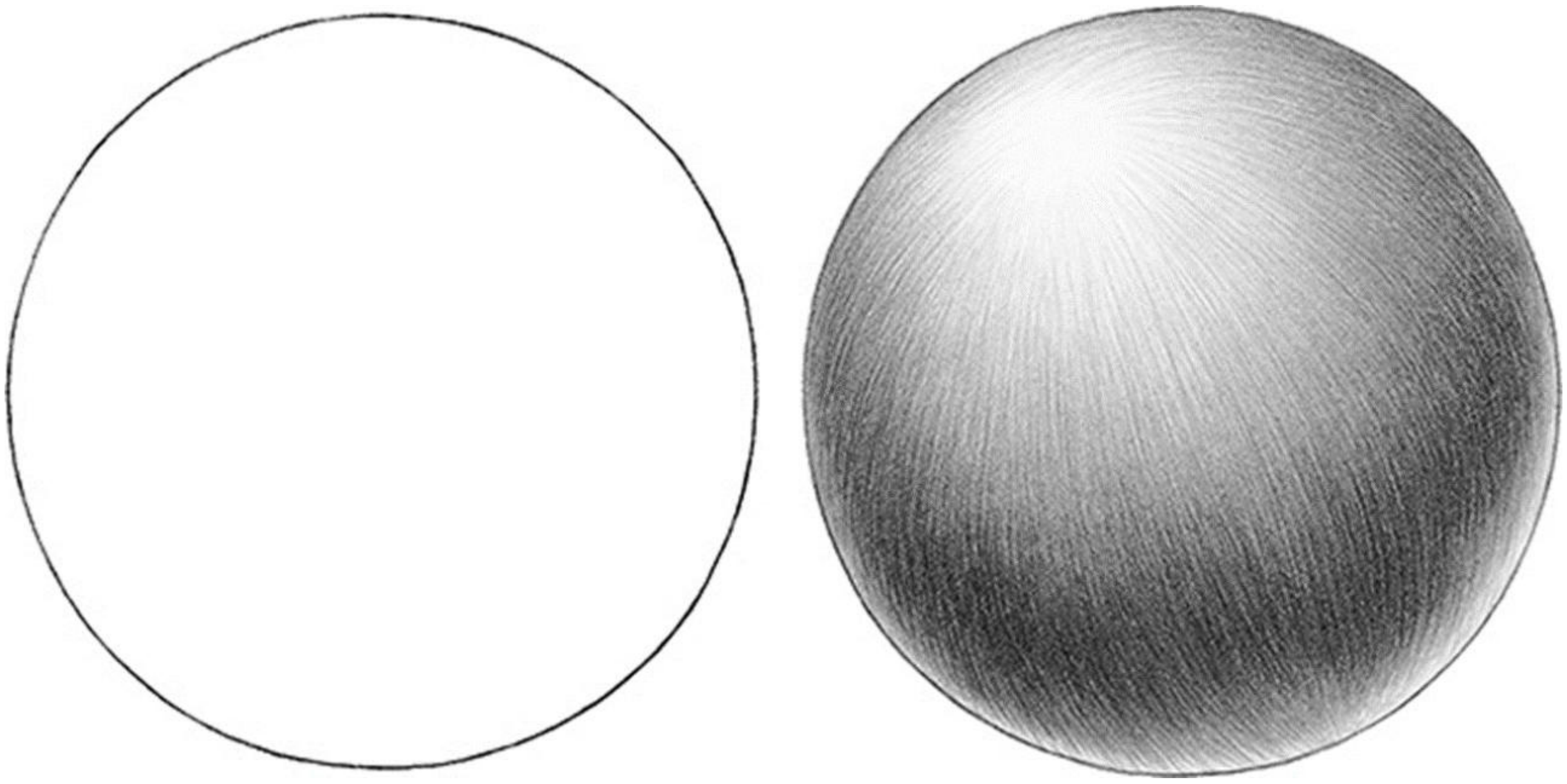
Photometric stereo



Shape from Shading and Photometric Stereo

- **How is this possible?**
  - The answer is that as the surface normal changes across the object, the apparent brightness changes as a function of the angle between the local surface orientation and the incident illumination.
- **Shape from shading**

  The problem of recovering the *shape of a surface* from this *intensity variation* is known as *shape from shading*.

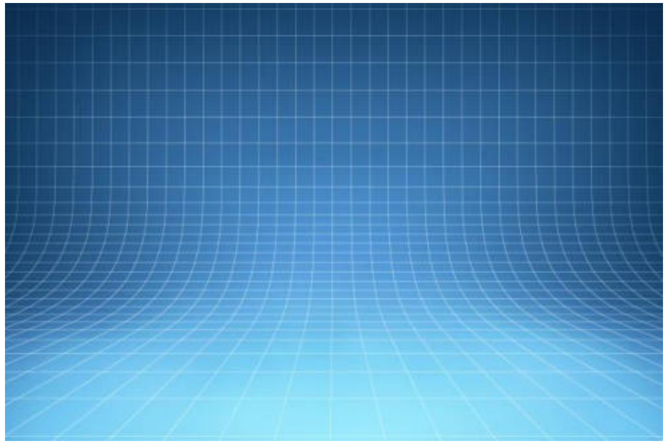


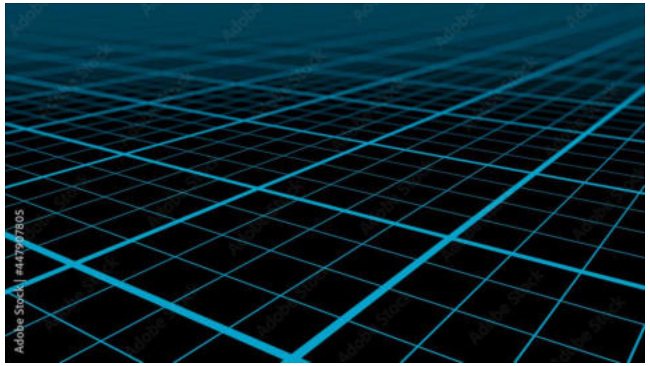
Shape from Texture

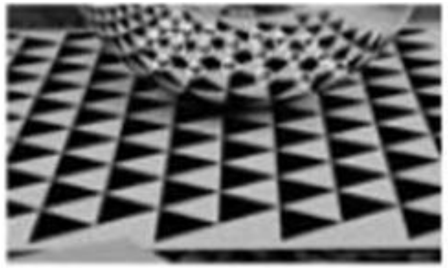
- The variation in foreshortening observed in regular textures can also provide useful information about local surface orientation.

“By using **foreshortening**, an object is presented so it appears to recess or emerge from a particular point in the picture plane. Foreshortening creates an illusion of depth in an image and makes the work more realistic.”

- The deformations induced in a regular pattern when it is viewed in the reflection of a curved mirror can be used to recover the shape of the surface.







Shape from Texture

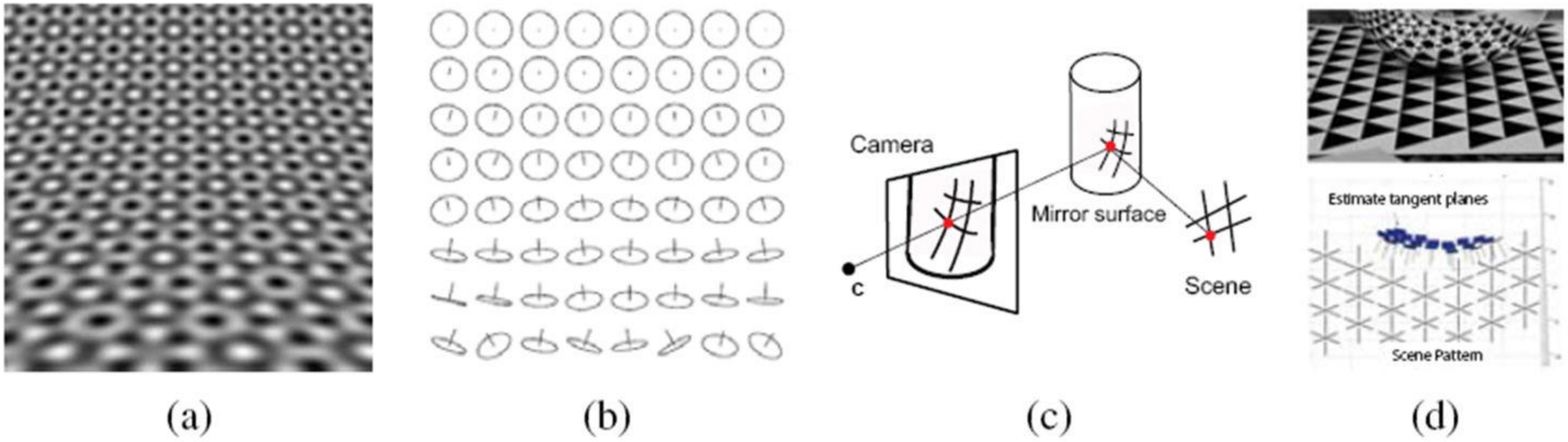
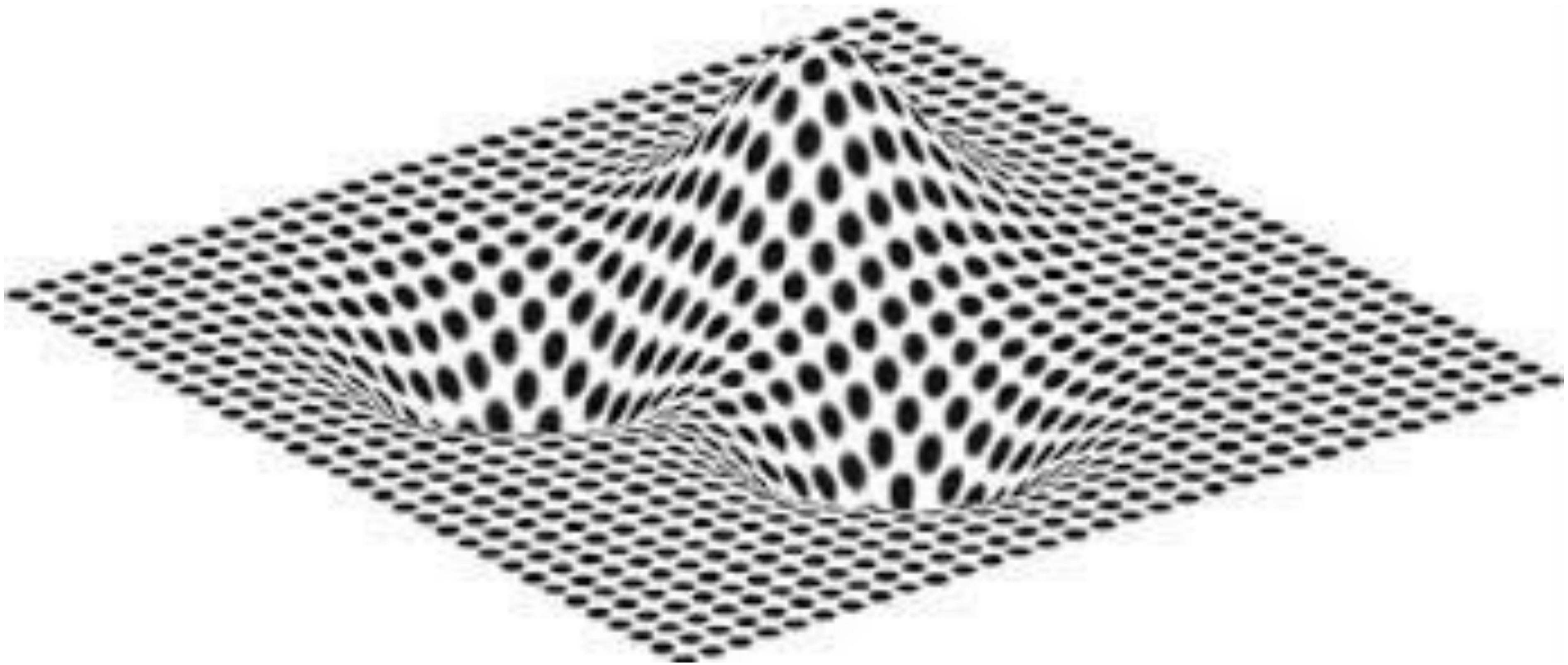


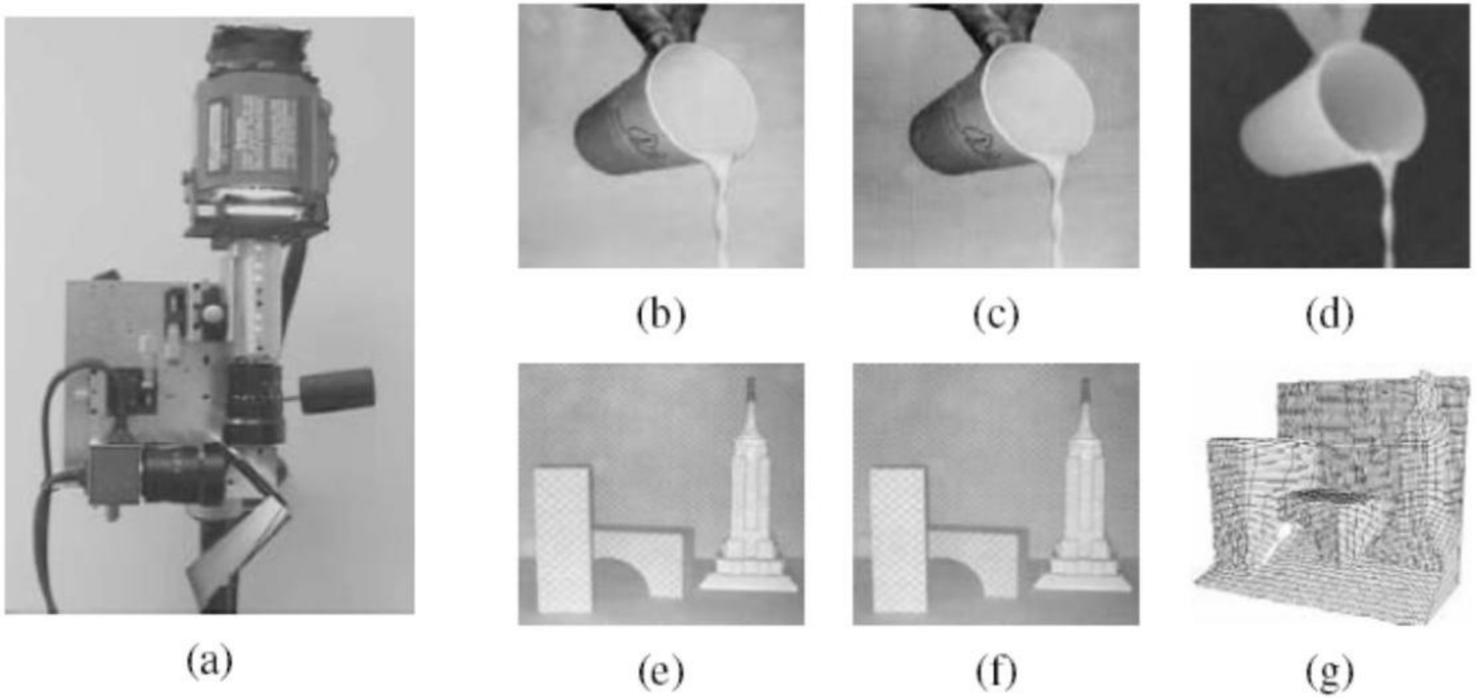
Figure 12.3 Synthetic shape from texture (Garding 1992) © 1992 Springer: (a) regular texture wrapped onto a curved surface and (b) the corresponding surface normal estimates. Shape from mirror reflections (Savarese, Chen, and Perona 2005) © 2005 Springer: (c) a regular pattern reflecting off a curved mirror gives rise to (d) curved lines, from which 3D point locations and normals can be inferred.

Shape from Texture



Shape from focus

- A strong cue for object depth is the amount of blur, which increases as the object's surface moves away from the camera's focusing distance.
- The amount of blur increases in both directions as you move away from the focus plane. Therefore, it is necessary to
  1. use two or more images captured with different focus distance settings
  2. or to translate the object in depth and look for the point of maximum sharpness.



(a) (b) (c) (d) (e) (f) (g)

Figure 12.4 Real time depth from defocus (Nayar, Watanabe, and Noguchi 1996) © 1996 IEEE: (a) the real-time focus range sensor, which includes a half-silvered mirror between the two telecentric lenses (lower right), a prism that splits the image into two CCD sensors (lower left), and an edged checkerboard pattern illuminated by a Xenon lamp (top); (b–c) input video frames from the two cameras along with (d) the corresponding depth map; (e–f) two frames (you can see the texture if you zoom in) and (g) the corresponding 3D mesh model.

Shape from light stripes

- 3D shape can also be estimated using active illumination techniques such as light stripes
- The partial surface models obtained using such techniques can then be merged into more coherent 3D surface models
- Such techniques have been used to construct highly detailed and accurate models of cultural heritage such as historic sites

Shape from light stripes

- One of the most popular active illumination sensors is a laser or light stripe sensor, which sweeps a plane of light across the scene or object while observing it from an offset viewpoint.
- As the stripe falls across the object, it deforms its shape according to the shape of the surface it is illuminating.
- knowledge of the 3D plane equation of the light stripe allows us to infer the 3D location corresponding to each illuminated pixel.
- 3D modeling can be more efficient and effective if we know something about the objects we are trying to reconstruct.

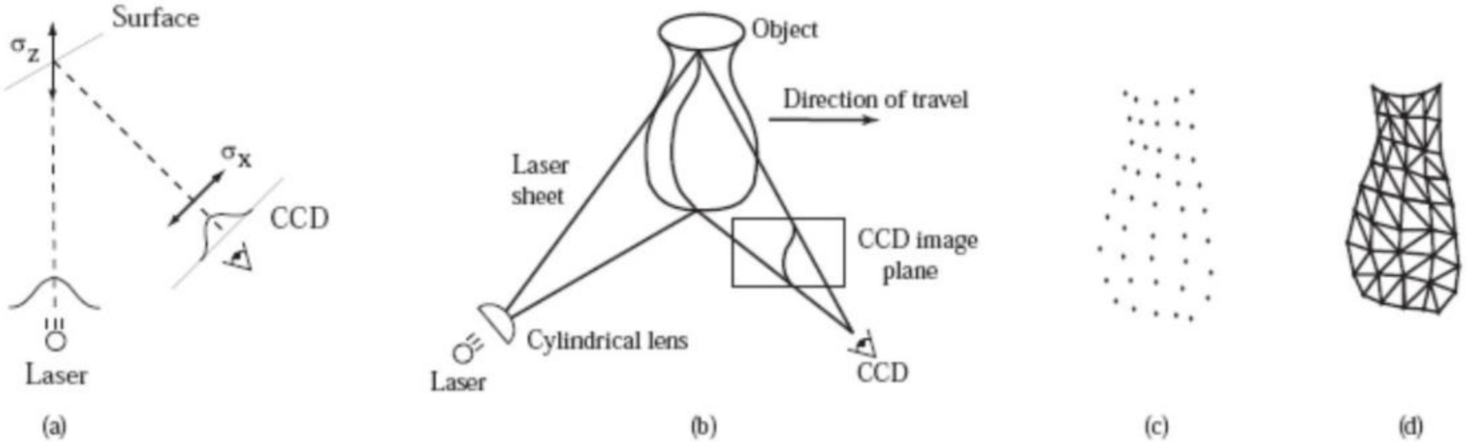


Figure 12.5 Range data scanning (Curless and Levoy 1996) © 1996 ACM: (a) a laser dot on a surface is imaged by a CCD sensor; (b) a laser stripe (sheet) is imaged by the sensor (the deformation of the stripe encodes the distance to the object); (c) the resulting set of 3D points are turned into (d) a triangulated mesh.

Shape from light stripes

- Instead of projecting a light stripe, they simply wave a stick casting a shadow over a scene or object illuminated by a point light source such as a lamp or the sun.
- As the shadow falls across two background planes whose orientation relative to the camera is known, the plane equation for each stripe can be inferred from the two projected lines, whose 3D equations are known

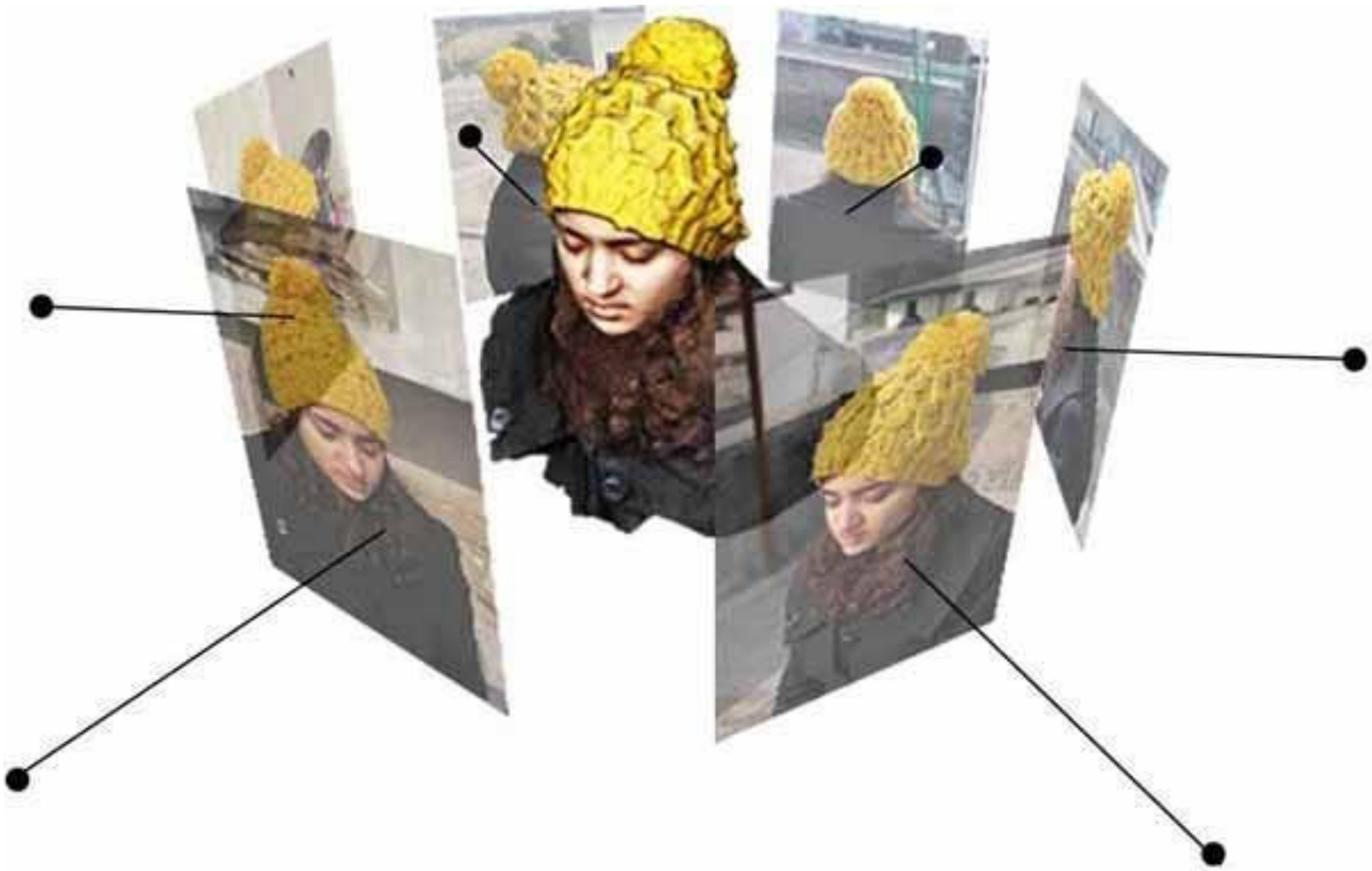
(a) (b) (c)

Figure 12.6 Shape scanning using cast shadows (Bouguet and Perona 1999) © 1999 Springer: (a) camera setup with a point light source (a desk lamp without its reflector), a hand-held stick casting a shadow, and (b) the objects being scanned in front of two planar backgrounds. (c) Real-time depth map using a pulsed illumination system (Iddan and Yahav 2001) © 2001 SPIE.

Multi-view stereo

- Multi-View Stereo (MVS) uses multiple images of a scene taken from different viewpoints to reconstruct a detailed 3D model.
- MVS algorithms take advantage of dense correspondence matching between image pixels to generate a high-resolution 3D point cloud.
- MVS focuses on generating a dense 3D reconstruction of the object. This can take the form of a dense point cloud, a faceted surface (mesh), or a set of planes. These results can be visualized as a realistic 3D rendering of the scene.

Multi-view stereo



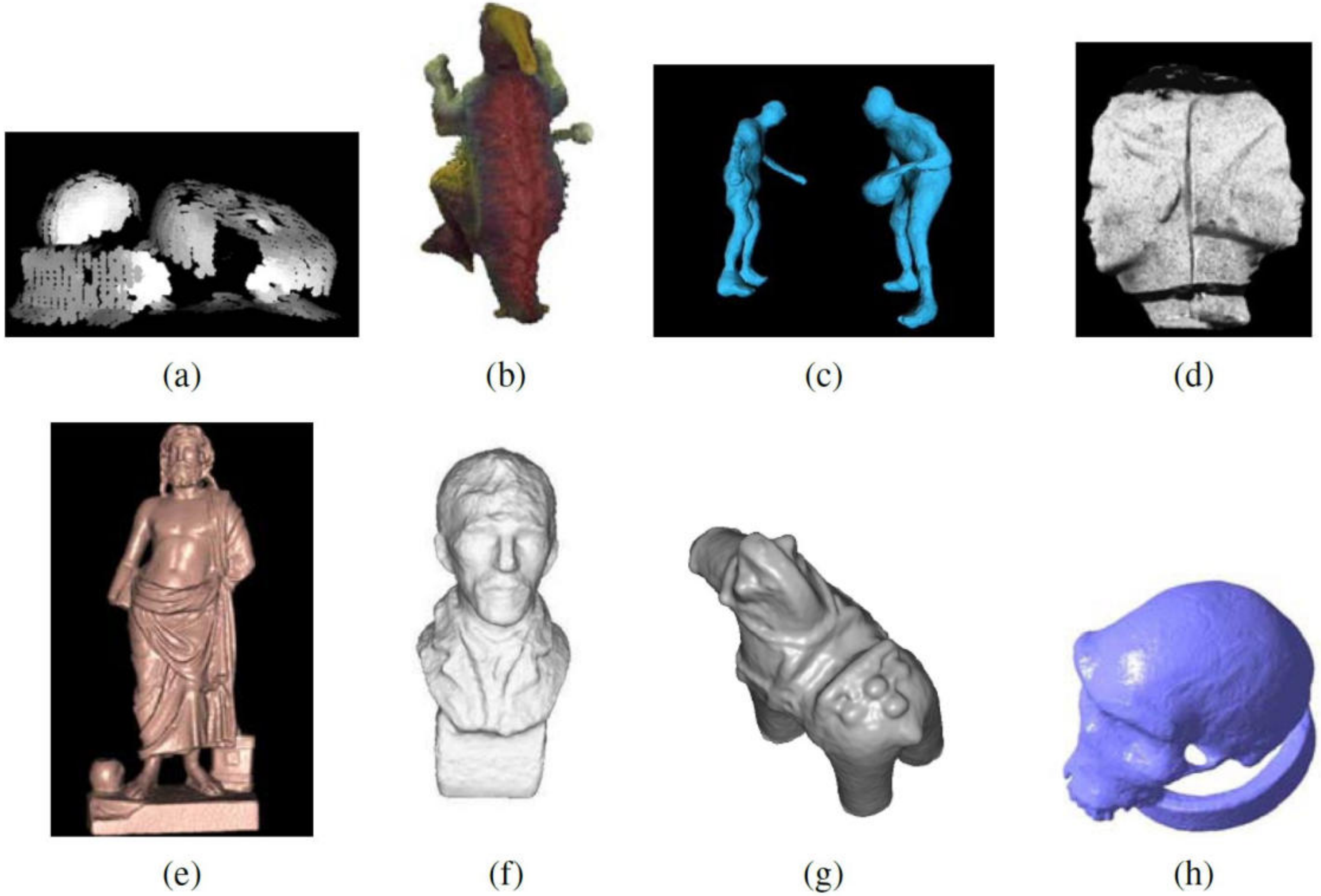


Figure 11.19 Multi-view stereo algorithms: (a) surface-based stereo (Fua and Leclerc 1995); (b) voxel coloring (Seitz and Dyer 1999) © 1999 Springer; (c) depth map merging (Narayanan, Rander, and Kanade 1998); (d) level set evolution (Faugeras and Keriven 1998) © 1998 IEEE; (e) silhouette and stereo fusion (Hernandez and Schmitt 2004) © 2004 Elsevier; (f) multi-view image matching (Pons, Keriven, and Faugeras 2005) © 2005 IEEE; (g) volumetric graph cut (Vogiatzis, Torr, and Cipolla 2005) © 2005 IEEE; (h) carved visual hulls (Furukawa and Ponce 2009) © 2009 Springer.

Image collection

- When reconstructing the 3D configuration of an object, it is essential to capture more than two high-quality images that overlap.
- The image should include crucial information such as the image’s width and height, the camera used, exposure time, and most importantly, the focal length.

Feature extraction

- After collecting overlapped images, the features from the raw data are exacted.
- A feature is a specific characteristic that represents relevant information. Features and their descriptors are extracted from each image. These features can include keypoints (corners, edges, and blobs), lines, or segments.
- Various algorithms are used for feature extraction such as Scale-Invariant Feature Transform.

Feature matching

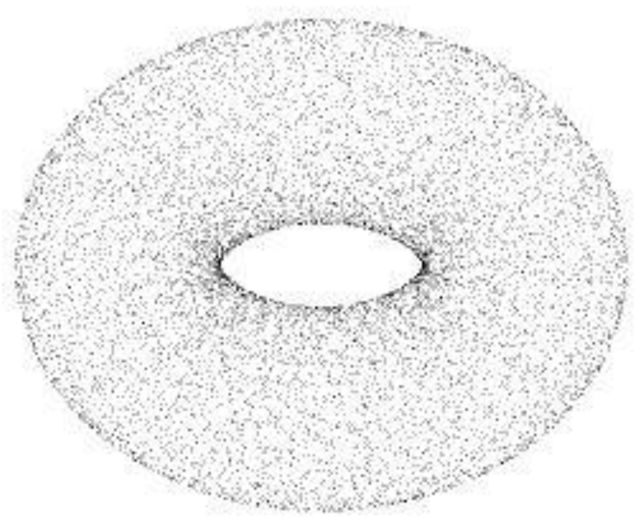
- The next step is identifying corresponding features in multiple images. This step is important in determining the relative orientation of the cameras
- Algorithms you can use for feature matching:
  1. **Exhaustive Matching** — involves comparing each feature in an image with all features in the other images based on a similarity measure
  2. **Nearest-Neighbor Search** — the closest matching feature is found based on a similarity score

Structure from motion

- The objective of the SFM algorithm is to estimate the camera poses and the 3D structure of the scene, given two or more images.
- Camera poses describe the position and orientation of each camera when the image was taken relative to a common coordinate system.
- The pixel coordinates in each image are obtained through feature matching. The ray vectors from each camera center through the pixel coordinates are calculated.
- The intersection point of these two rays in 3D space will be the 3D coordinate (x, y, z) of the feature point in the real world.

Sparse point cloud

- The significance of sparse point clouds lies in their ability to provide an initial framework for more detailed reconstruction. They offer a quick and lightweight representation of the scene, making them ideal for real-time applications like augmented reality, autonomous navigation, and so forth.

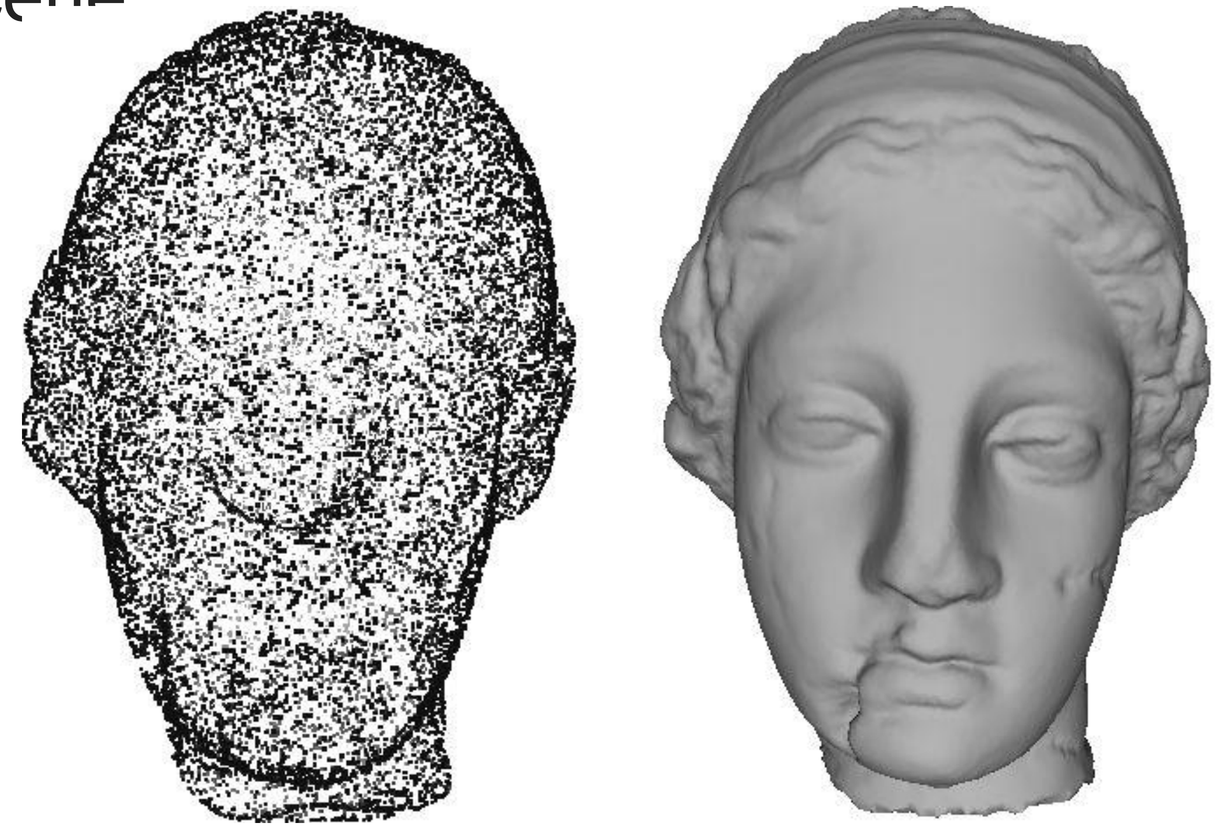


Dense point –cloud reconstruction

- To enhance the richness of a point cloud, depth estimation can be added to fill in the gaps and create denser representations of the scene.
- To compute the **depthmaps**, the images and the camera parameters (pixel size, frame rate, resolution, focal length etc.) are needed. In this step depthmaps are computed for all images, and they will be fused with the sparse point cloud into a 3D mesh.

Mesh reconstruction

- Surface reconstruction algorithms come into play here, creating a continuous surface mesh from the discrete points.
- They aim to smooth and interpolate the point cloud data to generate a surface that accurately represents the shape of the object or scene



Mesh refinement

- Mesh refinement is the process of improving the quality and accuracy of an existing 3D mesh.
- This is often necessary because the initial mesh generated through reconstruction may be rough or contain imperfections.
- This process can involve smoothing to reduce irregularities and noise in the mesh

Mesh texturing

- Mesh texturing allows us to have a final 3D model. Materials' properties and lighting condition are added to the refined mesh, resulting in 3D representation, similar to the real world.

3D reconstruction software (polycam)

https://poly.cam/tools/photogrammetry

Tutorial

https://www.youtube.com/watch?v=HZPAqL0JJiw