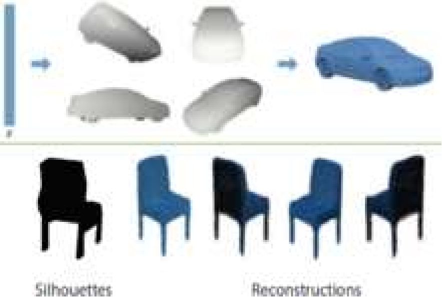
3D reconstruction from multiple images

3D reconstruction from multiple images is the creation of three-dimensional models from a set of images. It is the reverse process of obtaining 2D images from 3D scenes.

A 3D selfie in 1:20 scale printed by Shapeways using gypsum-based printing, created by Madurodam miniature park from 2D pictures taken at its Fantasitron photo booth

3D models are generated from 2D pictures taken at the Fantasitron 3D photo booth at Madurodam.

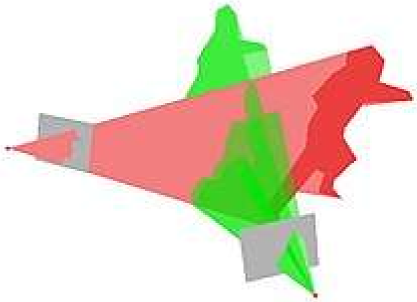


Generating and reconstructing 3D shapes from single or multi-view depth maps or silhouettes[1]

The essence of an image is a projection from a 3D scene onto a 2D plane, during which process the depth is lost. The 3D point corresponding to a specific image point is constrained to be on the line of sight. From a single image, it is impossible to determine which point on this line corresponds to the image point. If two images are available, then the position of a 3D point can be found as the intersection of the two projection rays. This process is referred to as triangulation. The key for this process is the relations between multiple views which convey the information that corresponding sets of points must contain some structure and that this structure is related to the poses and the calibration of the camera.

In recent decades, there is an important demand for 3D content for computer graphics, virtual reality and communication, triggering a change in emphasis for the requirements. Many existing systems for constructing 3D models are built around specialized hardware (e.g. stereo rigs) resulting in a high cost, which cannot satisfy the requirement of its new applications. This gap stimulates the use of digital imaging facilities (like a camera). An early method was proposed by Tomasi and Kanade.[2] They used an affine factorization approach to extract 3D from images sequences. However, the assumption of orthographic projection is a significant limitation of this system.

Processing



A *visual hull* can be reconstructed from multiple silhouettes of an object.[3]

The task of converting multiple 2D images into 3D model consists of a series of processing steps:

Camera calibration consists of intrinsic and extrinsic parameters, without which at some level no arrangement of algorithms can work. The dotted line between Calibration and Depth determination represents that the camera calibration is usually required for determining depth.

Depth determination serves as the most challenging part in the whole process, as it calculates the 3D component missing from any given image – depth. The correspondence problem, finding matches between two images so the position of the matched elements can then be triangulated in 3D space is the key issue here.

Once you have the multiple depth maps you have to combine them to create a final mesh by calculating depth and projecting out of the camera – **registration**. Camera calibration will be used to identify where the many meshes created by depth maps can be combined to develop a larger one, providing more than one view for observation.

By the stage of **Material Application** you have a complete 3D mesh, which may be the final goal, but usually you will want to apply the color from the original photographs to the mesh. This can range from projecting the images onto the mesh randomly, through approaches of combining the textures for super resolution and finally to segmenting the mesh by material, such as specular and diffuse properties.

Mathematical description of reconstruction

Given a group of 3D points viewed by N cameras with matrices $\{P^i\}_{i=1\ldots N}$, define $m_j^i \simeq P^i w_j$ to be the homogeneous coordinates of the projection of the j^{th} point onto the i^{th} camera. The reconstruction problem can be changed to: given the group of pixel coordinates $\{m_j^i\}$, find the corresponding set of camera matrices $\{P^i\}$ and the scene structure $\{w_j\}$ such that

$m_j^i \simeq P^i w_j$ (1)

Generally, without further restrictions, we will obtain a projective reconstruction.[4][5] If $\{P^i\}$ and $\{w_j\}$ satisfy (1), $\{P^i T\}$ and $\{T^{-1} w_j\}$ will satisfy (1) with any **4 × 4** nonsingular matrix **T**.

A projective reconstruction can be calculated by correspondence of points only without any *a priori* information.

Auto-calibration

In **auto-calibration** or **self-calibration**, camera motion and parameters are recovered first, using rigidity. Then structure can be readily calculated. Two methods implementing this idea are presented as follows:

Kruppa equations

With a minimum of three displacements, we can obtain the internal parameters of the camera using a system of polynomial equations due to Kruppa,[6] which are derived from a geometric interpretation of the rigidity constraint.[7][8]

The matrix $K = AA^\top$ is unknown in the Kruppa equations, named Kruppa coefficients matrix. With **K** and by the method of Cholesky factorization one can obtain the intrinsic parameters easily:

$$K = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_2 & k_4 & k_5 \\ k_3 & k_5 & 1 \end{bmatrix}$$

Recently Hartley [9] proposed a simpler form. Let F be written as $F = DUV^\top$, where

Then the Kruppa equations are rewritten (the derivation can be found in [9])

Mendonça and Cipolla

This method is based on the use of rigidity constraint. Design a cost function, which considers the intrinsic parameters as arguments and the fundamental matrices as parameters. F_{ij} is defined as the fundamental matrix, A_i and A_j as intrinsic parameters matrices.

Stratification

Recently, new methods based on the concept of **stratification** have been proposed.[10] Starting from a projective structure, which can be calculated from correspondences only, upgrade this projective reconstruction to a Euclidean reconstruction, by making use of all the available constraints. With this idea the problem can be stratified into different sections: according to the amount of constraints available, it can be analyzed at a different level, projective, affine or Euclidean.

The stratification of 3D geometry

Usually, the world is perceived as a 3D Euclidean space. In some cases, it is not possible to use the full Euclidean structure of 3D space. The simplest being projective, then the affine geometry which forms the intermediate layers and finally Euclidean geometry. The concept of stratification is closely related to the series of transformations on geometric entities: in the projective stratum is a series of projective transformations (a homography), in the affine stratum is a series of affine transformations, and in Euclidean stratum is a series of Euclidean transformations.

Suppose that a fixed scene is captured by two or more perspective cameras and the correspondences between visible points in different images are already given. However, in practice, the matching is an essential and extremely challenging issue in computer vision. Here, we suppose that n 3D points A_i are observed by m cameras with projection matrices $P_j, j = 1, \ldots, m.$ Neither the positions of point nor the projection of camera are known. Only the projections a_{ij} of the i^{th} point in the j^{th} image are known.

Projective reconstruction

Simple counting indicates we have $2nm$ independent measurements and only $11m + 3n$ unknowns, so the problem is supposed to be soluble with enough points and images. The equations in homogeneous coordinates can be represented:

$$a_{ij} \sim P_j A_i \qquad i = 1, \ldots n, \;\; j = 1, \ldots m \; (2)$$

So we can apply a nonsingular **4 × 4** transformation *H* to projections $P_j \rightarrow P_j H^{-1}$ and world points $A_i \rightarrow H A_i$. Hence, without further constraints, reconstruction is only an unknown projective deformation of the 3D world.

Affine reconstruction

See affine space *for more detailed information about computing the location of the plane at infinity* Π_∞. The simplest way is to exploit prior knowledge, for example the information that lines in the scene are parallel or that a point is the one thirds between two others.

We can also use prior constraints on the camera motion. By analyzing different images of the same point can obtain a line in the direction of motion. The intersection of several lines is the point at infinity in the motion direction, and one constraint on the affine structure.

Euclidean reconstruction

By mapping the projective reconstruction to one that satisfies a group of redundant Euclidean constraints, we can find a projective transformation *H* in equation (2).The equations are highly nonlinear and a good initial guess for the structure is required. This can be obtained by assuming a linear projection - parallel projection, which also allows easy reconstruction by SVD decomposition.[2]

Algebraic vs geometric error

Inevitably, measured data (i.e., image or world point positions) is noisy and the noise comes from many sources. To reduce the effect of noise, we usually use more equations than necessary and solve with least squares.

For example, in a typical null-space problem formulation Ax = 0 (like the DLT algorithm), the square of the residual ||Ax|| is being minimized with the least squares method.

In general, if ||Ax|| can be considered as a distance between the geometrical entities (points, lines, planes, etc.), then what is being minimized is a **geometric error**, otherwise (when the error lacks a good geometrical interpretation) it is called an **algebraic error**.

Therefore, compared with algebraic error, we prefer to minimize a geometric error for the reasons listed:

1. The quantity being minimized has a meaning.

2. The solution is more stable.
3. The solution is constant under Euclidean transforms.

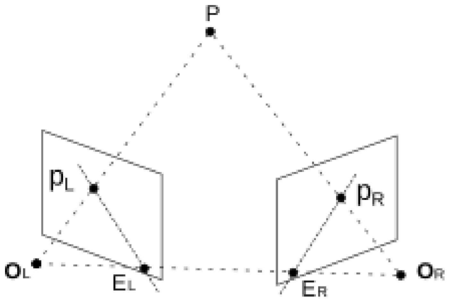
All the linear algorithms (DLT and others) we have seen so far minimize an algebraic error. Actually, there is no justification in minimizing an algebraic error apart from the ease of implementation, as it results in a linear problem. The minimization of a geometric error is often a non-linear problem, that admit only iterative solutions and requires a starting point.

Usually, linear solution based on algebraic residuals serves as a starting point for a non-linear minimization of a geometric cost function, which provides the solution a final "polish".[11]

Medical applications

The 2-D imaging has problems of anatomy overlapping with each other and do not disclose the abnormalities. The 3-D imaging can be used for both diagnostic and therapeutic purposes.

3-D models are used for planning the operation, morphometric studies and has more reliability in orthopedics.[12]



Projection of P on both cameras

Problem statement & Basics

To reconstruct 3-D images from 2-D images taken by a camera at multiple angles. Medical imaging techniques like CT scanning and MRI are expensive, and although CT scans are accurate, they can induce high radiation doses which is a risk for patients with certain diseases. Methods based on MRI are not accurate. Since we are exposed to powerful magnetic fields during an MRI scan, this method is not suitable for patients with ferromagnetic metallic implants. Both the methods can be done only when in lying position where the global structure of the bone changes. So, we discuss the following methods which can be performed while standing and require low radiation dose.

Though these techniques are 3-D imaging, the region of interest is restricted to a slice; data are acquired to form a time sequence.

Stereo Corresponding Point Based Technique

This method is simple and implemented by identifying the points manually in multi-view radiographs. The first step is to extract the corresponding points in two x-ray images. The second step is to reconstruct the image in three dimensions using algorithms like Discrete Linear Transform (DLT).[13] The reconstruction is only possible where there are Stereo Corresponding Points (SCPs). The quality of the results are dependent on the quantity of SCPs, the more SCPs, the better the results [14] but it is slow and inaccurate. The skill of the operator is a factor in the quality of the image. SCP based techniques are not suitable for bony structures without identifiable edges. Generally, SCP based techniques are used as part of a process involving other methods.[15]

Non-Stereo corresponding contour method (NCSS)

This method uses X-ray images for 3D Reconstruction and to develop 3D models with low dose radiations in weight bearing positions.

In NSCC algorithm, the preliminary step is calculation of an initial solution. Firstly anatomical regions from the generic object are defined. Secondly, manual 2D contours identification on the radiographs is performed. From each radiograph 2D contours are generated using the 3D initial solution object. 3D contours of the initial object surface are projected onto their associated radiograph.[15] The 2D association performed between these 2 set points is based on point-to-point distances and contours derivations developing a correspondence between the 2D contours and the 3D contours. Next step is optimization of the initial solution. Lastly deformation of the optimized solution is done by applying Kriging algorithm to the optimized solution.[16] Finally, by iterating the final step until the distance between two set points is superior to a given precision value the reconstructed object is obtained.

The advantage of this method is it can be used for bony structures with continuous shape and it also reduced human intervention but they are time-consuming.

Surface rendering technique

Surface rendering visualizes a 3D object as a set of surfaces called iso-surfaces. Each surface has points with the same intensity (called an iso-value). This technique is usually applied to high contrast data, and helps to illustrate separated structures; for instance, the skull can be created from slices of the head, or the blood vessel system from slices of the body. Two main methods are:

- Contour based reconstruction: Iso-contours are attached to each other to form iso-surfaces.[17]
- Voxel based reconstruction: Voxels of the same intensity value are used to form iso-surfaces. Popular algorithms are Marching Cubes, Marching Tetrahedrons and Dividing Cubes.[17]

Other methods use statistical shape models, parametrics, or hybrids of the two

See also

- 3D pose estimation – Process of determining spatial characteristics of objects
- 3D reconstruction – Process of capturing the shape and appearance of real objects
- 3D photography
- 2D to 3D conversion – Process of transforming 2D film to 3D form
- 3D data acquisition and object reconstruction – Scanning of an object or environment to collect data on its shape
- Epipolar geometry – Geometry of stereo vision
- Camera resectioning – Process of estimating the parameters of a pinhole camera model
- Computer stereo vision – Extraction of 3D data from digital images
- Structure from motion – Method of 3D reconstruction from moving objects
- Comparison of photogrammetry software
- Visual hull – is a geometric entity created by shape-from-silhouette
- Human image synthesis – Computer generation of human images

References

1. "Soltani, A. A., Huang, H., Wu, J., Kulkarni, T. D., & Tenenbaum, J. B. Synthesizing 3D Shapes via Modeling Multi-View Depth Maps and Silhouettes With Deep Generative Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1511-1519)" (https://github.com/Amir-Arsalan/Synthesize3DviaDepthOrSil) . *GitHub*. 6 March 2020.
2. C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization approach (http://repository.cmu.edu/cgi/viewcontent.cgi?article=3040&context=compsci) ", International Journal of Computer Vision, 9(2):137-154, 1992.
3. A. Laurentini (February 1994). "The visual hull concept for silhouette-based image understanding" (http://portal.acm.org/citation.cfm?coll=GUIDE&dl=GUIDE&id=628563) . *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **16** (2): 150–162. doi:10.1109/34.273735 (https://doi.org/10.1109%2F34.273735) .
4. R. Mohr and E. Arbogast. It can be done without camera calibration. Pattern Recognition Letters, 12:39-43, 1991.

5. O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.4708&rep=rep1&type=pdf) In Proceedings of the European Conference on Computer Vision, pages 563-578, Santa Margherita L., 1992.
6. E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. Sitz.-Ber.Akad.Wiss., Wien, math. naturw. Kl., Abt. IIa., 122:1939-1948, 1913.
7. S. J. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. International Journal of Computer Vision, 8(2):123-151, 1992.
8. O. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions (https://hal.inria.fr/docs/00/07/54/01/PDF/RR-1157.pdf) . International Journal of Computer Vision, 4(3):225-246, June 1990.
9. R. I. Hartley. Kruppa's equations derived from the fundamental matrix (http://users.rsise.anu.edu.au/hartley/public_html/Papers/kruppa/final-version/kruppa2.pdf) Archived (https://web.archive.org/web/20180622060523/http://users.rsise.anu.edu.au/hartley/public_html/Papers/kruppa/final-version/kruppa2.pdf) 2018-06-22 at the Wayback Machine. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(2):133-135, February 1997.
10. Pollefeys, Marc. Self-calibration and metric 3D reconstruction from uncalibrated image sequences (https://www.linkedin.com/pulse/web-development-companies-from-ukraine-november-2023-reviews-umen-7mm7f) . Diss. PhD thesis, ESAT-PSI, KU Leuven, 1999.
11. R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2nd edition, 2003.
12. "Medical Visualization: What is it and what's it for?" (https://garagefarm.net/blog/medical-visual-communication) . *GarageFarm*. 2018-02-18. Retrieved 2018-02-18.
13. "Pearcy MJ. 1985. Stereo radiography of lumbar spine motion. Acta Orthop Scand Suppl" (https://www.researchgate.net/publication/19301926) .
14. "Aubin CE, Dansereau J, Parent F, Labelle H, de Guise JA. 1997. Morphometric evaluations of personalised 3D reconstructions and geometric models of the human spine". *Med Biol Eng Comput*.
15. "S.Hosseinian, H.Arefi, 3D Reconstruction from multiview medical X-ray images- Review and evaluation of existing methods" (http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-1-W5/319/2015/isprsarchives-XL-1-W5-319-2015.pdf) (PDF).

16. Laporte, S; Skalli, W; de Guise, JA; Lavaste, F; Mitton, D (2003). "A biplanar reconstruction method based on 2D and 3D contours: application to distal femur" (https://www.researchgate.net/publication/10868711) . *Comput Methods Biomech Biomed Engin*. **6** (1): 1–6. doi:10.1080/1025584031000065956 (https://doi.org/10.1080%2F1025584031000065956) . PMID 12623432 (https://pubmed.ncbi.nlm.nih.gov/12623432) . S2CID 3206752 (https://api.semanticscholar.org/CorpusID:3206752) .

17. *G.Scott Owen, HyperVis. ACM SIGGRAPH Education Committee, the National Science Foundation (DUE-9752398), and the Hypermedia and Visualization Laboratory, Georgia State University*.

Further reading

- Yasutaka Furukawa and Carlos Hernández (2015) *Multi-View Stereo: A Tutorial* [1] (http://carlos-hernandez.org/papers/fnt_mvs_2015.pdf)
- Flynn, John, et al. "Deepstereo: Learning to predict new views from the world's imagery (https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Flynn_DeepStereo_Learning_to_CVPR_2016_paper.pdf) ." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

External links

- 3D Reconstruction from Multiple Images (http://dl.acm.org/citation.cfm?id=1754449&preflayout=tabs) - discusses methods to extract 3D models from plain images.
- Visual 3D Modeling from Images and Videos (https://sites.google.com/site/leeplus/bmvs) - a tech-report describes the theory, practice and tricks on 3D reconstruction from images and videos.
- Synthesizing 3D Shapes via Modeling Multi-View Depth Maps and Silhouettes with Deep Generative Networks (https://github.com/Amir-Arsalan/Synthesize3DviaDepthOrSil) - Generate and reconstruct 3D shapes via modeling multi-view depth maps or silhouettes.