Computer stereo vision

Computer stereo vision is the extraction of 3D information from digital images, such as obtained by a CCD camera. By comparing information about a scene from two vantage points, 3D information can be extracted by examination of the relative positions of objects in the two panels. This is similar to the biological process Stereopsis.

1 Outline

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 disparities, which are inversely proportional to the differences in distance to the objects.

To compare the images, the two views must be superimposed in a stereoscopic device, the image from the right camera being shown to the observer's right eye and from the left one to the left eye.

In real camera systems however, several pre-processing steps are required.[1]

1. The image must first be removed of distortions, such as barrel distortion to ensure that the observed image is purely projectional.
2. The image must be projected back to a common plane to allow comparison of the image pairs, known as image rectification.
3. An information measure which compares the two images is minimized. This gives the best estimate of the position of features in the two images, and creates a disparity map.
4. Optionally, the disparity as observed by the common projection, is converted back to the height map by inversion. Utilising the correct proportionality constant, the height map can be calibrated to provide exact distances.

2 Active stereo vision

The active stereo vision is a form of stereo vision which actively employs a light such as a laser or a structured light to simplify the stereo matching problem. The opposed term is passive stereo vision.

2.1 Conventional structured-light vision (SLV)

The conventional structured-light vision (SLV) employs a structured light or laser, and finds projector-camera correspondences.[2][3]

2.2 Conventional active stereo vision (ASV)

The conventional active stereo vision (ASV) employs a structured light or laser, however, the stereo matching is performed only for camera-camera correspondences, in the same way as the passive stereo vision.

2.3 Structured-light stereo (SLS)[4]

There is a hybrid technique, which utilizes both camera-camera and projector-camera correspondences.[4]

3 Applications

3D stereo displays finds many applications in entertainment, information transfer and automated systems. Stereo vision is highly important in fields such as robotics, to extract information about the relative position of 3D objects in the vicinity of autonomous systems. Other applications for robotics include object recognition, where depth information allows for the system to separate occluding image components, such as one chair in front of another, which the robot may otherwise not be able to distinguish as a separate object by any other criteria.

Scientific applications for digital stereo vision include the extraction of information from aerial surveys, for calculation of contour maps or even geometry extraction for 3D building mapping, or calculation of 3D heliographical information such as obtained by the NASA STEREO project.

4 Detailed definition

Main article: Triangulation (computer vision)

A pixel records color at a position. The position is identified by position in the grid of pixels (x, y) and depth to the pixel z.

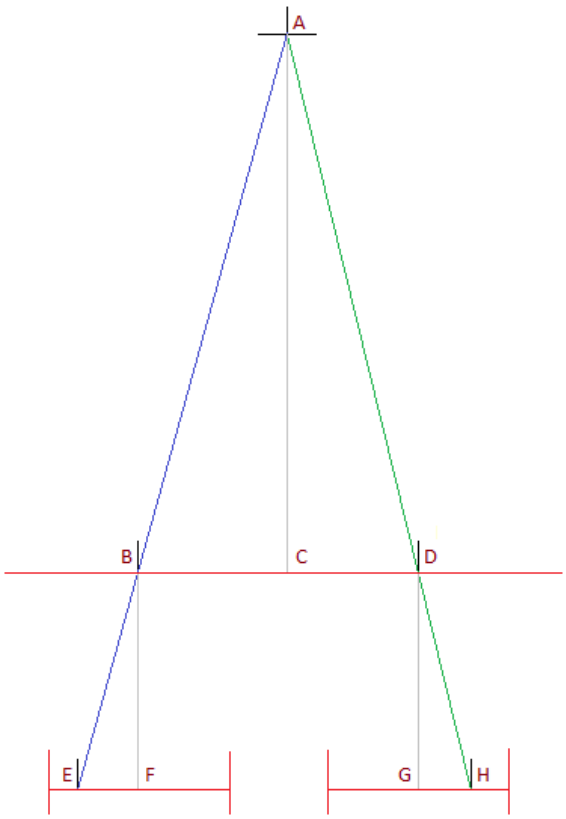


Diagram describing relationship of image displacement to depth with stereoscopic images, assuming flat co-planar images.

Stereoscopic vision gives two images of the same scene, from different positions. In the diagram on the right light from the point *A* is transmitted through the entry points of a pinhole cameras at *B* and *D*, onto image screens at *E* and *H*.

In the attached diagram the distance between the centers of the two camera lens is *BD* = *BC* + *CD*. The triangles are similar,

- *ACB* and *BFE*
- *ACD* and *DGH*

$$\begin{aligned}\text{displacement Therefore} d &= EF + GH \\ &= BF(\frac{EF}{BF} + \frac{GH}{BF}) \\ &= BF(\frac{EF}{BF} + \frac{GH}{DG}) \\ &= BF(\frac{BC + CD}{AC}) \\ &= BF\frac{BD}{AC} \\ &= \frac{k}{z}\text{where ,}\end{aligned}$$

- *k* = *BD BF*
- *z* = *AC* is the distance from the camera plane to the object.

So assuming the cameras are level, and image planes are flat on the same plane, the displacement in the y axis between the same pixel in the two images is,

$$d = \frac{k}{z}$$

Where k is the distance between the two cameras times the distance from the lens to the image.

The depth component in the two images are z_1 and z_2 , given by,

$$z_2(x,y) = \min\left\{v : v = z_1(x, y - \frac{k}{z_1(x,y)})\right\}$$

$$z_1(x,y) = \min\left\{v : v = z_2(x, y + \frac{k}{z_2(x,y)})\right\}$$

These formulas allow for the occlusion of voxels, seen in one image on the surface of the object, by closer voxels seen in the other image, on the surface of the object.

4.1 Image Rectification

Where the image planes are not co-planar image rectification is required to adjust the images as if they were co-planar. This may be achieved by a linear transformation.

The images may also need rectification to make each image equivalent to the image taken from a pinhole camera projecting to a flat plane.

4.2 Least squares information measure

The normal distribution is

$$P(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability is related to information content described by message length L,

$$P(x) = 2^{-L(x)}$$

$$L(x) = -\log_2 P(x)$$

so,

$$L(x, \mu, \sigma) = \log_2(\sigma\sqrt{2\pi}) + \frac{(x-\mu)^2}{2\sigma^2}\log_2 e$$

For the purposes of comparing stereoscopic images, only the relative message length matters. Based on this, the

information measure *I*, called the Sum of Squares of Differences (SSD) is,

$$I(x,\mu,\sigma) = \frac{(x-\mu)^2}{\sigma^2}$$

where,

$$L(x,\mu,\sigma) = \log_2(\sigma\sqrt{2\pi}) + I(x,\mu,\sigma)\frac{\log_2 e}{2}$$

4.3 Other measures of information content

Because of the cost in processing time of squaring numbers in SSD, many implementations use Sum of Absolute Difference (SAD) as the basis for computing the information measure. Other methods use normalized cross correlation (NCC).

4.4 Information measure for stereoscopic images

The least squares measure may be used to measure the information content of the stereoscopic images ,[5] given depths at each point $z(x,y)$. Firstly the information needed to express one image in terms of the other is derived. This is called I_m .

A color difference function should be used to fairly measure the difference between colors. The color difference function is written *cd* in the following. The measure of the information needed to record the color matching between the two images is,

$$I_m(z_1,z_2) = \frac{1}{\sigma_m^2}\sum_{x,y} \mathrm{cd}(\mathrm{color}_1(x,y+\frac{k}{z_1(x,y)}),\mathrm{color}_2(x,y))^2$$

An assumption is made about the smoothness of the image. Assume that two pixels are more likely to be the same color, the closer the voxels they represent are. This measure is intended to favor colors that are similar being grouped at the same depth. For example if an object in front occludes an area of sky behind, the measure of smoothness favors the blue pixels all being grouped together at the same depth.

The total measure of smoothness uses the distance between voxels as an estimate of the expected standard deviation of the color difference,

$$I_s(z_1,z_2) = \frac{1}{2\sigma_h^2}\sum_{i:\{1,2\}}\sum_{x_1,y_1}\sum_{x_2,y_2}\frac{\mathrm{cd}(\mathrm{color}_i(x_1,y_1),\mathrm{color}_i(x_2,y_2))^2}{(x_1-x_2)^2+(y_1-y_2)^2+(z_i(x_1,y_1)-z_i(x_2,y_2))^2}$$

The total information content is then the sum,

$$I_t(z_1,z_2) = I_m(z_1,z_2) + I_s(z_1,z_2)$$

The z component of each pixel must be chosen to give the minimum value for the information content. This will give the most likely depths at each pixel. The minimum total information measure is,

$$I_{\min} = \min\{i : i = I_t(z_1,z_2)\}\}$$

The depth functions for the left and right images are the pair,

$$(z_1,z_2) \in \{(z_1,z_2) : I_t(z_1,z_2) = I_{\min}\}$$

4.5 Smoothness

Smoothness is a measure of how similar colors that are close together are. There is an assumption that objects are more likely to be colored with a small number of colors. So if we detect two pixels with the same color they most likely belong to the same object.

The method described above for evaluating smoothness is based on information theory, and an assumption that the influence of the color of a voxel influencing the color of nearby voxels according to the normal distribution on the distance between points. The model is based on approximate assumptions about the world.

Another method based on prior assumptions of smoothness is auto-correlation.

Smoothness is a property of the world. It is not inherently a property of an image. For example an image constructed of random dots would have no smoothness, and inferences about neighboring points would be useless. Theoretically smoothness, along with other properties of the world should be learnt. This appears to be what the human vision system does.

5 Methods of implementation

The minimization problem is NP-complete. This means a general solution to this problem will take a long time to reach a solution. However methods exist for computers based on heuristics that approximate the result in a reasonable amount of time. Also methods exist based on neural networks.[6] Efficient implementation of stereoscopic vision is an area of active research.

6 See also

- 3D scanner

- Autostereoscopy
- Epipolar geometry
- Computer vision
- Stereo camera
- Stereoscopic Depth Rendition
- Stereopsis
- Stereophotogrammetry
- 3D reconstruction from multiple images

7 References

[1] Bradski, Gary and Kaehler, Adrian. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly.

[2] C. Je, S. W. Lee, and R.-H. Park. High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging. Computer Vision – ECCV 2004, LNCS 3021, pp. 95–107, Springer-Verlag Berlin Heidelberg, May 10, 2004.

[3] C. Je, S. W. Lee, and R.-H. Park. Colour-Stripe Permutation Pattern for Rapid Structured-Light Range Imaging. Optics Communications, Volume 285, Issue 9, pp. 2320-2331, May 1, 2012.

[4] W. Jang, C. Je, Y. Seo, and S. W. Lee. Structured-Light Stereo: Comparative Analysis and Integration of Structured-Light and Active Stereo for Measuring Dynamic Shape. Optics and Lasers in Engineering, Volume 51, Issue 11, pp. 1255-1264, November, 2013.

[5] Lazaros, Nalpantidis; Sirakoulis, Georgios Christou; Gasteratos1, Antonios (2008). "REVIEW OF STEREO VISION ALGORITHMS: FROM SOFTWARE TO HARDWARE". *International Journal of Optomechatronics* **2**: 435–462. doi:10.1080/15599610802438680.

[6] WANG, JUNG-HUA; HSIAO, CHIH-PING (1999). *Proc. Natl. Sci. Counc. ROC(A)* **23** (5): 665–678. Missing or empty |title= (help)

8 External links

- Tutorial on uncalibrated stereo vision
- Learn about stereo vision with MATLAB
- Stereo Vision and Rover Navigation Software for Planetary Exploration

9 Text and image sources, contributors, and licenses

9.1 Text

- **Computer stereo vision** *Source:* https://en.wikipedia.org/wiki/Computer_stereo_vision?oldid=677195163 *Contributors:* Richie, Sgauria, Toussaint, Kri, Banus, KYN, Gilliam, Mion, Magioladitis, User A1, Cubic333, 1ForTheMoney, WikHead, Fgnievinski, Fotaun, Gvitaly, Onoes, Gwestheimer, Thepigdog, Fluffystar, Avi.nehemiah, JuliusHuelsmann and Anonymous: 13

9.2 Images

- **File:Stereoscopic_images,_depth_to_displacement_relationship_assuming_flat_co-planar_images..png** *Source:* https://upload.wikimedia.org/wikipedia/commons/b/b9/Stereoscopic_images%2C_depth_to_displacement_relationship_assuming_flat_co-planar_images..png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Thepigdog

9.3 Content license

- Creative Commons Attribution-Share Alike 3.0